52520UD01A01
62306UD01A01

Unidad Didáctica

Ingeniería Industrial

COMUNICACIONES INDUSTRIALES: PRINCIPIOS BÁSICOS

Manuel-Alonso Castro Gil
Gabriel Díaz Orueta
Francisco Mur Pérez
Rafael Sebastián Fernández
Elio Sancristóbal Ruiz
Víctor Miguel Sempere Paya
Javier Silvestre Blanes
Josep Maria Fuertes Armengol
Pau Marti Colom
José Gregorio Yepez Castillo
Perfecto Mariño Espiñeira
Miguel Ángel Domínguez Gómez
Ricardo Mayo Bayón

UNIVERSIDAD NACIONAL DE EDUCACIÓN A DISTANCIA

COMUNICACIONES INDUSTRIALES: PRINCIPIOS BÁSICOS

UNIDADES DIDÁCTICAS

Manuel-Alonso Castro Gil, Gabriel Díaz Orueta,
Francisco Mur Pérez, Rafael Sebastián Fernández,
Elio Sancristóbal Ruiz, Víctor Miguel Sempere Paya,
Javier Silvestre Blanes, Josep Maria Fuertes Armengol,
Pau Marti Colom, José Gregorio Yepez Castillo,
Perfecto Mariño Espiñeira, Miguel Ángel Domínguez Gómez
y Ricardo Mayo Bayón

COMUNICACIONES INDUSTRIALES: PRINCIPIOS BÁSICOS

UNED

Universidad Nacional de Educación a Distancia

UNIDADES DIDÁCTICAS (52520UD01A01)
COMUNICACIONES INDUSTRIALES: PRINCIPIOS BÁSICOS

Quedan rigurosamente prohibidas, sin la autorización escrita de los titulares del «Copyright», bajo las sanciones establecidas en las leyes, la reproducción total o parcial de esta obra por cualquier medio o procedimiento, comprendidos la reprografía y el tratamiento informático, y la distribución de ejemplares de ella mediante alquiler o préstamo públicos.

© UNIVERSIDAD NACIONAL
DE EDUCACIÓN A DISTANCIA - Madrid 2007

Librería UNED: c/ Bravo Murillo, 38 - 28015 Madrid.
Tels:. 91 398 75 60 / 73 73, e-mail: libreria@adm.uned.es

© Manuel-Alonso Castro Gil, Gabriel Díaz Orueta, Francisco Mur Pérez, Rafael Sebastián Fernández, Elio Sancristóbal Ruiz, Víctor Miguel Sempere Paya, Javier Silvestre Blanes, Josep María Fuertes Armengol, Pau Martí Colom, José Gregorio Yepez Castillo, Perfecto Mariño Espiñeira, Miguel Ángel Domínguez Gómez y Ricardo Mayo Bayón

ISBN: 978-84-362-5460-0
Depósito legal: M. 5.275-2007

Primera edición: febrero de 2007

Impreso en Fernández Ciudad, S. L.
Coto de Doñana, 10. 28320 Pinto (Madrid)
Impreso en España - Printed in Spain

A nuestros alumnos de las Universidades
de Oviedo, de Vigo,
de la Politécnica de Cataluña,
de la Politécnica de Valencia
y de la Universidad Nacional de Educación a Distancia,
para los que a lo largo de muchos años
hemos desarrollado los conocimientos plasmados
en estos libros.

A nuestros nuevos alumnos y lectores,
que esperamos sepan valorar
el gran esfuerzo realizado en escribir estos libros,
pioneros en su tema en castellano.

ÍNDICE

Presentación 25
Contenido 27

1. INTRODUCCIÓN A LAS REDES DE COMUNICACIONES ANALÓGICAS Y DIGITALES. ESTRUCTURA BÁSICA 29
   1.1. Introducción y orientaciones para el estudio 31
   1.2. Objetivos 31
   1.3. Fundamentos de la comunicación 31
   1.4. Introducción a los medios de transmisión de datos 35
      1.4.1. Características físicas del medio 35
         1.4.1.1. Características de las señales 36
         1.4.1.2. Propagación de las señales en el medio 39
            1.4.1.2.1. Atenuación 39
            1.4.1.2.2. Retardo 40
            1.4.1.2.3. Ruido 41
         1.4.1.3. Capacidad de transmisión del medio 42

1.4.2. Medios de transmisión 43
1.4.2.1. Medios guiados 44
1.4.2.2. Medios no guiados 46
1.5. Estructuras básicas en la comunicación 47
1.5.1. Transmisión de datos en paralelo 48
1.5.2. Transmisión de datos en serie 48
1.5.2.1. Modos de comunicación punto a punto 49
1.5.2.1.1. Modo simplex 49
1.5.2.1.2. Modo half duplex 49
1.5.2.1.3. Modo full duplex 50
1.5.2.2. Topologías de redes multipunto 50
1.5.2.2.1. Bus de datos (canal de distribución) 51
1.5.2.2.2. Estrella 51
1.5.2.2.3. De jerarquía o de árbol . 52
1.5.2.2.4. Anillo 52
1.5.2.2.5. Malla 52
1.5.2.2.6. Bucle 52
1.6. Introducción a la codificación de datos 53
1.6.1. Codificación en las comunicaciones analógicas 53
1.6.1.1. Información analógica 53
1.6.1.2. Información digital 55
1.6.1.2.1. Codificación en amplitud (ASK) 55
1.6.1.2.2. Codificación en frecuencia (FSK) 56
1.6.1.2.3. Codificación en fase (PSK) 56

1.6.2. Codificación en las comunicaciones digitales 58
1.6.2.1. Información digital 58
1.6.2.1.1. Codificaciones polares .. 59
1.6.2.1.2. Codificaciones bifase 59
1.6.2.1.3. Codificaciones bipolares 60
1.6.2.2. Información analógica 61
1.6.2.2.1. Teorema de muestreo de Shannon..... 61
1.6.2.2.2. Modulación por codificación de impulsos 64
1.6.2.2.3. Modulación delta..... 65
1.7. Principios generales de la multiplexación 66
1.7.1. Multiplexación por división de frecuencia (FDM) 66
1.7.2. Multiplexación por división en el tiempo (TDM) . 67
1.8. Protocolos y control de enlace de datos 68
1.8.1. Acceso al medio..... 69
1.8.1.1. Acceso controlado 70
1.8.1.2. Reserva 70
1.8.1.3. Acceso aleatorio 71
1.8.2. Control de enlace de datos 72
1.9. Conclusiones 72
1.10. Bibliografía 73
1.11. Evaluación 73
1.11.1. Ejercicios teóricos 73
1.11.2. Evaluación objetiva 76

2. MODELO OSI. ESTÁNDARES. PROTOCOLOS. MODELO TCP/IP..... 79
2.1. Introducción y orientación en el estudio..... 81

2.2. Objetivos 81
2.3. Arquitectura de protocolos 81
2.3.1. Arquitectura de protocolos normalizada 83
2.4. Modelo OSI 84
2.4.1. Capa física 86
2.4.2. Capa de enlace de datos 86
2.4.3. Capa de red 87
2.4.4. Capa de transporte 87
2.4.5. Capa de sesión 88
2.4.6. Capa de presentación 88
2.4.7. Capa de aplicación 88
2.5. Modelo TCP/IP 89
2.5.1. Internet y TCP/IP 92
2.5.2. Protocolo de Internet o IP 92
2.5.3. Protocolo de control de transmisión o TCP 96
2.5.4. Otros protocolos de TCP/IP 100
2.6. IP versión 6 101
2.6.1. Encabezado base 102
2.6.2. Encabezado de extensión 103
2.7. Conclusiones 104
2.8. Bibliografía 104
2.9. Evaluación 105
2.9.1. Evaluación objetiva 105

3. MEDIOS Y MODOS DE TRANSMISIÓN 109
3.1. Introducción y orientación en el estudio 111
3.2. Objetivos 111
3.3. Conceptos básicos de los medios físicos 112

3.3.1. El espectro electromagnético 113
3.3.1.1. Unidades y nomenclatura 117
3.3.2. Las bandas de comunicación y sus aplicaciones 118
3.3.2.1. Radiofrecuencia 119
3.3.2.2 Microondas 124
3.3.2.3. Banda óptica 126
3.4. Los conductores metálicos 130
3.4.1. Características 130
3.4.2. Cableado estructurado 133
3.4.3. Certificación de cables 135
3.5. Las fibras ópticas 137
3.5.1. Características 137
3.5.2. Tipos de modos 140
3.5.3. Nomenclatura de las portadoras ópticas 141
3.6. Los satélites espaciales de comunicaciones 143
3.6.1. Conceptos básicos 143
3.6.2. Tipos de órbitas 146
3.6.3. Servicios VSAT 148
3.7. La multiplexación y la modulación 150
3.7.1. Las señales analógicas y digitales 150
3.7.2. Multiplexación 154
3.7.3. El concepto de modulación 161
3.8. Modulación en banda base y banda ancha 162
3.8.1. Modulación en banda base 162
3.8.2. La modulación en banda ancha 163
3.8.2.1. Moduladora analógica 164
3.8.2.2. Moduladora digital 167

3.9. Aplicaciones de la multiplexación y modulación 169
3.9.1. Redes de cable CATV y PLC 169
3.9.2. Redes de cable xDSL 171
3.9.3. Redes de satélites 173
3.9.4. Buses de campo 179
3.10. Conclusiones 183
3.11. Bibliografía 184
3.12. Evaluación 184
3.12.1. Ejercicios resueltos 184
3.12.2. Evaluación objetiva 187

4. ACCESO AL MEDIO 191
4.1. Introducción y orientación para el estudio 193
4.2. Objetivos 194
4.3. Capa de enlace: enlace lógico y acceso al medio 194
4.4. Estructura del mensaje 196
4.5. Detección de errores 198
4.5.1. Códigos de palabras 198
4.5.2. Códigos de tramas 200
4.6. Control de errores 202
4.7. Control de flujo 204
4.8. Control de acceso al medio 204
4.8.1. Maestro/Esclavo 207
4.8.2. Paso de testigo delegado 208
4.8.3. Paso de testigo 209
4.8.4. Acceso múltiple por división de tiempo (TDMA) 209
4.8.5. Acceso múltiple por detección de portadora (CSMA) 211

4.8.5.1. CSMA/CD 211
4.8.5.2. CSMA/CR/CA 212
4.9. Evaluación de redes 213
4.10. Conclusiones 215
4.11. Bibliografía 216
4.12. Evaluación 216
4.12.1. Ejercicios resueltos 216
4.12.2. Evaluación objetiva 222

5. REDES DE ÁREA LOCAL. COMPONENTES 225
5.1. Introducción y orientaciones al estudio 227
5.2. Objetivos 227
5.3. Características generales de las LAN 228
5.4. Topologías y medios de transmisión 228
5.4.1. Topologías 228
5.4.1.1. Topología en bus 229
5.4.1.2. Topología en árbol 231
5.4.1.3. Topología en anillo 232
5.4.1.4. Topología en estrella 235
5.4.2. Medios de transmisión 236
5.4.2.1. Cable de par trenzado 236
5.4.2.2. Cable coaxial 237
5.4.2.3. Fibra óptica 238
5.4.2.4. Transmisión inalámbrica 239
5.5. Modelo IEEE 802, IEEE 802.3 y ETHERNET 240
5.5.1. Modelo de referencia IEEE 802 240
5.5.1.1. Control de enlace lógico 241

5.5.1.1.1. Protocolo de control de enlace lógico 242
5.5.1.2. Control de acceso al medio 243
5.5.1.2.1. Algoritmos de asignación de canal de acceso múltiple 244
5.5.1.2.2. La trama MAC 246
5.5.2. IEEE 802.3 y ETHERNET 246
5.5.3. Redes de área local virtuales 248
5.6. Dispositivos de interconexión de redes 249
5.6.1. Repetidores 249
5.6.2. Puentes 249
5.6.3. Encaminadores o Routers 251
5.6.4. Pasarela o Gateway 252
5.7. Redes de área local de alta velocidad 252
5.7.1. IEEE 802.3 100 Mbps (FAST ETHERNET) 253
5.7.2. GIGABIT ETHERNET 253
5.7.3. Canal de Fibra 255
5.8. Redes inalámbricas 257
5.8.1. Requisitos de las redes de área local inalámbricas 258
5.8.2. Clasificación de LAN inalámbricas 259
5.8.3. Arquitectura del 802.11 260
5.8.3.1. Pila de protocolos 260
5.8.3.1.1. Capa física 260
5.8.3.1.2. Capa de enlace de datos 262
5.8.3.2. Servicios 263
5.9. Aplicaciones 264
5.9.1. Bus de campo o Fieldbus 264

5.9.1.1. Profibus 265
5.9.2. Redes LAN de ordenadores personales 265
5.9.3. Redes de respaldo y almacenamiento 266
5.10. Conclusiones 266
5.11. Bibliografía 267
5.12. Evaluación 267
5.12.1. Evaluación objetiva 267

6. REDES DE ÁREA AMPLIA. COMPONENTES. REDES IP 271
6.1. Introducción y objetivos 273
6.2. Características generales de las redes de área amplia .. 273
6.3. Protocolos de redes de área amplia: HDLC, PPP, FR, etc. 275
6.3.1. Protocolo HDLC 275
6.3.2. Protocolo PPP 276
6.3.3. Protocolo *Frame Relay* 277
6.3.4. Dos aproximaciones al mismo problema: RDSI y ADSL 279
6.4. Las redes IP: redes WAN públicas 283
6.5. Direccionamiento IP y encaminamiento en IP 284
6.6. Los encaminadores como componentes típicos de WAN 289
6.7. Comunicación IP cliente/servidor mediante *sockets* 292
6.8. Operación de aplicaciones TCP, orientadas a conexión 294
6.9. Operación de aplicaciones UDP 296
6.10. Operación de aplicaciones del nivel de red 296
6.11. Conclusiones 299
6.12. Bibliografía 299
6.13. Evaluación 300

6.13.1. Evaluación objetiva 300

7. SISTEMAS DE COMUNICACIONES INALÁMBRICOS Y MÓVILES 303

7.1. Introducción y orientaciones para el estudio 305

7.2. Objetivos 305

7.3. Conceptos básicos de la propagación inalámbrica 306

7.3.1. Efectos de la propagación 306

7.3.2. Reducción de los efectos de la propagación 309

7.3.3. Métricas de prestaciones y factor de mérito 313

7.4. Diseño de redes inalámbricas y móviles 316

7.4.1. Elementos básicos 316

7.4.2. Planificación y dimensionamiento de una red de radiocomunicación 319

7.4.3. Evolución tecnológica 321

7.4.3.1. Sistemas MIMO 322

7.4.3.2. Sistemas de radio definido por programación 323

7.5. Redes de área local móviles (WLAN) 323

7.5.1. Introducción 323

7.5.2. La Norma IEEE 802.11 330

7.5.3. HIPERLAN 333

7.5.4. Óptica del espacio libre (FSO) 334

7.6. Redes de área personal (WPAN) 336

7.6.1. Introducción 336

7.6.2. Bluetooth (IEEE 802.15.1) 336

7.6.3. UBW (IEEE 802.15.3) 337

7.6.4. ZigBee (IEEE 802.15.4) 338

7.6.5. IrDA 342

7.7. Redes fijas de acceso inalámbrico (WMAN) 343
7.7.1. Banda ancha (MMDS y LMDS) 343
7.7.2. Normas ETSI 344
7.7.3. WiMAX (IEEE 802.16) 345
7.7.4. IEEE 802.20 e IEEE 802.22 348
7.7.5. Banda estrecha (WLL) 349
7.8. Redes de acceso celular (WWAN) 350
7.8.1. Configuración de los sistemas celulares 350
7.8.2. Normas internacionales 354
7.9. Redes de satélites espaciales 358
7.9.1. Sistemas celulares basados en satélites 358
7.9.2. Servicios de navegación global 359
7.10. Las redes inalámbricas y los buses de campo 363
7.10.1. Tecnologías inalámbricas en redes industriales 363
7.10.2. Redes en malla (*Mesh*) 365
7.10.3. Ingeniería de protocolos de capa cruzada (*Cross-Layer*) 366
7.11. Conclusiones 367
7.12. Bibliografía 367
7.13. Evaluación 368
7.13.1. Ejercicios resueltos 368
7.13.2. Evaluación objetiva 370

8. SEGURIDAD EN LAS COMUNICACIONES. CALIDAD DEL SERVICIO 373
8.1. Introducción y objetivos 375
8.2. Factores de inseguridad en sistemas y en dispositivos .. 376
8.2.1. Problemas de seguridad física 377

8.2.2. Problemas generales de seguridad en sistemas operativos y aplicaciones 379
8.2.2.1. Mala codificación 380
8.2.2.2. Ingeniería social 381
8.2.3. Problemas comunes de seguridad en sistemas operativos 382
8.2.4. Problemas comunes de seguridad en aplicaciones 383
8.2.4.1. Implementación incorrecta de protocolos 384
8.2.4.2. Diseño peligroso de protocolos y aplicaciones 385
8.2.4.3. Incorrecta selección de protocolos de aplicación 387
8.2.4.4. Otros 387
8.2.5. Factores de inseguridad en dispositivos de comunicaciones 388
8.3. Soluciones actuales en sistemas y dispositivos de seguridad 390
8.4. Política de seguridad para redes 392
8.5. Defensas no criptográficas 395
8.5.1. Cortafuegos 395
8.5.2. Sistemas de detección de intrusiones 398
8.5.3. Detectores de vulnerabilidades 399
8.6. Defensas criptográficas 400
8.6.1. Algoritmos criptográficos: Clave privada, funciones *Hash* y clave pública 402
8.6.2. Protocolos criptográficos: SSL, IPSec y otros 407
8.6.3. Redes privadas virtuales 410
8.7. Seguridad en protocolos industriales 411
8.7.1. OPC 411

8.7.2. MMS 412
8.7.3. IEC 61850 413
8.7.4. IICP 413
8.7.5. Recomendaciones de políticas de seguridad 414
8.8. Introducción al concepto de calidad del servicio en redes 414
8.9. Aproximaciones estándar a la calidad del servicio en redes 415
8.10. Los encaminadores y la calidad del servicio..... 418
8.11. Conclusiones 419
8.12. Bibliografía 420
8.13. Evaluación 420
8.13.1. Evaluación objetiva 420

9. INTRODUCCIÓN A LAS COMUNICACIONES INDUSTRIALES Y A LOS SISTEMAS DE TIEMPO REAL..... 423
9.1. Introducción y orientaciones para el estudio..... 425
9.2. Objetivos 426
9.3. Características de las redes de comunicación industrial 427
9.4. Modelo de integración de redes de comunicación..... 429
9.4.1. Integración de la red de comunicaciones 429
9.4.2. Redes de control y redes de datos 431
9.5. Comunicación y sistemas de tiempo real 432
9.5.1. Sistemas de tiempo real 432
9.5.2. Comunicación en tiempo rea 433
9.5.3. Comunicación en los sistemas distribuidos de control en tiempo real 434
9.6. Sistemas disparados por tiempo y disparados por evento 435
9.6.1. Mensajes críticos y no críticos 437

9.6.2. Mensajería periódica y aperiódica 439
9.7. Planificación de mensajes de tiempo real 439
9.7.1. Políticas de planificación..... 440
9.7.2. Test de planificabilidad 441
9.7.3. Planificación con desalojo 441
9.7.4. Planificación sin desalojo 442
9.8. Prioridades estáticas y dinámicas 442
9.8.1. Planificación cíclica 443
9.8.2. Planificación estática 444
9.8.3. Planificación dinámica 445
9.8.4. Algoritmo de planificación *Rate Monotonic*..... 445
9.8.5. Algoritmo de planificación *Deadline Monotonic* . 447
9.8.6. Algoritmo de planificación *Earliest Deadline First*..... 451
9.9. Análisis de tiempo real en redes de comunicación..... 453
9.10. Conclusiones 454
9.11. Bibliografía..... 455
9.12. Evaluación 455
9.12.1. Evaluación objetiva..... 455

10. INTRODUCCIÓN A LOS BUSES DE CAMPO Y A LOS SISTEMAS DISTRIBUIDOS 459
10.1. Introducción y orientaciones para el estudio..... 461
10.2. Objetivos..... 462
10.3. Los buses de campo. Definición..... 462
10.4. Los sistemas distribuidos..... 462
10.5. Estructuración de los sistemas distribuidos industriales. 463
10.6. Características de distribución..... 465

10.7. Características de la mensajería de los buses de campo 467
10.8. Normalizaciones en los buses de campo 469
10.9. Modelo general de operación de los buses de campo ... 470
10.10. Funcionalidades comunes 477
10.11. Conclusiones..... 479
10.12. Bibliografía 480
10.13. Evaluación..... 480
10.13.1. Ejercicios resueltos..... 480
10.13.2. Evaluación objetiva..... 481

ANEXO.

SOLUCIONES A LAS PREGUNTAS DE EVALUACIÓN OBJETIVA..... 485

PRESENTACIÓN

Las comunicaciones industriales son una de las áreas en auge dentro del amplio mundo de las comunicaciones, al unirse en un mismo entorno los temas empresariales (y más en detalle, los temas de fabricación más ligados a la industria en su más amplio y básico aspecto) y las comunicaciones, como completo (y complejo) soporte para la implantación tecnológica dentro de la empresa.

¿Y que complemento mejor para la Ingeniería Industrial y nuestra industria, que dotar a sus lectores, aficionados, estudiantes, amigos o practicantes, de los conocimientos precisos de las tecnologías de las comunicaciones, listas para ser integradas en el mundo industrial?

Este ha sido el objetivo que nos hemos marcado un grupo de profesores de cinco Universidades Públicas Españolas (Universidad Nacional de Educación a Distancia, Universidad Politécnica de Valencia, Universidad Politécnica de Cataluña, Universidad de Vigo y Universidad de Oviedo) coordinados desde la experiencia en trabajo colaborativo desde la propia Universidad Nacional de Educación a Distancia, la UNED, para elaborar dos libros, que aporten estos conocimientos dentro de una profesión en la que, hasta ahora, las comunicaciones ha sido una de las materias menos desarrollada.

Ambos libros presentan un objetivo común, cubrir los temas de las comunicaciones industriales, pero de una forma complementaria entre sí.

Así, el primer libro que está leyendo en este momento, **Comunicaciones Industriales: Principios básicos**, pretende dotar a su lector de los conocimientos previos necesarios para poder entender, conocer y asimilar las comunicaciones, desde las bases de las comunicaciones analógicas y digitales, la codificación, medios y modos de transmisión, los protocolos y estándares existentes en este dinámico y cambiante mundo, los componentes y elementos existentes en las redes locales, amplias, personales o inalámbricas, sin olvidarse de la seguridad y el concepto más avanzado actualmente como es la calidad en el servicio. Por último este primer libro tiende un puente al segundo, estableciendo los principios de las comunicaciones industriales, los sistemas de tiempo real, los buses de campo y los sistemas distribuidos

El segundo libro, **Comunicaciones Industriales: Sistemas Distribuidos y Aplicaciones,** presenta una visión más aplicada para un lector ya "introducido" en las comunicaciones, pero interesado en entrar como complemento en el mundo de las comunicaciones industriales.

Así, comienza este libro complementando la visión de los procesos industriales, la fabricación y los sistemas en tiempo real, ya esbozados en el primer libro, para complementar la introducción general de las comunicaciones industriales, los sistemas distribuidos y su programación, para centrase y entrar en profundidad en los estándares de las redes industriales y los buses de campo, y a continuación describir en detalle varios sistemas de comunicaciones de amplio uso en los entornos industriales: PROFIBUS, WorldFIP, CAN y Ethernet industrial.

Sin olvidar amplios sectores de la automatización y el control, como son la domótica e inmótica (en pleno auge igualmente en nuestro país), una revisión final de aplicaciones en otras áreas, como son la automoción, la navegación marítima o el sector aéreo y espacial, para acabar con los sistemas con una vertiente más electrónica (instrumentación) o de control.

Manuel-Alonso Castro Gil

Coordinador de los libros

CONTENIDOS

Este primer libro, **Comunicaciones Industriales: Principios básicos**, incluye los conocimientos previos necesarios para poder entender, conocer y asimilar las comunicaciones en general, y las comunicaciones industriales como objetivo y aplicación.

El tema 1 presenta de forma general, pero con un suficiente nivel de detalle, la estructura básica de las comunicaciones analógicas y digitales, incluyendo la transmisión de datos y los distintos medios existentes para ello, así como los diferentes métodos de codificación y multiplexación, para mejorar las prestaciones del canal de comunicación, así como los distintos sistemas de control con el enlace que se realiza para establecer la comunicación con el medio. El tema 3, modos y medios de transmisión, profundiza en detalle en cada uno de los temas y técnicas propuestas, mientras que el tema 4, acceso al medio, hace lo propio con la estructura de mensajes y el control de flujo.

El tema 2 formaliza la estructuración de los modelos y protocolos empleados en los sistemas de comunicaciones, básicamente OSI y TCP/IP, el primero, el estándar de comunicaciones por excelencia, y el segundo, el estándar de facto usado en la mayoría de las redes, con Internet como ejemplo fundamental.

Los temas 5, 6 y 7 desarrollan los componentes y elementos existentes en diferentes tipos de redes, organizando estas según su distribución geográfica, dentro de un entorno de área local, en el primer caso, y de un entorno más

amplio, en el segundo caso, o de una forma deslocalizada, cubriendo el emergente mundo de las comunicaciones inalámbricas y móviles.

El tema 8 cubre uno de los aspectos actualmente más candentes de las comunicaciones, la seguridad, sobre todo al difundirse el modelo de Internet y su acceso de forma omnipresente en las comunicaciones industriales. Se cubre igualmente el concepto actual de la calidad en el servicio, así como varios protocolos nuevos recientemente desarrollados o en desarrollo, sobre las implementaciones de la seguridad en las comunicaciones.

Los dos últimos temas, el 9 y 10, establecen los principios de las comunicaciones industriales, los sistemas de tiempo real y sus requerimientos específicos, los buses de campo y los sistemas distribuidos, como un todo que se aplica de una forma completa (y como sería deseable, transparente) en las industrias y empresas.

El segundo libro, **Comunicaciones Industriales: Sistemas Distribuidos y Aplicaciones,** presenta una visión más aplicada para un lector ya "introducido" en las comunicaciones, ayudándole a entrar de pleno en el mundo de las comunicaciones industriales.

En cada capítulo se ha incluido una serie de preguntas de autoevaluación, cuyas soluciones se han recogido en un Anexo al final de cada libro. En muchos de los capítulos también se han intercalado diversos temas y ejercicios, tanto prácticos como teóricos, para que el lector pueda ir conociendo su asimilación de forma estructurada.

TEMA 1

INTRODUCCIÓN A LAS REDES DE COMUNICACIONES ANALÓGICAS Y DIGITALES. ESTRUCTURA BÁSICA

1.1. Introducción y orientaciones para el estudio
1.2. Objetivos
1.3. Fundamentos de la comunicación
1.4. Introducción a los medios de transmisión de datos
1.5. Estructuras básicas en la comunicación
1.6. Introducción a la codificación de datos
1.7. Principios generales de la multiplexación
1.8. Protocolos y control de enlace de datos
1.9. Conclusiones
1.10. Bibliografía
1.11. Evaluación

1.1. INTRODUCCIÓN Y ORIENTACIÓN PARA EL ESTUDIO

En este capítulo se exponen los aspectos fundamentales de la comunicación. Introduciendo conceptos que posteriormente serán ampliados y tratados con más detalle en otros capítulos del libro. El planteamiento comienza con el análisis, en grandes bloques, de un proceso de comunicación cotidiano, planteando la problemática y necesidades que se presentan en cada uno de estos grandes bloques. Desde este análisis previo se plantean los conceptos fundamentales, empezando por el medio físico y terminando con los primeros conceptos y necesidades de protocolos en las comunicaciones. Por el camino se habrán tratado temas de topologías de red, codificación, multiplexación, control de enlace, sincronización o control de errores.

1.2. OBJETIVOS

Los objetivos de este capítulo son plantear la base conceptual que permita al lector abordar con éxito los temas 2, 3 y 4 del libro, en donde se tratará con mayor amplitud muchos de los temas tratados a continuación.

1.3. FUNDAMENTOS DE LA COMUNICACIÓN

La necesidad de comunicación entre dos sistemas, más o menos distantes (tomando un concepto amplio de sistema), está presente desde el origen de la vida y es intrínseco a multitud de procesos, tanto creados por el hombre como naturales. Se puede pensar en sonidos, olores, señales luminosas u otras formas de comunicación que son utilizadas en el reino animal, para mostrar multitud de formas de comunicación. También se puede pensar en sistemas fisiológicos internos de los seres vivos, como el sistema nervioso y, por supuesto, en todos los sistemas tecnológicos que el hombre ha desarrollado, desde el teléfono de Alexander Graham Bell hasta la red de comunicaciones más compleja que se pueda imaginar. Todos estos procesos de comunicación tienen básicamente la misma estructura y el mismo objetivo: intercambiar información entre dos entidades. Para iniciar el estudio de “la comunicación” no es necesario buscar procesos complejos y, por ello, se analizará el proceso de comunicación oral, abstrayendo sus fundamentos para representarlo en el diagrama de la Figura 1.1.

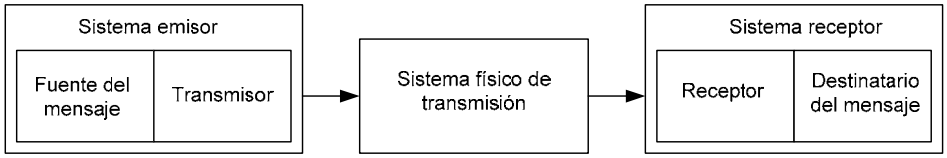


FIGURA 1.1. Modelo conceptual de comunicación.

La primera impresión es que el sistema de comunicación está compuesto por tres grandes bloques: Un elemento que necesita comunicar algo, sistema emisor (la persona que habla); un elemento que soporta la transmisión, sistema de transmisión (en el caso que se plantea como modelo, será el aire) y un elemento que recibe la señal, el sistema receptor (la persona que escucha). Si se analiza con más detalle el proceso, se comprenderá que estos grandes bloques son en realidad sistemas notablemente complejos. Así el sistema emisor consta de dos grandes subsistemas: el elemento que produce la información, por ejemplo, el cerebro del orador, que es la **fuente del mensaje** y los elementos que trasforman la idea en las señales que se desplazan por el aire, constituye el **transmisor**. En el otro extremo se produce el proceso contrario, hay un **receptor** que interpreta las señales que llegan por el aire y un elemento que reconoce la información recogida por el receptor, el **destinatario del mensaje**. A continuación se analizan estos elementos.

La **fuente del mensaje** es el dispositivo que genera los datos y da servicio a la necesidad de transmitirlos, decidiendo también el destinatario de los mismos. En el ejemplo de la comunicación oral, es el cerebro, en los sistemas informáticos o en entornos industriales, se encuentran distintos dispositivos electrónicos, sensores u ordenadores. Estos complejos dispositivos no sólo generan la información que se transmite sino que intervienen en el proceso de comunicación en las tareas más de mayor nivel del proceso de comunicación, es decir de las funciones más alejadas del medio físico de transmisión. El ordenador que contiene los datos a transmitir decide a donde los manda y para qué aplicaciones, entre otras múltiples funciones.

El transmisor. El formato de los datos es propio del sistema que los genera y no tiene por qué ser compatible con el medio de transmisión. Los elementos que componen el transmisor transforman y codifican la información, generando señales compatibles con el medio de transmisión que se va a utilizar. El transmisor ha de utilizar todas las técnicas que posibilitan el intercambio correcto de datos y utilizar los procedimientos adecuados para aprovechar la capacidad del medio de transmisión. Estos aspectos se desarrollarán con más profundidad en los epígrafes que recogen la **modulación de la señal**, la

codificación, el **control de enlace** y la **multiplexación** de señales. Los ordenadores utilizan dos elementos de transmisión típicos: el módem y las tarjetas de red. El módem es un buen ejemplo de este tipo de dispositivos ya que transforma la información digital generada por el ordenador en señales analógicas aptas para ser transmitidas por una línea telefónica convencional.

Los aspectos a estudiar en el **sistema físico de transmisión** son tanto al soporte físico de la trasmisión por donde se propagan las señales (el **medio físico**: aire, par de cobre trenzado, cable coaxial, fibra óptica, etc.), como a la estructura y **topología** del sistema de transmisión, incluyendo los dispositivos que forman parte de esta red. En redes grandes y complejas se utilizan distintos equipos para regenerar las señales (**repetidores**), para transformarlas y hacerlas aptas para otros soportes físicos (**puentes**), para dirigir los paquetes de información por la red (**encaminadores**) o para conectar sistemas que utilizan sistemas de comunicación y lenguajes completamente diferentes (**pasarelas**). La comunicación oral directa o los teléfonos móviles utilizan el aire como soporte, una red Ethernet utiliza cables de cobre. El proceso de comunicación tiene algunas similitudes en estos ejemplos, considerando que hay un elemento transmisor que “habla” al medio, difundiéndose el mensaje por todo él y todos los demás sistemas “oyen” aunque normalmente sólo habrá uno que “escuche”, que es el receptor; todas ellas se basan en **redes de difusión**, en contraposición a las **comunicaciones punto a punto**. Sin embargo, también hay grandes diferencias en el medio de transmisión, pudiendo hacer otra clasificación como **medios guiados** y **medios no guiados**.

Otra clasificación de la red de interconexión existente entre el emisor y el receptor se puede realizar en función de que se establezca físicamente la conexión o no. La comunicación telefónica tradicional, para establecer la comunicación, conecta físicamente un circuito entre el emisor y el receptor. Esta técnica se denomina **conmutación de circuitos**. En otras técnicas de comunicación, que se parecen más al servicio de correo postal, el emisor hecha un “paquete de datos” en la red indicando la dirección de destino y dinámicamente se establece el recorrido óptimo, sin que exista una ruta preestablecida, esto se denomina **conmutación de paquetes**. Este sistema aporta la ventaja de que en caso de algún fallo en la red, los dispositivos que la forman son capaces de establecer una ruta alternativa para el envío del paquete.

Para la gestión y el funcionamiento de la red se intercambia gran cantidad de información que posibilita y facilita la comunicación de los sistemas conectados a ella. A todo este conjunto de mensajes, procedimientos y mecanismos de control y gestión es a lo que se denomina **protocolos** de la red.

El receptor. Una vez que los datos han llegado “al otro lado” es necesario captarlos, decodificarlos y recomponer el mensaje original que produjo la fuente, realizando el proceso inverso al transmisor. El proceso de captura de la información desde el medio, implica una tarea de reconocimiento del comienzo y del final de la transmisión, y la capacidad de recibir los datos al ritmo que los produce el transmisor, todo este proceso se engloba en los mecanismos de **sincronización** de la comunicación. Una vez que el transmisor ha decidido el destinatario, hay unos mecanismos para que el receptor identifique el mensaje como suyo. Este proceso es distinto si pensamos en una llamada telefónica o en una conversación en voz alta que se produce en la calle. En el primer caso es el transmisor el que identifica el receptor en una comunicación punto a punto con un único receptor posible. El segundo caso, el emisor también se dirige a un receptor único pero escuchan todos, por lo que tendrá que ser el receptor el que identifique el mensaje como propio.

Sin embargo aunque el receptor escuche el mensaje puede no recibirlo correctamente, por lo que el receptor debe tener la capacidad para reconocer que el mensaje ha llegado íntegro con un sistema de de **detección o corrección de errores**. También es importante prever sistemas de recuperación de la comunicación, para solucionar posibles interrupciones de la misma y continuar en el punto en el que se interrumpió, cuando esto sea posible. Todo esto entra en lo que se denomina la gestión del intercambio de información y **control del flujo**. En una comunicación de un ordenador por la línea telefónica el módem será el encargado de realizar estas tareas, recibirá la señal analógica de la línea telefónica, reconocerá la información que llega, verificará la integridad de los datos y la transformará en la información digital que salió del ordenador fuente.

El destinatario del mensaje toma los datos del receptor y los interpreta. Evidentemente, la fuente y el destinatario del mensaje deben utilizar el mismo lenguaje, el mismo sistema de codificación de caracteres y los mismos formatos de mensaje para entenderse. También se consideran cuestiones de **seguridad**, **privacidad** y **autenticidad** en la comunicación. La red debe permitir una difusión limitada de la información garantizando que la recibe únicamente el destinatario, permitir identificar con garantía el origen de los datos y la autenticidad del contenido del mensaje.

El diagrama de bloques, mostrado en la Figura 1.1, presenta un modelo, que por cotidiano es casi trivial, del proceso de comunicación. A pesar de esto, el análisis un poco más detenido de las necesidades inherentes al proceso, permite entrever la complejidad del mismo y los distintos aspectos a tener en cuenta, aspectos en los que se profundiza en el resto de capítulos de este texto.

1.4. INTRODUCCIÓN A LOS MEDIOS DE TRANSMISIÓN DE DATOS

Las redes de comunicaciones, locales, metropolitanas o de área amplia, en la actualidad utilizan soportes físicos muy diversos para adaptarse a las características específicas de comunicación y en función del momento tecnológico en el que se realiza. Estos medios se refieren, por ejemplo, a los que enlazan los ordenadores de una oficina entre sí, a los que enlazan los domicilios de los abonados con las centrales telefónicas, a los que interconectan redes de área local entre sí o centrales de conmutación telefónica, incluyendo redes de distintos países. Como consecuencia de la diversidad de redes existente, de las distintas necesidades de sus usuarios, incluso de la titularidad de la red (si la red es pública o privada), los requisitos que se imponen a la red son muy diferentes y por tanto los medios físicos utilizados son muy diversos. A pesar de esto, las características físicas de los soportes que permiten la propagación de las señales electromagnéticas (incluyendo también la información a transmitir) vienen a ser cualitativamente las mismas, aunque no cuantitativamente.

1.4.1. Características físicas del medio

El medio físico es el que finalmente realiza la transmisión de datos entre el emisor y el receptor. Por tanto, interesa conocer las características de propagación de estas señales electromagnéticas por él, para determinar su capacidad de transmisión.

Las propiedades de propagación de las señales electromagnéticas por los materiales que hacen de soporte de la comunicación, producen una limitación física que imposibilita superar ciertos valores en la velocidad de transmisión de datos. Como anécdota, comentar que el “bucle de abonado” (el cable que une el teléfono de nuestro domicilio con la central de conmutación telefónica) tiene un límite relativamente bajo que restringe la capacidad de comunicación a través de este medio.

Esta es la limitación que se encuentra un dispositivo como el módem para comunicarse con otro dispositivo exterior. De hecho, los módem, según se fueron desarrollando nuevas técnicas de modulación multinivel, fueron mejorando su velocidad de comunicación llegando a velocidades de 56 kbps (bits por segundo) con líneas de buena calidad (poco ruido). Valor raramente alcanzable aun cuando el módem tiene capacidad para hacerlo. Sin embargo

actualmente se fabrican dispositivos capaces de transmitir por ese medio muchos más datos por segundo que el módem (hasta 20 Mbps). Son los moduladores que utilizan tecnología xDSL ("lazo digital de abonado" o *Digital Subscriber Line*). Evidentemente el medio es el mismo para ambos dispositivos y sus limitaciones físicas no se pueden modificar, entonces ¿qué ha cambiado? El "truco" está en aprovechar mejor el medio, explotando mejor su capacidad de transmisión, utilizando otros algoritmos de codificación. Esto es posible gracias a que hoy se dispone de mejores y más rápidos procesadores para el tratamiento digital de la señal.

1.4.1.1. Características de las señales

La forma habitual de representar y estudiar las señales (formas de onda), es a partir de una función analítica dependiente del tiempo, g(t), en la que se presenta en el eje de abscisas el tiempo y en el eje de ordenadas la amplitud. En este **dominio temporal** se pueden diferenciar dos tipos de señales (Fig. 1.2): las **señales continuas**, que se corresponden con magnitudes analógicas y que toman valores continuos, es decir sin saltos bruscos o discontinuidades en la amplitud, y las **señales discretas**, utilizadas por los sistemas digitales y que presentan un conjunto reducido y limitado de valores de amplitud, generalmente dos, variando bruscamente entre dichos valores con el tiempo[1].

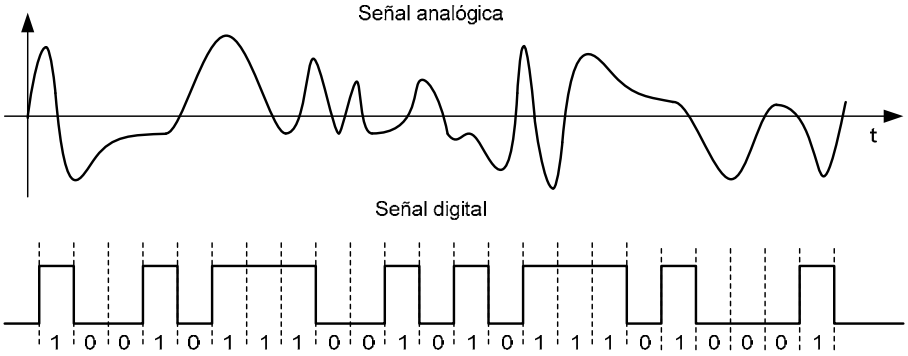


FIGURA 1.2. Señales analógicas y digitales.

[1] Las propiedades físicas del proceso de difusión de las ondas electromagnéticas impiden que la señal cambie de valor en tiempo cero (que implica una pendiente infinita), como se muestra en la Figura 1.2, pero la variación será tan rápida como permita el medio de propagación.

La señal continua más típica es una señal sinusoidal pura (Fig. 1.3), representada por la ecuación:

$$g_s(t) = A.sen\ (2.\pi.f.t+\varphi) \qquad (1.1)$$

presentando esta función tres parámetros que la caracterizan y que en general permiten caracterizar a todas las señales periódicas:

A: Amplitud de la onda, que es valor máximo que puede alcanzar la señal. Las unidades para la amplitud son las mismas que se utilizan para la magnitud física representada.

f: Frecuencia. Al tratarse de una señal periódica (que repite un patrón periódicamente) este parámetro representa el número de veces que se repite ese patrón (o ciclo) en un segundo. La unidad de medida son ciclos por segundo también llamados Hertzios o Hercios [Hz], que es una unidad equivalente a 1/segundos o segundos^{-1}. Relacionado con la frecuencia está el **periodo,** T = 1/f [segundos], y la **frecuencia angular,** $\omega = 2.\pi.f$ [radianes].

φ: Fase. Este parámetro produce un adelanto (si es positivo) o retraso (si es negativo) en el tiempo de la señal. Si la fase es cero, la onda pasa por el origen de coordenadas. En la ecuación 1.1, el ángulo de fase debe estar en radianes, y con la conversión adecuada, se podría dar la fase en grados o el tiempo de desfase t_d en segundos, midiendo el tiempo desde que la señal se hace cero ($g_s(t) = 0$) hasta el instante t = 0. El desfase será $\varphi = 2.\pi.f\ .t_d$.

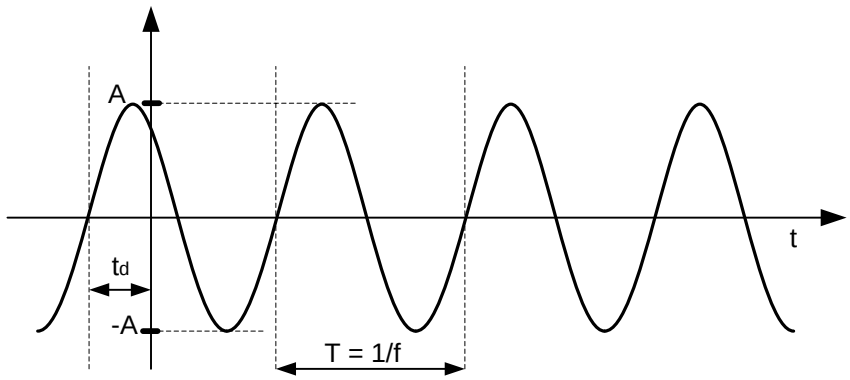


FIGURA 1.3. Señal sinusoidal pura y sus parámetros.

Para estudiar funciones periódicas cualesquiera, en las que cada T segundos se repite un patrón no sinusoidal, se recurre a la descomposición en serie de Fourier. Joseph Fourier demostró que cualquier función periódica se puede descomponer en una suma de funciones sinusoidales, con una senoide fundamental de frecuencia f = 1/T y un conjunto de senoides "armónicas" de frecuencias múltiplo de la fundamental (2.f, 3.f, 4.f, etc.). Habrá un armónico principal o primero de frecuencia f y amplitud A_1 $\{A_1.\text{sen}\,(2.\pi.f.t)\}$, un armónico segundo de frecuencia 2.f y amplitud A_2 $\{A_2.\text{sen}\,(4.\pi.f.t)\}$, etc., y un armónico enésimo de frecuencia n.f y amplitud A_n $\{A_n.\text{sen}\,(2.\pi.n.f.t)\}$, teniendo en cuenta que pueden presentarse infinitos armónicos en la descomposición.

A partir de esta descomposición matemática se puede representar y, lo que es más importante, estudiar cualquier función periódica como la suma de un conjunto de funciones sinusoidales. Esto se realiza en el **dominio de la frecuencia**, en el que las funciones vendrán representadas por su frecuencia fundamental f y el conjunto de valores de las amplitudes de sus armónicos A_1, A_2, …, A_n. También puede aparecer el coeficiente A_0 que representa una señal continua (invariante en el tiempo).

$$g_2 = A_0 + A_1.\text{sen}\,(2.\pi.f.t) + A_2.\text{sen}\,(2.\pi.2.f.t) + \ldots + A_n.\text{sen}\,(2.\pi.n.f.t) \qquad (1.2)$$

A cada uno de estos elementos sinusoidales de les denomina **armónicos**, y al conjunto de frecuencias de todos los armónicos se le denomina **espectro de frecuencias**, que puede ser infinito para funciones teóricas. El intervalo que cubren estas frecuencias es el **ancho de banda** de la señal. Dado que este ancho de banda puede ser infinito o al menos extenderse mucho, se habla del ancho de banda efectivo de la señal, que agrupa las frecuencias que recogen la mayor parte de la energía de la señal.

Una onda cuadrada periódica tiene una descomposición en funciones sinusoidales infinita y por tanto un ancho de banda infinito. Es interesente ver como se forma la onda cuadrada según se van sumando nuevos armónicos a la serie (constituida por infinitos armónicos impares).

$$g_c = \Sigma\ 1/n.\text{sen}\,(2.\pi.n.f.t);\ \text{con}\ n = 1, 3, 5, 7, \ldots, \infty \qquad (1.3)$$

La conclusión es que se puede estudiar cualquier onda periódica, ya sea continua o discreta (analógica o digital) como si fueran conjuntos de ondas sinusoidales puras, y esto también es aplicable a las fenómenos de propagación por el medio, conociendo como se comportan las ondas sinusoidales de distintas frecuencias en él (Fig. 1.4).

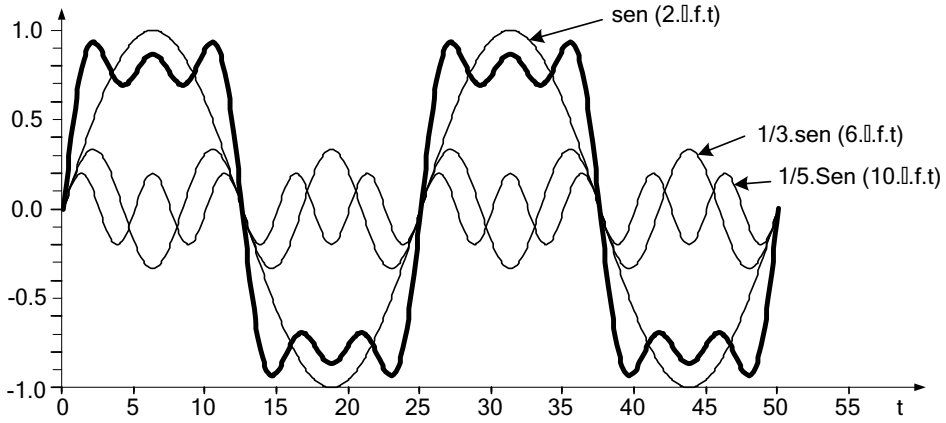


FIGURA 1.4. Composición de una señal cuadrada a partir de sus tres primeros armónicos.

1.4.1.2. Propagación de las señales en el medio

Cualquier medio físico conocido alterará las señales que se propagan por él. En general, estas modificaciones producen cierta degradación de la señal. Si se están difundiendo señales analógicas, éstas se distorsionarán perdiendo calidad. Si son señales digitales, se pueden producir alteraciones en la información que se transmite. Para determinar la capacidad de transmisión de un medio es necesario conocer como altera las señales y como se podrán reconstruir una vez hayan llegado al receptor. Las alteraciones más importantes que se producen en la señal son **atenuación, retardo y ruido**.

1.4.1.2.1. Atenuación

Por atenuación se entiende la disminución de la amplitud de la señal. Esta atenuación es función de la distancia que recorre la señal por el medio. La atenuación, en los medios guiados, tiene un comportamiento logarítmico y se expresa en decibelios por unidad de longitud. En la atmósfera y los medios no guiados en general, la atenuación es función de la distancia y de las condiciones atmosféricas (presión, temperatura, humedad). Esta circunstancia hace que las señales se vayan perdiendo con la distancia recorrida y por tanto, esto limita la distancia entre el emisor y el receptor, si no se emplean repetidores.

$$\text{Atenuación} = -20.\log_{10} (A_{\text{salida}}/A_{\text{entrada}})/L \text{ [dB/m]} \qquad (1.4)$$

Dado que todo medio introduce cierto ruido sobre la señal (como se verá a continuación), la amplitud de la señal debe conservar un nivel lo suficientemente destacado sobre el ruido para que los dispositivos electrónicos o los repetidores puedan reconocerla y regenerarla. El uso de repetidores cada cierta longitud de cable, garantizan la integridad de la señal cuando la información que se transmite es digital. Sin embargo, cuando la información transmitida es analógica (por ejemplo, una conversación telefónica), los repetidores no solo amplifican la señal sino que además hacen lo mismo con el ruido superpuesto, limitando la capacidad para regenerar la señal.

Si la atenuación únicamente dependiera de la distancia recorrida los repetidores paliarían casi completamente el problema. Sin embargo, la atenuación es también función de la frecuencia de la señal que transmite, haciéndose más importante cuando aumenta la frecuencia de la señal transmitida: a más frecuencia, más atenuación. Como las señales que transportan la información no son sinusoides puras, la propagación produce distinto efecto sobre los múltiples armónicos de la señal y mucho más acentuado para los armónicos altos. Este es verdaderamente uno de los factores que limita el ancho de banda de la señal que se puede transmitir el medio.

Si se analiza el problema de la atenuación creciente con la frecuencia sobre señales analógicas, se concluye que la señal que llega el otro extremo del medio estará distorsionada, reduciéndose la inteligibilidad de la información que transporta. Para aminorar este problema, existen técnicas que permiten ecualizar la atenuación en una banda de frecuencias. Otra alternativa es la utilización de amplificadores no lineales que presenten mayores coeficientes de amplificación para las frecuencias más altas.

1.4.1.2.2. Retardo

El retardo de la señal por sí solo no es un problema importante, sin embargo, los medios guiados presentan diferente velocidad de propagación para distintas frecuencias, lo que se conoce como **distorsión de retardo de propagación**. Al estudiar el retardo que produce un medio para distintas frecuencias, se aprecia que la velocidad es mayor en la frecuencia central del ancho de banda que presenta el medio y disminuye al acercarse a los extremos. Esto hace que las diferentes componentes armónicas de la señal lleguen al receptor en distintos instantes de tiempo, distorsionando la señal. El efecto de esto es que, si se está transmitiendo una secuencia de bits usando una señal digital (sucedería lo mismo con una señal analógica), debido a la distorsión de retardo, algunas de las componentes de un bit cualquiera se desplazarán hacia otras posiciones,

solapándose con las componentes de los bits adyacentes. Esto produce una distorsión que aumenta con la frecuencia de las señales que se transmiten llegando a hacerlas irreconocibles y, por tanto, limita la frecuencia máxima de transmisión. Hay técnicas de ecualización que pueden corregir, en buena medida, la distorsión de este retardo.

El retardo de propagación, junto con la atenuación, son las características del medio que limitan la capacidad de transmisión por el bucle de abonado de la línea telefónica, haciendo que un módem no supere los 56 kbps. La técnica que utilizan los sistemas xDSL para aprovechar mejor el ancho de banda es utilizar muchas señales con anchos de banda muy pequeños, con frecuencias relativamente elevadas. Con estas características de señal las atenuaciones son importantes pero las distorsiones, tanto por retardo como por atenuación, son pequeñas, alcanzando velocidades de transmisión sorprendentemente elevadas para el medio que se esta utilizando.

1.4.1.2.3. Ruido

De los tres factores que afectan a la comunicación en el medio, el ruido es el factor que más influye para limitar las prestaciones de un sistema de comunicación. Siempre que se propaga una señal por un medio, la señal recibida en el otro extremo estará alterada, por las distorsiones introducidas por la atenuación y el retardo, y por una serie de ruidos que se solapan a la señal. El ruido tiene diferentes orígenes y en función de éste se puede clasificar en: ruido térmico, diafonía, ruido de intermodulación y ruido impulsivo.

El **ruido térmico** está presente en el medio y es debido a la agitación térmica de los electrones que componen el medio. Como la agitación de los electrones aumenta con la temperatura el ruido térmico será también función de ésta y se produce en todos los dispositivos electrónicos y medios de transmisión. El ruido térmico es lo que se denomina ruido blanco y tiene una distribución uniforme en el espectro de frecuencias. El ruido térmico esta presente siempre y por tanto limita las prestaciones de los medios de transmisión de datos. El nivel de ruido debe ser inferior al de la señal y esta se va atenuando con la distancia recorrida mientras que el ruido permanece constante.

La **diafonía** es un acoplamiento no deseado entre las líneas que transportan dos señales distintas. La perturbación que produce esta interferencia es del mismo orden de magnitud que el ruido térmico. Es el típico cruce de líneas telefónico en el que se escucha otra conversación solapada. Esto ocurre cuando se acoplan eléctricamente dos pares de cables cercanos, aunque en ocasiones

también se produce en líneas de cable coaxial con varias canales multiplexados o en antenas de microondas.

Otro efecto que se produce en la transmisión es el denominado **ruido de intermodulación** que se produce entre señales de distintas frecuencias que comparten el medio de transmisión y por deficiencias del sistema se generan otras señales de frecuencias suma o diferencia de las dos frecuencias originales, o múltiplos de éstas. Si se transmiten dos señales f_1 y f_2 se pueden producir señales a frecuencias $f_2 – f_1$ y $f_1 + f_2$ que interferirán sobre otras señales con esas frecuencias. Estas componentes pueden aparecer debido al mal funcionamiento o por saturación de los sistemas cuando se utilizan niveles inadecuados de señal, lo que produce alinealidades en el transmisor o en el receptor.

Los tres fenómenos anteriores son razonablemente predecibles y se caracterizan por presentar magnitud constante, haciendo posible diseñar dispositivos electrónicos que minoren los problemas que producen.

El **ruido impulsivo** son perturbaciones constituidas por impulsos o picos irregulares de corta duración y de amplitud relativamente grande. Se generan por descargas atmosféricas o por perturbaciones electromagnéticas producidas por fallos o defectos en los sistemas de comunicación o de alimentación de los equipos. Este tipo de perturbaciones es muy irregular y difícilmente predecible por lo que es complicado abordarlo. En cualquier caso, el ruido impulsivo no tiene efectos muy negativos cuando se trata de transmisiones analógicas. Una conversación telefónica se puede perturbar con chasquidos cortos sin que afecte a la comunicación. Sin embargo, el ruido impulsivo es una de las fuentes principales de error en la comunicación digital de datos. Lo que para la comunicación analógica representa un pequeño ruido, por ejemplo un impulso de 0,01 s de duración, en una transmisión de 2 Mbps afectaría a 20.000 bits.

1.4.1.3. Capacidad de transmisión del medio

Los factores que intervienen para limitar la capacidad de transmisión de un medio, ya están básicamente expuestos y son: La **velocidad de transmisión** de los datos, que se expresa en bits por segundo (bps o baudios) y que está íntimamente relacionada con el **ancho de banda** del medio y del transmisor, y el **ruido** que está directamente relacionado con la **tasa de errores** que se producen el la transmisión.

Considerando un medio exento de ruido y una comunicación digital binaria, Nyquist estableció que la limitación en la velocidad de los datos está impuesta

simplemente por el ancho de banda del canal W. En esta situación la velocidad mayor de transmisión de la señal que se puede conseguir es 2.W. Si se considera la línea telefónica con un ancho de banda de 3400 Hz para transmitir señales digitales, entonces la capacidad del canal es de 6800 bps. Sin embargo, se pueden utilizar señales codificadas con más de dos niveles para transmitir la información. Nyquist planteó la ecuación 1.5 para calcular la capacidad del canal con codificación multinivel.

$$C = 2.W.\log_2 M \tag{1.5}$$

siendo M el número de codificaciones utilizadas en la comunicación. El inconveniente de incrementar M es que se reduce la "distancia" entre niveles pudiendo llegar a ser del mismo orden que el ruido que presenta el canal. Momento en el cual la información se corrompe. Este factor no está considerado en la ecuación 1.5.

El matemático Claude Shannon estudió y cuantificó la capacidad de transmisión de un determinado medio en función del ancho de banda y de la relación señal/ruido que presenta el canal. La ecuación que planteó Shannon establece que la capacidad máxima del canal, en bits por segundo:

$$C = W.\log_2 (1+S/N) \tag{1.6}$$

S/N es la relación entre la potencia de la señal a transmitir y la potencia del ruido del canal, y está expresado en decibelios. Esta ecuación representa el límite máximo teórico que se puede conseguir, teniendo en cuenta que existe un ruido blanco asociado al medio. Sin embargo, en la práctica, también se produce ruido impulsivo, atenuación y retardo, por lo que las velocidades de transmisión que se consiguen son muy inferiores a las obtenidas con la ecuación 1.6.

1.4.2. Medios de transmisión

Los medios de transmisión deben cubrir todas las necesidades que plantean las distintas necesidades de comunicación. Para dejar patente la variedad de medios de transmisión que están en uso sólo hay que analizar algunos casos: la distribución de una red de telecomunicaciones en un núcleo urbano, la interconexión entre ciudades distintas o la problemática que se plantea cuando la comunicación se establece entre dispositivos portátiles en movimiento. Los medios de transmisión físicos son soporte en muchos casos de miles de circuitos y aunque los dispositivos nuevos que se instalan son completamente digitales aún conviven con grandes redes de comunicaciones analógicas.

Para iniciar el estudio de los medios de transmisión se clasifican éstos en dos grandes grupos: medios guiados y no guiados.

1.4.2.1. Medios guiados

Los medios guiados son los que confinan la señal en el medio y guían las ondas electromagnéticas a lo largo de él. Hay tres grandes grupos dentro de este tipo de medios: pares trenzados, cables coaxiales y cables de fibra óptica. A continuación se describen las características esenciales de estos medios, empezando por los más antiguos, hasta los enlaces de fibra óptica, aunque todos ellos se estudiarán con mayor profundidad en el capítulo tercero.

Los **portadores de hilo desnudo** están completamente desaparecidos. Las redes de transmisión telefónica a principios del siglo XX estaban constituidas por conductores de cobre desnudo que proporcionaban circuitos de baja frecuencia. La disposición del hilo se efectuaba inicialmente sobre unos aisladores situados en los laterales de los postes de madera o sobre crucetas. Admitía frecuencias del orden de los 300 kHz, pudiendo proporcionar hasta 28 circuitos sobre dos hilos, sin embargo presentaba importantes problemas de transmisión, unas veces por las roturas ocasionales del portador y otras por la baja calidad que ofrecían en situaciones meteorológicas adversas.

Los **cables de pares trenzados** están constituidos por un conjunto de conductores metálicos cilíndricos, aislados entre sí y protegidos del exterior por una cubierta común. Los hay apantallados (STP) y sin apantallar (UTP). Se denominan de pares simétricos no por su disposición geométrica en el interior del cable cilíndrico, sino por el comportamiento eléctrico de los conductores con respecto a la pantalla metálica de la cubierta, generalmente conectada a tierra. La disposición interior de los pares permite clasificarlos como cable de cuadretes estrella y como cuadretes DM en función de cómo estén trenzados los conductores entre sí (un cuadrete está constituido por cuatro conductores trenzados). Este tipo de cables proporciona tres circuitos en frecuencia vocal sobre cada cuadrete.

Actualmente, los cables de pares que se usan para telefonía en la red de conexión de los domicilios de los abonados con la central. También se utiliza como soporte para enlaces de redes de área local. El cable UTP es el más utilizado para el cableado de edificios ya que es el más barato y más fácil de instalar, como contra partida también es el más sensible a interferencias electromagnéticas, presentando problemas de diafonía.

Un **cable coaxial** está constituido por uno o más tubos coaxiales, estando cada tubo formado por dos conductores de cobre: uno exterior en forma de cilindro hueco, y dentro de él y en su eje un conductor interior macizo, separados ambos por un dieléctrico.

Este tipo de conductor se utiliza para distribución de señales de vídeo, en redes de área local y en telefonía de larga distancia. El cable coaxial es mucho más inmune a interferencias que el par trenzado, permitiendo cubrir mayores distancias y transportar señales con mayor ancho de banda.

Los tubos coaxiales, que se instalaban en las grandes líneas de comunicación telefónica, pueden transportar hasta 10.800 circuitos de tipo analógico con frecuencia máxima en línea de 60 MHz y secciones de repetición cada 1,5 km. Estos cables coaxiales posibilitan la transmisión de sistemas digitales de 140 y 565 Mbps con capacidades de 1920 y 7680 circuitos de 64 kbps. Los cables coaxiales se dejaron de instalar para telefonía en el año 1985 a favor de los cables de fibra óptica aunque siguen siendo muy utilizados en otras aplicaciones.

Los **cables de fibra óptica** se componen de un hilo flexible de óxido de silicio (vidrio), recubiertos con otra capa de vidrio con un índice de refracción menor, y protegido por una cubierta opaca que absorbe la luz. Estos cables conducen con facilidad por su interior un haz luminoso que puede utilizarse como soporte para la información. Para ello el equipo transmisor convertirá la señal eléctrica en haces luminosos, realizándose el proceso inverso en el equipo receptor.

Puede producir distorsión en la señal propagada, sobre todo en largas distancias, debido a que los armónicos del haz siguen caminos distintos en el interior de la fibra y por tanto llegan al extremo final en tiempos diferentes. Para reducir este inconveniente se fabrican fibras cuyo núcleo tiene un índice de refracción variable, que compensa la diferencia de recorrido de los rayos luminosos, haciendo que la llegada de todos los rayos se produzca al mismo tiempo, a este tipo de fibras se las denomina fibras multimodo.

Además de las fibras multimodo de alto índice y de índice gradual, las redes de telecomunicación utilizan fibras monomodo que consiguen que el rayo luminoso se propague por un único camino, teniendo este tipo de fibras un núcleo de un diámetro reducido (10 micras). Las fibras monomodo se utilizan en enlaces interurbanos con elementos de repetición. Las fibras ópticas multimodo, con un diámetro de 50 micras, se utilizan en distancias cortas en las que no se necesite regenerar la señal y en enlaces entre instalaciones urbanas.

Los cables de fibra óptica están desplazando al resto de portadores, se utilizan para enlaces urbanos e interurbanos, como cable submarino y como cable de distribución para acceso hacia el domicilio de los abonados. La constitución y número de portadores del cable dependerá de la aplicación y tipo de enlace que se realice. Los cables de fibra que más habitualmente se utilizan llevan entre 4 y 128 fibras y los tipos de protecciones externas e internas dependerán de por donde discurra la instalación. Es necesario recordar el enorme ancho de banda y el poder de transporte de información que ofrecen los cables de fibra óptica ya que la luz transmitida corresponde a una frecuencia de cien billones de hercios.

Los emisores de luz que se utilizan normalmente son diodos fotoemisores (LED) y emisores láser, éste último con una fuente de radiación coherente en la que todos los fotones se encuentran en fase y tienen una misma longitud de onda, siendo los diodos fotoemisores más económicos y de mayor duración aunque con menor potencia. Para la recepción se utilizan detectores ópticos que transforman la señal luminosa en señales eléctricas convencionales.

La atenuación de los conductores de fibra óptica ha ido reduciéndose gracias a los avances en la calidad de las fibras y a la mejora de los métodos de empalme. Actualmente se sitúan los equipos repetidores sobre la fibra cada 30 kilómetros ya que se han conseguido atenuaciones en torno a los 0,2 dB/km. El desarrollo de nuevas fibras ópticas apunta a que estas atenuaciones sean del orden de 0,01 y 0,005 dB/km.

Los cables de fibra óptica tienen innumerables ventajas frente a otros medios de transmisión como son los cables coaxiales, permiten anchos de banda muy elevados, con diámetros y pesos notablemente inferiores (dos órdenes de magnitud), lo que redunda en que haya mayor longitud de cable en las bobinas, además son inmunes frente a interferencias electromagnéticas, no presentan problemas de diafonía, aumentan las secciones de repetición. Todo esto conlleva una mayor seguridad, y por lo tanto mayor calidad.

1.4.2.2. Medios no guiados

Radioenlaces. La radio se utiliza como medio de transmisión en las redes de telecomunicación tanto de telefonía como de datos, y no sólo en la estructura de la red para enlazar centros de conmutación sino que ofrecen directamente servicio al abonado.

En este caso el medio físico de transmisión de la señal entre dos puntos (antenas) es la atmósfera. La propagación de las ondas de radio se puede producir por reflexión de las ondas en las capas altas de la atmósfera, entre 10 y 400 kilómetros, como sucede en el caso de las emisiones de onda corta, o directamente. En este segundo caso o bien las señales siguen la curvatura de la Tierra, como sucede en emisiones de onda media o bien necesitan la visibilidad directa entre las antenas para su detección, este es el caso de la televisión o las emisiones de radio en frecuencia modulada.

En la transmisión de servicios de telefonía móvil, de comunicaciones marítimas y los básicos de telefonía y datos se utilizan las comunicaciones radioeléctricas. Las empresas operadores tienen radioenlaces como medio de transmisión. Muchas ciudades están comunicadas por radioenlaces que actúan de sistemas alternativos para conexiones con cables coaxiales o de fibra óptica.

Los radioenlaces analógicos, ya obsoletos, permiten mantener entre 300 y 2700 comunicaciones simultáneas con frecuencias de línea entre 2 y 8 GHz para el transporté de señales de voz, datos, televisión, etc. Los radioenlaces digitales permiten comunicaciones de hasta 140 Mbps que pueden incorporar ocho portadoras con 1920 circuitos cada una, normalmente siete operativas y una en reserva.

Un **satélite** de comunicaciones es un equipo repetidor de un enlace terrestre de microondas, que se encuentra situado a 36.000 kilómetros de altura.

Las señales que se transmiten al satélite se trasladan a una banda de frecuencias de 6 a 14 GHz. El satélite regenera la señal y la reenvía en una banda de frecuencias de 4 a 12 GHz. Los paneles solares del satélite suministran la potencia necesaria, y en periodos de eclipse, utilizan baterías recargadas durante el resto del tiempo. A través de enlaces microondas, fibra óptica o cables coaxiales se incorpora la señal recibida en la estación de seguimiento a la red terrestre. De este modo las estaciones de seguimiento de satélites acceden y comparten los canales del satélite, que se definen por el ancho de banda y la potencia de la antena.

1.5. ESTRUCTURAS BÁSICAS EN LA COMUNICACIÓN

En este epígrafe se estudian las configuraciones básicas de los sistemas de comunicaciones desde el punto de vista del tipo de transmisión de datos y la estructura de interconexión de la red.

La transmisión de datos se puede realizar por un solo par de conductores, una vía, o por varias simultáneamente dando lugar a dos tipos básicos de transmisión: paralelo o serie.

1.5.1. Transmisión de datos en paralelo

Los ordenadores en sus buses de comunicaciones internos y en algunos casos para la comunicación con algunos dispositivos periféricos (como la impresora), utilizan transmisión de datos en paralelo. Los buses de datos en paralelo transmiten simultáneamente 8, 16 o 32 bits. En este tipo de comunicación cada bit de datos y cada señal de control dispone de una línea dedicada del bus. Para transmitir una palabra completa de 8 bits se necesitan 8 líneas de datos. Junto con estas líneas de datos, también son necesarias líneas de control de flujo. La transmisión de datos en paralelo permite alcanzar altas velocidades en la transferencia de datos, pero su cableado e interfaces resultan más costosos que los de una transmisión serie. La contrapartida es que este tipo de conexiones es muy vulnerable a las interferencias electromagnéticas por lo que se utiliza en distancias muy cortas de transmisión de datos.

1.5.2. Transmisión de datos en serie

La transmisión de datos se realiza bit a bit, secuencialmente, por la misma línea, junto con los bits de control de la transmisión, utilizando sólo dos conductores. Dado que los bloques de información se transmiten de manera secuencial y no simultánea, la velocidad de transferencia de los datos es mucho menor que en el caso de la transmisión de datos en paralelo, para la misma tasa de transferencia de bits. Sin embargo, este tipo de transmisión resulta más barata.

Las redes de comunicación utilizan enlaces serie, con un único soporte para establecer la comunicación. En ordenadores personales también se están imponiendo las comunicaciones serie (USB e IEEE1394) frente a las paralelo y únicamente para conexiones próximas de los sistemas microprocesadores se utiliza la comunicación paralelo. Así, entornos que siempre han utilizado comunicaciones paralelo, como el acceso a memoria de un procesador, en muchos casos también utilizan la comunicación serie, existiendo estándares de comunicación para dispositivos microelectrónicos, como el bus I^2C.

A continuación se estudia la problemática de enviar una secuencia de caracteres a través de un enlace en serie, la **sincronización** del transmisor con el

receptor. El receptor necesita identificar el comienzo y el final de la secuencia de datos que compone cada carácter que se transmite. Para realizar esto se utilizan dos procedimientos denominados transmisión serie asíncrona y transmisión serie síncrona.

En la **transmisión asíncrona** el ordenador transmisor y el receptor tienen señales de temporización independientes. El tiempo que transcurre entre un carácter y el siguiente es arbitrario. Para identificar el inicio de la transmisión de un carácter, este está precedido por un bit de inicio, y en algunos casos también de un bit de paro para señalar el final del carácter. Este método tiene la desventaja de que requiere la transmisión de bits adicionales que acompañen a cada uno de los caracteres, siendo estas cadenas de datos lo suficientemente cortas para garantizar la sincronización, lo cual disminuye la eficiencia de la comunicación.

En la **transmisión síncrona** no se necesitan bits de inicio ni de paro, puesto que el transmisor y el receptor tienen una señal de temporización común. Como contrapartida, es necesaria una línea de transmisión adicional que sincroniza el dispositivo transmisor con el receptor, o una codificación especial en las señales transmitidas que incorpore las marcas de sincronización. Este tipo de señalización ocupa parte del ancho de banda del canal disponible.

1.5.2.1. Modos de comunicación punto a punto

Según el sentido de la información, la transmisión de datos puede ser de tres tipos (Fig. 1.5).

1.5.2.1.1. Modo simplex

La transmisión sólo es posible en una dirección, desde la estación transmisora a la estación receptora y ésta no puede transmitir en la otra dirección. Este método sólo se usa para comunicar dispositivos como, sensores, dispositivos de medida o periféricos de entrada o salida que sólo transmiten información en una dirección.

1.5.2.1.2. Modo half duplex

Los datos se transmiten en ambas direcciones, pero en distintos instantes de tiempo. Las estaciones de ambos extremos del enlace deben cambiar de estar en

modo transmisión a modo recepción alternativamente. Así, la estación A puede transmitir a la estación B y viceversa, pero en distintos instantes de tiempo. Un *walkie-talkie* es un ejemplo de transmisión en modo *half duplex* (o alternado). Cada uno de los dispositivos puede transmitir o recibir, pero no a la vez.

1.5.2.1.3. Modo full duplex

Los datos se transmiten en forma simultánea en ambas direcciones entre las estaciones A y B. El sistema telefónico es un ejemplo de modo *full duplex* (también denominado bidireccional o en muchos textos, simplemente *duplex*), ya que una persona puede hablar y escuchar al mismo tiempo.

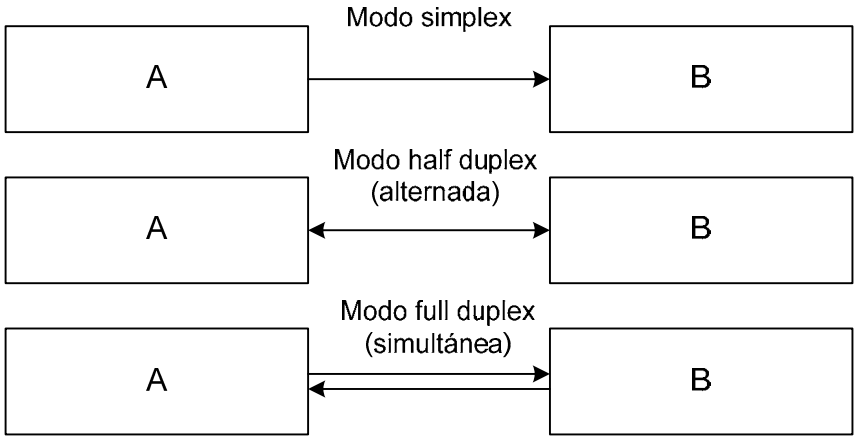


FIGURA 1.5. Modos de comunicación punto a punto.

1.5.2.2. Topologías de redes multipunto

Las redes locales utilizadas en oficinas o en entornos industriales, son ejemplos de redes multipunto en las que cualquier dispositivo puede comunicarse con cualquier otro de su entorno. En la representación de estas configuraciones (Fig. 1.6), se muestran las **líneas de comunicación** y los **nodos**, que son los puntos de la red donde confluyen una o varias líneas de comunicación y también los puntos de conexión de los ordenadores.

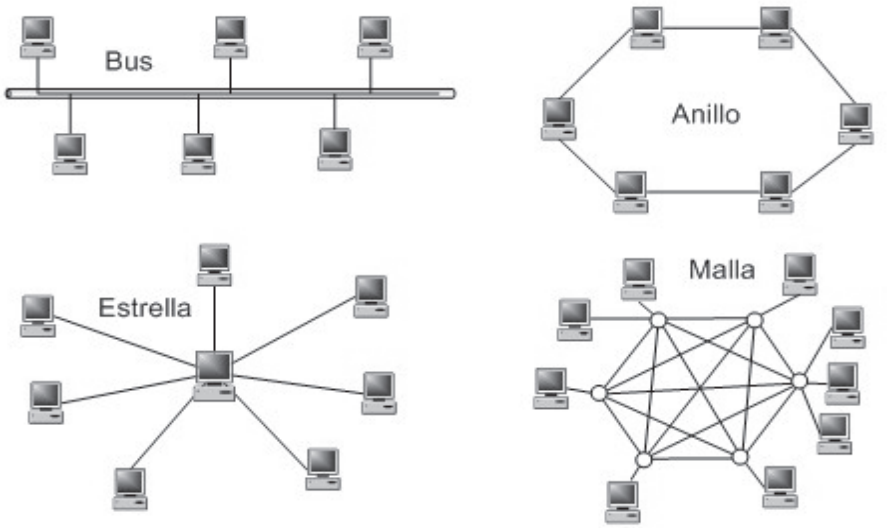


FIGURA 1.6. Topologías de redes multipunto.

1.5.2.2.1. Bus de datos (canal de distribución)

Consta de un bus lineal con el cual se conectan todas las estaciones, de forma que todas ellas comparten el mismo medio físico de difusión. Con frecuencia, este sistema se utiliza en agrupamientos de terminales multipunto. Su ventaja radica es que su instalación es sencilla y barata, constituyendo una red pasiva, en la que todos los elementos activos están en las estaciones. El fallo de una estación no afecta a la red pero un fallo en el bus la paraliza completamente.

1.5.2.2.2. Estrella

Esta configuración tiene canales bidireccionales asignados entre cada estación y un anillo conmutador central, a través del cual deben pasar todas las comunicaciones. Este tipo de red se utiliza en los sistemas telefónicos de muchas empresas (centrales privadas, PBXs), en los cuales todas las líneas pasan a través de una central telefónica. Este sistema con frecuencia también se utiliza para conectar terminales remotas y locales con una computadora principal central aportando gran flexibilidad en la gestión y el control de la conexión. Las desventajas de este sistema es que si hay un fallo en el nodo central, toda la red queda bloqueada y las longitudes de cableado son elevadas.

1.5.2.2.3. De jerarquía o de árbol

Esta configuración consiste en una serie de derivaciones que en general convergen en un punto. En este sistema sólo hay una ruta de transmisión entre dos estaciones. La configuración se obtiene con varias redes en bus vinculadas entre sí mediante repetidores.

1.5.2.2.4. Anillo

Todas las estaciones de la red de área local se conectan entre sí formado un lazo cerrado. Las distancias entre los nodos por lo general son menores a 100 m. Los datos que se introducen en el sistema de anillo circulan a su alrededor hasta que algún sistema los retira. Todas las estaciones tienen acceso a los datos y se puede realizar un reparto equilibrado de la capacidad de transmisión, proporcionando un tiempo de respuesta limitado y buena gestión de las averías.

Las desventajas son que al estar todos lo dispositivos involucrados en la comunicación, el fallo de cualquiera de ellos paraliza la red. Esta topología requiere mecanismos de control que permitan reconocer la información defectuosa que circula por la red. En algunos casos se dedica una estación a las tareas de supervisión y control de la red.

1.5.2.2.5. Malla

En este método no existe una configuración formal para las conexiones entre estaciones, y entre éstas puede haber varias rutas de datos.

1.5.2.2.6. Bucle

Es una mezcla entre la configuración en estrella y en anillo. Se necesita una estación central para controlar las demás estaciones, que están interconectadas formando un lazo cerrado sobre la estación controladora. Presenta los inconvenientes de ambas configuraciones en cuanto al bloque de la red y presenta las ventajas de economía de cableado y la facilidad de ampliación de estaciones. Se utiliza en aplicaciones de bajo coste y velocidad. Requiere un controlador con pocas conexiones externas.

1.6. INTRODUCCIÓN A LA CODIFICACIÓN DE DATOS

En los epígrafes anteriores se han diferenciado los datos y las señales analógicas de las digitales. En este apartado se explica como se pueden codificar información analógica o digital utilizando señales analógicas o digitales, indistintamente. Se presentan cuatro situaciones diferentes: se dispone de una línea analógica por la que se quiere transmitir información analógica, se dispone de información digital para ser transmitida por esa misma línea analógica o se dispone de una línea digital por la que se puede transmitir información analógica o información digital.

1.6.1. Codificación en las comunicaciones analógicas

Aun cuando se pudiera pensar que las nuevas infraestructuras de red han dejado en el pasado la transmisión de datos analógica, esto no es así. Algunos de los medios físicos de transmisión presentados en el epígrafe 1.4, sólo permiten la propagación de señales analógicas. Este es el caso de la fibra óptica y, en general, de todos los medios no guiados.

1.6.1.1. Información analógica

Las señales analógicas, una vez transformadas en señales eléctricas o electromagnéticas, se pueden transmitir por canales analógicos fácilmente sin codificación alguna, y sin alterar las frecuencias naturales de la señal, esto es lo que se denomina difusión en **banda base**. Este proceso se realiza para transmitir las señales eléctricas que genera un teléfono convencional por el par trenzado que llega a la centralita telefónica local. Los equipos de telefonía agruparán las comunicaciones para transferirlas por un único cable a otra centralita remota que distribuya cada una de ellas a sus destinatarios finales, de nuevo en banda base. El proceso de modulación permite así el agrupar varios canales de comunicación (multiplexación) por un cable. A la transmisión que se realiza simultáneamente con distintas señales moduladas a distintas frecuencias por un cable, normalmente coaxial, se le denomina transmisión en **banda ancha**.

El proceso de **modulación** (Figuras 1.7 y 1.8), consiste en desplazar el ancho de banda de la señal en banda base hacia otra zona del espectro de frecuencia. Si se realiza este desplazamiento con varios canales a distintas zonas del espectro, estas comunicaciones podrán compartir el mismo medio de transmisión, sin interferirse.

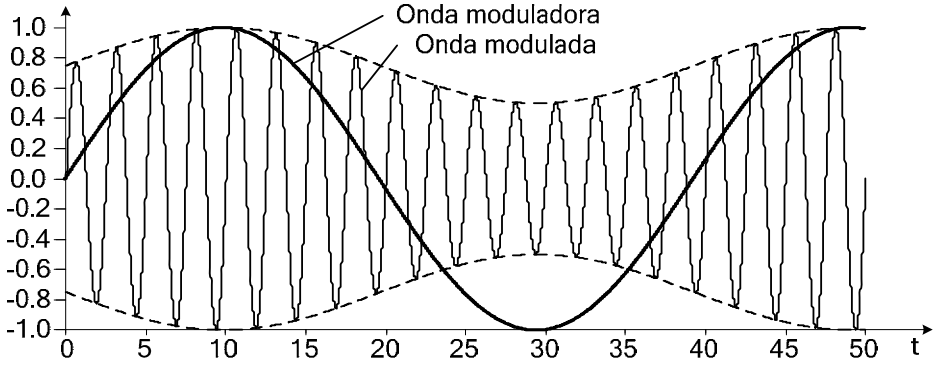


FIGURA 1.7. Modulación en amplitud.

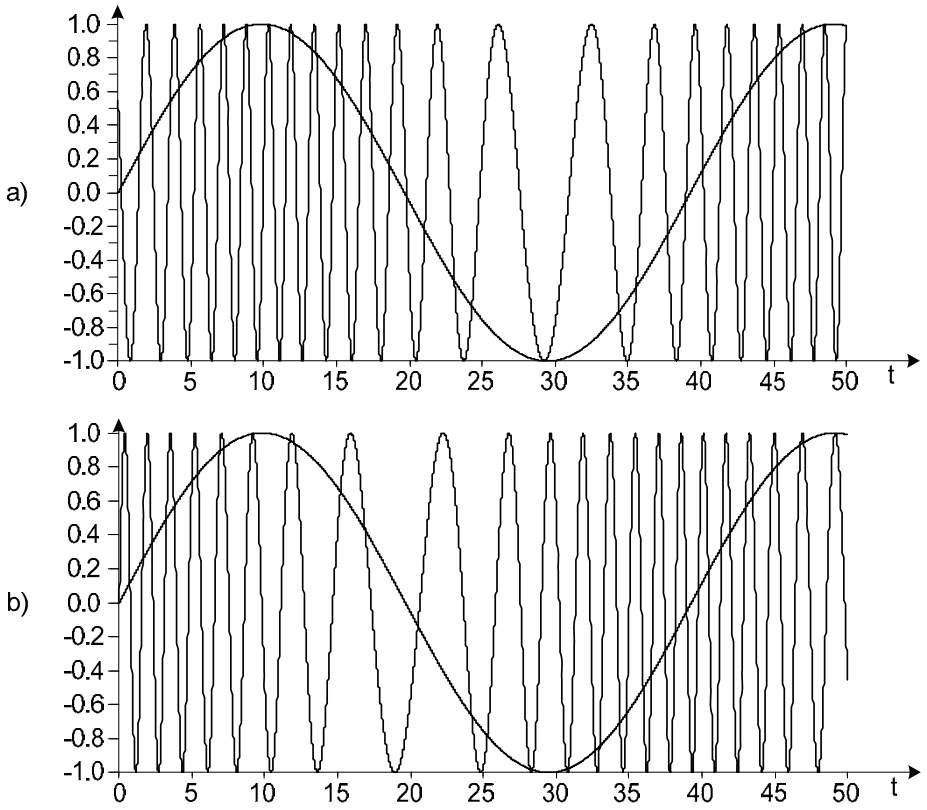


FIGURA 1.8. Modulación en a) frecuencia y b) fase.

Este proceso utiliza una **onda portadora** distinta para cada uno de los canales a modular. Siendo esta banda portadora una frecuencia sinusoidal pura, de amplitud, frecuencia y fase constante. Se entiende por modulación al proceso de variar la amplitud (amplitud modulada, como puede verse en la Figura 1.7), frecuencia (frecuencia modulada, como se ve igualmente en la Figura 1.8) o fase de una onda portadora en función de las variaciones que presenta la señal que contiene la información. Esta señal es la **onda moduladora**. En la transmisión telefónica el procedimiento más utilizado es la modulación de amplitud, por su economía y por la sencillez técnica de su proceso. En contrapartida esta modulación es la más vulnerable a las perturbaciones que se producen en el medio, que afectan fundamentalmente a la amplitud de la señal transportada.

1.6.1.2. Información digital

Esta es la que se presenta cuando es necesario conectar un ordenador personal a la red telefónica conmutada. Como ya se ha comentado, esta función la realiza un dispositivo denominado módem, término que procede de las palabras (modulación y demodulación). Dado que la transmisión de datos se debe producir sobre una señal analógica, para la modulación se utiliza una onda portadora, sinusoidal pura. Sobre esta onda se modifican algunos de sus parámetros básicos (amplitud, frecuencia o fase) para codificar la información en ella. La codificación sobre cada uno de estos parámetros da una técnica de codificación: ASK, FSK y PSK respectivamente.

1.6.1.2.1. Codificación en amplitud (ASK)

La información digital está codificada en amplitud (ASK, *Amplitude Shift Keying*), en dos niveles de tensión diferentes: un nivel bajo o cero lógico y un nivel alto o uno lógico. Por tanto, será necesario representar estas dos situaciones codificadas sobre la onda portadora de frecuencia f_p. Una posible representación codificada en amplitud será la mostrada en la Tabla 1.1 y en la Figura 1.9.

TABLA 1.1. Codificación en amplitud

Dato digital	Señal analógica
1	$A.\text{sen}\ (2.\pi.f_p.t)$
0	0

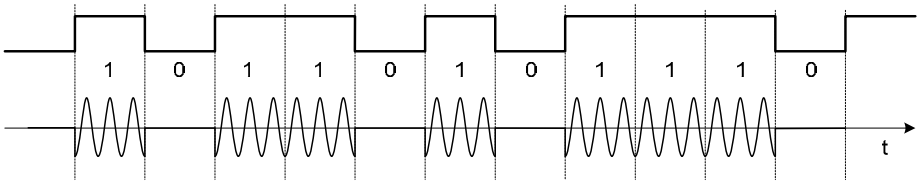


FIGURA 1.9. Codificación en amplitud.

1.6.1.2.2. Codificación en frecuencia (FSK)

La codificación en frecuencia (FSK, *Frequency Shift Keying*) se realiza eligiendo una onda portadora con frecuencia f_p, se seleccionaran dos desplazamientos (+d y –d) de frecuencia simétricos sobre la portadora para representar los dos códigos binarios (Tabla 1.2 y Figura 1.10).

TABLA 1.2. Codificación en frecuencia

Dato digital	Señal analógica
0	A.sen [2.π.(f_p-d)t]
1	A.sen [2.π.(f_p+d)t]

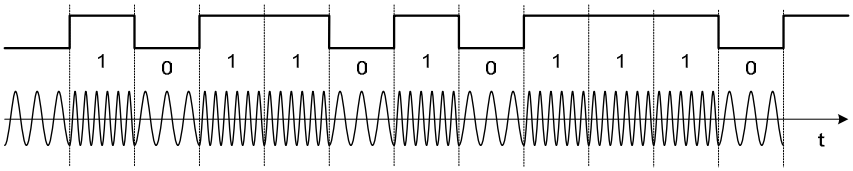


FIGURA 1.10. Modulación en frecuencia.

1.6.1.2.3. Codificación en fase (PSK)

Para la codificación en fase (PSK, *Phase Shift Keying*) los dos valores binarios utilizan un desfase de 180 grados (π radianes) entre las señales analógicas que las representan (Tabla 1.3 y Figura 1.11).

Una variante típica de la codificación en fase es la codificación en cuadratura QPSK (*Quadrature Phase Shift Keying*), que utiliza codificación multinivel para representar más de un dígito binario con cada código. En este

caso se utilizan desplazamientos de la fase de 90º ($\pi/2$) entre dos códigos consecutivos, como se muestra en la Tabla 1.4 y que está representada en la Figura 1.12 en un diagrama polar.

TABLA 1.3. Codificación en fase

Dato digital	Señal analógica
0	A.sen ($2.\pi.f_p.t$)
1	A.sen ($2.\pi.f_p.t + 180º$)

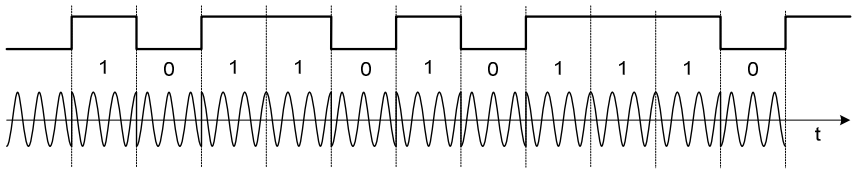


FIGURA 1.11. Modulación en fase.

TABLA 1.4. Codificación QPSK

Dato digital	Señal analógica
11	A.sen ($2.\pi.f_p.t + \pi/4$)
10	A.sen ($2.\pi.f_p.t + 3.\pi/4$)
00	A.sen ($2.\pi.f_p.t + 5.\pi/4$)
01	A.sen ($2.\pi.f_p.t + 7.\pi/4$)

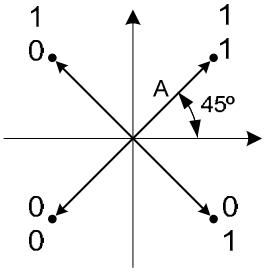


FIGURA 1.12. Representación de los códigos QPSK.

Esta técnica se puede ampliar y mezclada con la modulación en amplitud, posibilitando utilizaciones más eficientes del ancho de banda del canal (Fig. 1.13).

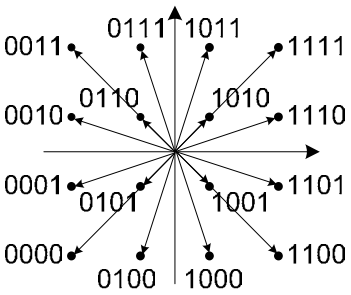


FIGURA 1.13. Codificación para cuatro bits, combina 3 amplitudes con 12 desfases.

1.6.2. Codificación en las comunicaciones digitales

El tipo de señales que se puede transmitir por un canal digital es únicamente información digital, codificada con dos o más niveles discretos de tensión. Cualquier información que se quiera transmitir por este medio debe ser convertida a este formato. El objetivo es el aprovechamiento óptimo del medio de transmisión y por tanto, a la hora de elegir un sistema de codificación, se ha de tener en cuenta el ancho de banda de la señal resultante y la tasa de error que esta conlleva por pérdida de sincronización y vulnerabilidad al ruido. También es muy importante tener capacidad para detectar los errores que se producen en el medio durante la transmisión y, por cuestiones de diseño hardware en las que no se va a entrar, evitar las componentes continuas en el espectro de la señal resultante.

1.6.2.1. Información digital

Las señales digitales que se transmiten por el medio físico utilizan las mismas frecuencias que la información digital que se pretende transmitir, por lo que estrictamente hablando, no se realiza ningún tipo de modulación. Por este motivo a este tipo de comunicación se la denomina modulación en banda base.

Los métodos digitales de codificación de los datos (Fig. 1.14), aportan a las señales transmitidas capacidades de sincronización, de detección de errores con

buenas características de inmunidad al ruido y, todo esto, sin incrementar innecesariamente el espectro de la señal transmitida.

Los formatos de codificación más habituales son: Polar sin retorno a cero (NRZ), Polar sin retorno a cero invertido (NRZI), Manchester, Manchester diferencial, AMI Bipolar (*Alternate Mark Inversion*) y Pseudoternaria, existiendo diversas variantes sobre cada uno de estos métodos.

1.6.2.1.1. Codificaciones polares

La codificación NRZ utiliza dos niveles distintos de tensión para representar los dos valores binarios 1 y 0. La variante NRZI también opera entre dos niveles distintos de tensión pero codifica un 1 lógico con una transición de la señal (desde el nivel de tensión en el que se encuentra, al otro) y un 0 lógico con una ausencia de transición. NRZI es un ejemplo de codificación diferencial. Este tipo de codificación es mucho más inmune al ruido que la codificación por nivel. Las codificaciones polares son la más fáciles de implementar pero presentan dos importantes inconvenientes: la ausencia de señales específicas de sincronización y la presencia de componentes continuas en el espectro de señal. Ambos problemas se agudizan cuando se presentan cadenas largas de ceros o de unos.

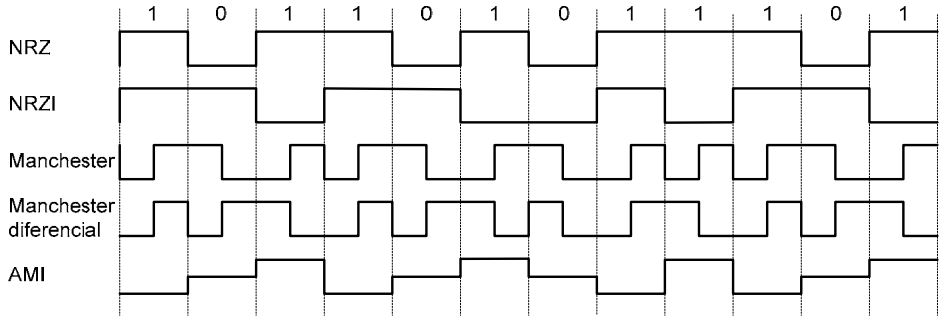


FIGURA 1.14. Codificación de señales digitales.

1.6.2.1.2. Codificaciones bifase

Dentro de este grupo de técnicas de codificación están el código Manchester y el código Manchester diferencial.

Este tipo de codificación supera las dos principales desventajas de los códigos NRZ, incorporando una transición a mitad de cada bit. El código Manchester representa un 1 lógico con una transición desde el nivel bajo al alto y un 0 lógico con una transición desde el nivel alto al bajo. El código Manchester diferencial presenta siempre una transición a mitad de bit con el propósito de mantener la sincronización. Codifica un cero con la presencia de una transición al principio del bit y un uno con la ausencia de dicha transición.

La principal desventaja de este tipo de código es que necesita el doble de ancho de banda para ser transmitido que los códigos NRZ. Sin embargo aporta las siguientes ventajas: es un código autosincronizado, no tiene componente de continua y proporciona la capacidad de detección de algunos errores (ante la ausencia de una transición esperada).

1.6.2.1.3. Códigos bipolares

Este tipo de códigos se caracteriza porque utilizan tres niveles de tensión distintos: cero y dos niveles de tensión simétricos, uno positivo y otro negativo. La codificación en el caso del código AMI bipolar es la siguiente, un cero se representa con el nivel de tensión cero y un uno se representa con un pulso positivo o negativo que se van alternando con la secuencia de unos. El código pseudoterciario codifica los ceros y unos al revés que el código AMI bipolar, en todo lo demás es equivalente al AMI.

Éste tipo de códigos presenta las siguientes ventajas: la señal transmitida no presenta componente continua, proporciona un método sencillo para detectar errores comprobando la alternancia de los pulsos y necesita un ancho de banda menor que el que necesita la codificación NRZ. Los problemas de sincronización de la señal se han solucionado cuando se presentan cadenas largas de unos, pero siguen existiendo cuando aparecen cadenas largas de ceros.

Para mejorar el problema con las cadenas de ceros en la codificación AMI bipolar se han desarrollado algunas variantes de este código. Por ejemplo, el código "*Bipolar with & zero substitution*" (B&ZS) propone cambiar las cadenas de ocho ceros por las dos siguientes secuencias: 000+-0-+ y 000-+0+-, alternándolas con cada nueva cadena de ocho ceros. Claramente estas secuencias incorporan una violación de código, lo que permite al receptor detectarlas y sustituirlas por los correspondientes ceros. Es interesante observar que para evitar las componentes continuas que introduce la violación de código sobre el espectro de frecuencias, se implementan dos codificaciones alternadas que compensan entre ellas esa posible incidencia.

1.6.2.2. Información analógica

En este caso el proceso consistirá en convertir las señales analógicas a transmitir en señales digitales en un proceso de "digitalización" de la señal, transformando una señal continua en una secuencia de valores que se obtiene muestreando la señal a intervalos regulares de tiempo y transformando esa secuencia de impulsos en códigos digitales. El proceso de comunicación se reduce a transmitir esos códigos con señales digitales, situación estudiada en el epígrafe anterior.

Se van a desarrollar a continuación dos técnicas diferentes para el proceso de digitalización de la señal: Modulación por codificación de impulsos (PCM, *Pulse Code Modulation*) y Modulación Delta (DM, *Delta Modulation*). Existen más técnicas de digitalización y algunas variantes sobre las dos anteriores. Antes de abordar las técnicas de digitalización se hace una revisión del teorema de muestreo.

1.6.2.2.1. Teorema de muestreo de Shannon

El primer paso en la codificación de la señal analógica, consiste en generar una señal discreta a partir de "muestras" que se recogen de la señal analógica original (Fig. 1.15). El proceso de obtención de muestras por simplicidad y eficiencia se realiza a intervalos regulares de tiempo.

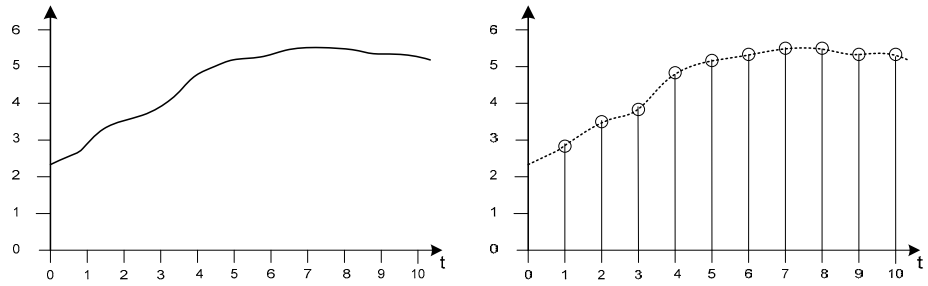


FIGURA 1.15. Muestreo.

Las muestras obtenidas deben contener toda la información importante de la señal analógica original. Es decir que la señal analógica y la secuencia de datos generada contienen básicamente la misma información. En las Figuras 1.16 y

1.17 se puede apreciar que hay pérdidas de información que pueden resultar transcendentes, entre la señal original y la señal que se infiere del muestreo.

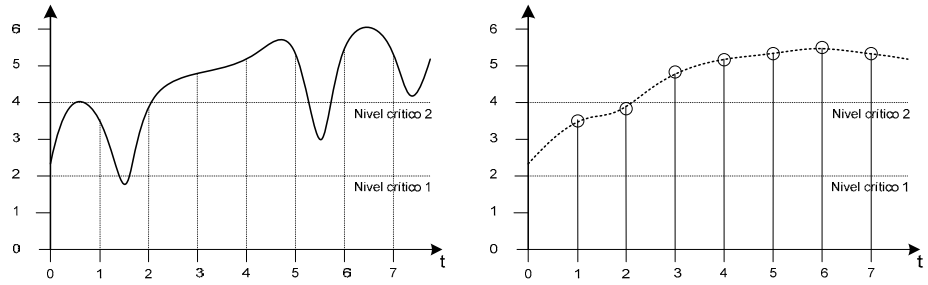


FIGURA 1.16. Posibles efectos del proceso de muestreo.

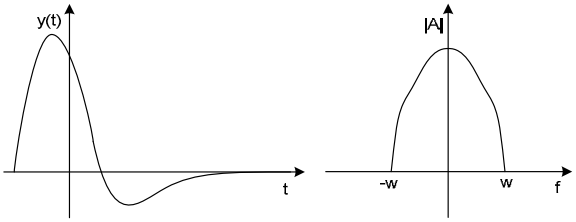


FIGURA 1.17. Espectro de una señal analógica. La máxima frecuencia que presenta la señal y(t) es w que se corresponde con el ancho de banda de la señal.

Si se analiza el problema visualizando los espectros de frecuencia de las señales implicadas, tanto de las analógicas originales como las que se pueden inferir de las muestras obtenidas. Se observa que cuando la frecuencia de obtención de muestras f_m es mayor del doble del ancho de banda W de la señal analógica muestreada (ecuación 1.7) no hay interferencias en los espectros resultantes, mientras que si no se cumple esta condición, $f_m < 2W$, los espectros de frecuencias se solapan[2] (Fig. 1.18).

$$f_m > 2W \tag{1.7}$$

[2] El autor considera que estos resultados son lo suficientemente ilustrativos para incluirlos sin desarrollar la base teórica de los mismos. Para ampliar los conocimientos en esta materia deberá recurrir a textos sobre teoría de la señal.

El efecto de realizar un muestreo con $f_m < 2W$ sobre la señal que se infiere del muestreo se muestran en la Figura 1.19. La consecuencia directa es que los armónicos superiores de la señal, que no pueden ser caracterizados con los datos obtenidos, producen un armónico ficticio. Este efecto se denomina *aliasing*.

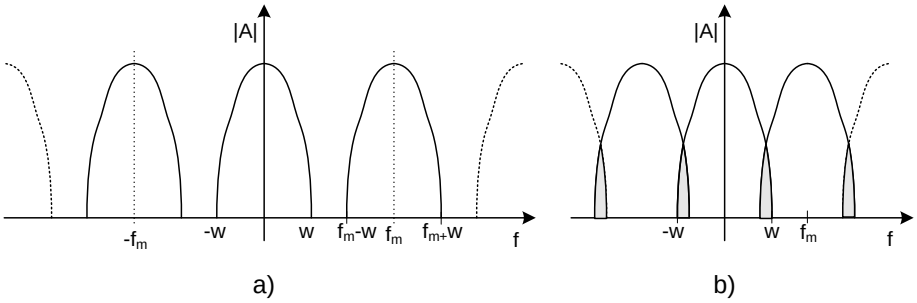


FIGURA 1.18. Espectros de la señal inferida tras el muestreo con a) $f_m > 2W$ y b) $f_m < 2W$, en donde se muestra el solape de frecuencias que se produce.

En realidad el espectro de frecuencias de una señal analógica cualquiera es infinito y por tanto no tiene un ancho de banda limitado, como se ha dado a entender en las Figuras 1.17 y 1.18. En transmisiones analógicas en banda base este "recorte" del espectro lo realizaría el propio medio, sin embargo, para garantizar que tras el muestreo de la señal no aparecen "alias" o armónicos ficticios, como el que se muestra en la Figura 1.18, siempre se realiza un **filtrado antialiasing** de la señal previo al muestreo. Este filtrado recorta el especto (Fig. 1.20), por debajo de doble de la frecuencia de muestreo f_m evitando los problemas de *aliasing*.

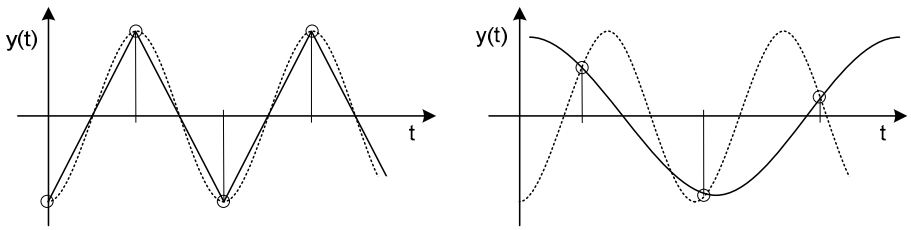


FIGURA 1.19. Efecto del muestreo en el dominio del tiempo con a) $f_m = 2W$ y b) $f_m < 2W$ (*aliasing*).

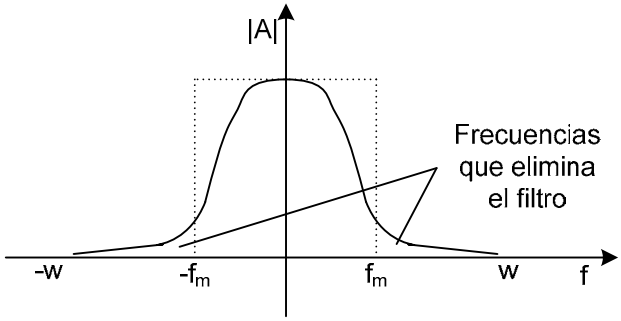


FIGURA 1.20. Filtrado antialiasing.

1.6.2.2.2. Modulación por codificación de impulsos

Las muestras que se han recogido de la señal analógica, son impulsos de señal (Fig. 1.21), que deben ser codificadas en formato digital para poder ser transmitidas. Este proceso se denomina de cuantificación (Fig. 1.22): se le asigna un valor digital a cada impulso analógico. Dado que los códigos digitales son un conjunto limitado (si se codifica con 8 bits hay 2^8 = 256 códigos distintos), este proceso implica cierta pérdida de información por redondeo, que será menor según se vaya incrementando el número de bits con que se cuantifican los impulsos.

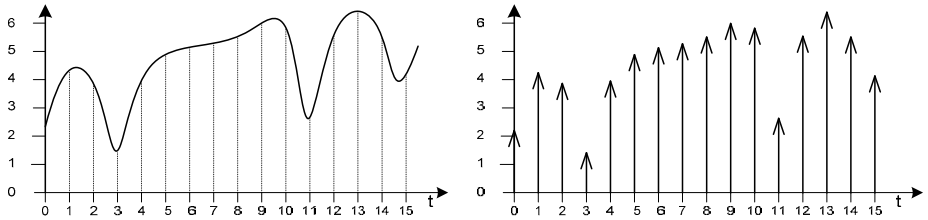


FIGURA 1.21. Impulsos de señal.

Evidentemente, la calidad y fidelidad de la señal resultante tras el muestreo y la cuantificación, implica seleccionar frecuencias de muestreo altas y codificaciones con muchos bits. Cuanto más alta y más bits más fiel será el resultado, pero más ancho de banda será necesario para transmitirlos. La solución buena siempre será un compromiso entre la fidelidad de la digitalización y el ancho de banda necesario.

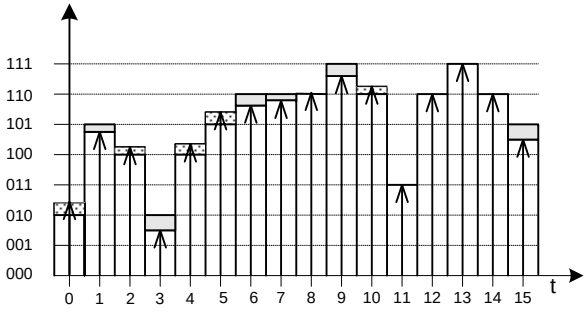


FIGURA 1.22. Cuantificación. Se indican los errores por redondeo que se producen. En gris por exceso y punteado por defecto.

A esta técnica de digitalización se la conoce por sus siglas en ingles PCM (*Pulse Code Modulation*).

1.6.2.2.3. Modulación delta

Otra alternativa de codificación que reduce la cantidad de información a transmitir y la complejidad del proceso de cuantificación es la modulación delta (DM).

La señal analógica que entra en el modulador se compara con la señal digitalizada que sale de él, y con esta información, la amplitud de la señal digitalizada se incrementa o decrementa para acercarse a la señal de entrada. Los incrementos o decrementos se realizan en una cantidad fija δ (delta) en cada nuevo periodo de muestreo. En la Figura 1.23 se muestra el esquema de bloques del modulador delta. La señal de entrada analógica es x(t), la señal de salida binaria a transmitir y(t) y la señal reconstruida r(t).

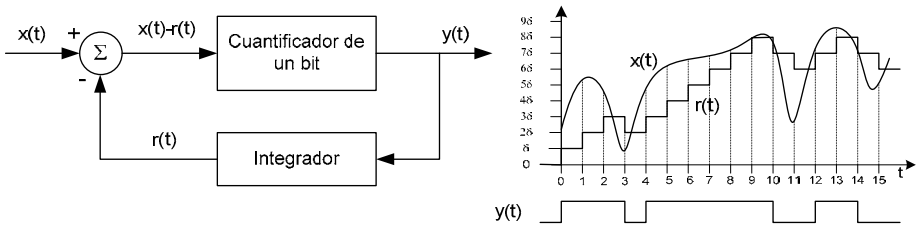


FIGURA 1.23. Modulación delta.

1.7. PRINCIPIOS GENERALES DE LA MULTIPLEXACIÓN

La comunicación ente dos estaciones remotas, normalmente no hará un uso exclusivo de la capacidad de comunicación del canal, sino que por cuestiones de eficiencia de la red compartirá la capacidad del canal con otras comunicaciones. A los procedimientos para compartir el ancho de banda disponible se los denomina multiplexación.

Hay dos técnicas básicas de multiplexación que se pueden combinar entre sí y que son válidas para el envío tanto de voz como de datos: Multiplexación por división de frecuencia (FDM) y Multiplexación por división en el tiempo (TDM). La utilización de cualquiera de las dos técnicas mencionadas implica la realización de dos procesos: la multiplexación al principio de la línea y la demultiplexación, para separar cada comunicación original en el otro extremo del cable.

1.7.1. Multiplexación por división de frecuencia (FDM)

Las líneas de comunicación y las señales sobre las que se aplica multiplexación por división de frecuencia son de naturaleza analógica, así como la señal resultante que se transmite por la línea.

Este procedimiento consiste en repartir el ancho de banda del canal entre varios canales de comunicación y por tanto, el espectro de frecuencias de cada uno de estos canales deberá ser único, sin superposiciones. Para esto se utilizan bandas portadoras sobre las que se modula la señal que contiene la información.

Un ejemplo de esto, es la modulación que se realiza con las comunicaciones telefónicas en banda vocal (entre 0,3 y 3,4 kHz). Si se utiliza una portadora de 20 kHz que se modula con un canal en banda vocal, se obtiene una nueva señal con frecuencias comprendida entre 16,6 y 19,7 kHz, conteniendo la misma información inicial, frecuencia vocal, pero desplazada a otro espectro de frecuencias.

Este mismo proceso se realiza con varias señales telefónicas, desplazando sus bandas de frecuencia a otras superiores y utilizando como portadoras otras tantas señales convenientemente distanciadas en frecuencia. El conjunto así obtenido habrá que agruparlo sumando las señales moduladas en un único ancho de banda que permita su transmisión a través de un solo medio portador (proceso de multiplexado).

Los cables de fibra óptica y la utilización de técnicas de modulación digital han desplazado a los sistemas FDM sobre cables coaxiales que constituyen aun en la actualidad gran parte de la red telefónica.

1.7.2. Multiplexación por división en el tiempo (TDM)

La multiplexación por división en el tiempo (Fig. 1.24), se utiliza sobre canales con transmisión digital, por lo que necesita un proceso de digitalización previo utilizando la técnica de modulación por codificación de impulsos, descrita anteriormente.

A continuación se analiza el proceso como se realiza en los canales de telefonía. Ya se ha indicado que el espectro de frecuencias de estas señales está entre 0,3 y 3,4 kHz, situando la frecuencia de muestreo mínima en 6,8 kHz, según indica el teorema de Shannon. La frecuencia de muestreo convencional utilizada en la red telefónica europea es de 8 kHz que garantiza la correcta codificación de la banda vocal. El empleo de esta frecuencia significa que la señal analógica se muestrea 8000 veces en un segundo (una vez cada 125 microsegundos). El proceso de cuantificación se realiza con códigos de 8 bits, en formato signo-magnitud dando 255 valores de cuantificación, 127 positivos, 127 negativos y cero

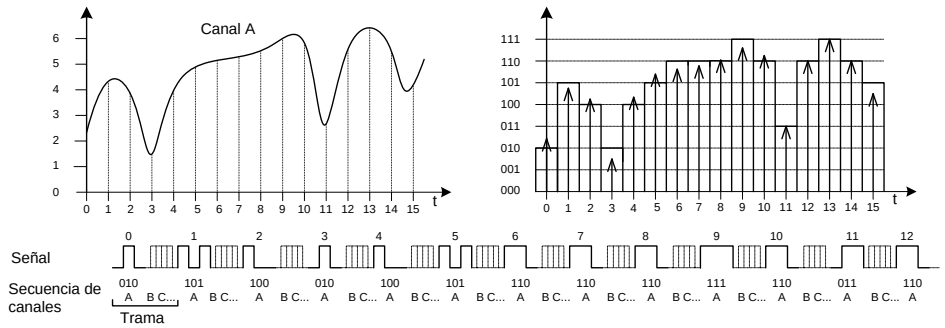


FIGURA 1.24. Multiplexación por división en el tiempo. Muestreo, cuantificación y codificación de un canal. Se muestra la señal multiplexada con los canales A, B, C.

La técnica de modulación de impulsos codificados permite aprovechar los tiempos existentes entre dos muestras consecutivas, pudiéndose incorporar entre aquéllas las muestras de otros canales, formando un conjunto homogéneo. Este procedimiento de reparto en el tiempo de un conjunto de canales PCM se denomina Multiplexación por división en el tiempo (TDM), técnica

generalizada en los sistemas de transmisión de información que actualmente se utilizan. En el caso de los canales de telefonía, la TDM se realiza sobre 32 canales, muestreando (cada uno de ellos) cada 125 microsegundos. Lo que forma el conjunto de información que se pone en la línea durante este periodo de tiempo. Se denomina **trama** a un ciclo que contenga los muestreos consecutivos de los 32 canales.

Cuando las tramas llegan al extremo receptor se realiza la demultiplexación de las señales, en la cual se separan las muestras de cada trama y, tras efectuar un proceso inverso al realizado en transmisión (decodificación y conversión digital–analógica), se distribuyen individualmente sobre cada canal. Las señales resultantes serán analógica, casi idénticas a la originales.

1.8. PROTOCOLOS Y CONTROL DE ENLACE DE DATOS

En un proceso de comunicación, lo que se transmite, cómo se transmite, donde (o por donde) se transmite y cuando se transmite, debe ajustarse a un conjunto de reglas formales que regulan el intercambio de información. Estos acuerdos que deben existir entre los distintos dispositivos que forman la red es lo que se denomina protocolo. Un protocolo es un conjunto de reglas formales mediante las que se establece el formato de los datos, la temporización, la secuenciación, el control del acceso, el control de errores, entre otras muchas cosas. La arquitectura de la red se establece por una constitución física de la misma, pero también por la estructura lógica que establecen los protocolos.

Un protocolo está caracterizado por su:

- **Sintaxis**, define el formato de los datos, la codificación y los niveles de señal volcanes y terremotos en la superficie de la tierra.
- **Semántica**, se ocupa de la sincronización, el control y tratamiento de errores.
- **Temporización**, controla la secuencia de datos y selecciona la velocidad con que los datos de van poniendo en la red.

Cuando dos estaciones establecen una comunicación, ambas deben utilizar idéntico protocolo para que la transferencia de información tenga lugar. Pero estos protocolos no solo se establecen entre la estación emisora y receptora, también entre éstas y los dispositivos de la red y entre estos mismos.

En una comunicación simplex, el transmisor envía el bloque de datos en forma directa al receptor. En este caso los protocolos son sencillos, pero tanto el emisor como el receptor deben utilizar un "lenguaje común". En una comunicación half dúplex, cada bloque de datos transmitido, si es válido, debe ser reconocido por el receptor antes del envío del siguiente bloque de datos. Si se considera que dicho paquete no es válido, se devuelve una señal para que el emisor retransmita el paquete, o de reconocimiento negativo. En este tipo de transmisión, se espera la validación de cada bloque y como se utiliza un único canal de transmisión, la comunicación de los datos no se puede realizar de manera continua. El mecanismo de validación de datos puede realizarse con códigos de redundancia cíclica (CRC) que permiten detectar errores en la recepción. Este tipo de códigos se añade en la cola del paquete de datos.

En comunicaciones dúplex, los protocolos de comunicación se hacen más complejos ya que los datos se pueden mandar sin esperar las señales de reconocimiento. En caso de pérdida de información o errores en la comunicación, debe ser el receptor el que solicite la repetición de envío de los bloques de información que hayan llegado dañados, identificando cuales son.

En toda comunicación, cualquier paquete enviado debe incluir los protocolos de comunicación establecidos. Si se analizan las comunicaciones serie asíncronas más sencillas, se comprueba que todos los paquetes enviados, en este caso caracteres deben incluir el inicio y el final de los datos. En la transmisión síncrona se incorporan señales de sincronización sobre la señal de datos transmitida que permite preparar al receptor para recibir los datos. Según se incrementa la complejidad de los sistemas de comunicación se complica el sistema de protocolos y se hace más necesario.

1.8.1. Acceso al medio

Los métodos para controlar el acceso al medio son necesarios para garantizar que sólo un usuario de la red pueda transmitir en cada momento evitando conflictos y errores. El protocolo de acceso al medio condiciona las características más importantes de la red como la disponibilidad, la fiabilidad, el rendimiento y la gestión de la propia red. Para el estudio de los métodos de acceso al medio se van a clasificar en: de acceso controlado, de acceso aleatorio, de reserva.

1.8.1.1. Acceso controlado

Este tipo de técnicas se basan en conceder a cada estación, por turnos, el permiso para transmitir. El control del acceso puede estar centralizado: técnica de sondeo (*polling*), en el que una estación central va dando acceso a cada uno de los dispositivos periféricos, conectados a un bus; o distribuido, como es el caso de las técnicas de paso de testigo (*token*). Estos métodos regulan el acceso al medio de las estaciones evitando posibles transmisiones simultáneas de dos o más estaciones, colisiones.

El paso por testigo es el método que se utiliza en redes con topología en anillo de tipo Token-Ring. Consiste en hacer circular un patrón de bits especiales que se denomina testigo. Si una estación desea transmitir, debe esperar hasta recibir el testigo; entonces transmite los datos manteniendo el testigo en su extremo final. Si otra estación desea transmitir, retira el testigo del paquete de datos y transmite sus propios datos con el testigo añadido a su extremo final. Este método de acceso es aplicable a topologías en anillo y en bus.

Las principales ventajas del paso por testigo son:

- Aumenta le rendimiento con la carga de la red, sin perder operatividad por saturación.
- Permite establecer prioridades en el acceso.
- Es un método de acceso determinativo. Con un límite superior en el tiempo de espera para la transmisión.

Todas estas características junto con la flexibilidad que aporta esta configuración hacen que sea una de las más utilizadas en procesos de automatización en entornos industriales.

1.8.1.2. Reserva

Este proceso se caracteriza porque las estaciones solicitan la transmisión y el sistema de gestión de la red concede los permisos de inicio de transmisión. Este procedimiento, al igual que el anterior, está libre de colisiones. El sistema de control de la red suele estar centralizado.

El anillo ranurado es un ejemplo de sistema de reserva. En este caso por el anillo circulan unas ranuras, que se corresponden con paquetes de datos vacíos. Si una estación desea transmitir datos, los deposita en la primera ranura vacía que aparezca.

Las ventajas de los métodos de reversa radican en su simplicidad, sin embrago, cuando la carga de la red es baja su rendimiento es malo, dejando vacías la mayoría de las ranuras sin posibilidad de que las utilicen los dispositivos que están transmitiendo.

1.8.1.3. Acceso aleatorio

Las técnicas de acceso aleatorio o contienda, son procesos que compiten por el acceso el medio. Cada estación es libre de comenzar una transmisión en cualquier momento. Esto puede dar lugar a colisiones producidas por transmisiones simultáneas. Básicamente hay dos procedimientos: de transmisión sin escucha y con escucha.

Para las redes de bus o de árbol, el método que en general se emplea es el **Método de acceso múltiple con detección de portadora y detección de colisión** (CSMA/CD). En general este método se relaciona con el bus Ethernet. En el método de acceso CSMA/CD, antes de transmitir las estaciones están pendientes de otras transmisiones en curso, y cualquier estación puede obtener el control de la red y transmitir; por ello el término acceso múltiple. Si no hay actividad, se procede a transmitir. Si hay actividad, el sistema debe esperar hasta que no exista ninguna actividad. A pesar de la escucha necesaria antes de transmitir, es posible que dos o más sistemas empiecen a transmitir al mismo tiempo y causar la colisión de los datos transmitidos por el bus. Esto provoca la corrupción de los datos. Las estaciones que transmiten detectarán la colisión por lo que interrumpirán su transmisión y reintentarán la transmisión transcurrido un tiempo aleatorio.

Éstas técnicas están muy probadas y son muy utilizadas en redes de área local. Sus principales ventajas radican en la simplicidad, su bajo coste y su fiabilidad. Presentando buenos rendimientos hasta niveles de carga medios. También presentan algunos inconvenientes, como mal funcionamiento con cargas de red elevadas, el bajo aprovechamiento del medio cuando los mensajes transmitidos son cortos, debido a los tiempos de espera que impone la difusión por el medio de las señales, y esta misma circunstancia, también limita la longitud máxima de la red.

1.8.2. Control de enlace de datos

El epígrafe anterior se ha centrado en el envío de las señales a través del enlace de transmisión, buscando una comunicación efectiva, un buen control y gestión del enlace. En este epígrafe se centra la atención en la transmisión de datos a través del enlace. Para realizar estas funciones es necesaria una capa lógica por encima de la interfaz física vista anteriormente. A esta lógica se la denomina protocolo de control del enlace de datos. Estos protocolos están encargados de la realización de las siguientes funciones:

- Sincronización de la trama. Los datos se agrupan en bloques que se denominan tramas, que deben tener identificado su principio y su final.

- Control de flujo. Controla la velocidad del transmisor para no saturar la capacidad del receptor.

- Control de errores. Detección y corrección de errores.

- Direccionamiento. Es el mecanismo mediante el cual se identifica la estación de destino de la trama.

- Gestión del enlace. Es el inicio, mantenimiento y finalización del intercambio de datos.

- Control del enlace. Existen tramas que circulan por la red dedicadas al control del enlace que son independientes de las tramas de datos y que deben poder ser diferenciadas por los equipos receptores.

1.9. CONCLUSIONES

Se han desarrollado desde un punto de vista conceptual y descriptivo las características y aspectos fundamentales de la comunicación. Entendiendo este concepto como un proceso universal que tiene lugar en cualquier intercambio de información. El análisis permite comprender que la problemática que presenta es independiente del proceso de comunicación y, por tanto, los principios básicos de las comunicaciones son universales.

Desde estos principios básicos se han aportado las soluciones tecnológicas dadas a las necesidades que plantea una comunicación efectiva.

Así, se han expuesto las característica fundamentales de las señales (entendiendo estas como ondas electromagnéticas) y la problemática que presentan los medios físicos para difundir estas señales (aspectos que se desarrollan en el tema 3); las estructuras y configuraciones topológicas que permiten establecer una comunicación; las necesidades intrínsecas de la propias señales para que estas lleguen a su destino y sean interpretadas correctamente que nos llevan a unas formas de codificación eficientes; la necesidad de optimizar los medios y como consecuencia el coste de los mismos; y por último, el concepto de protocolo que será indispensable entender para abordar los temas sobre los Modelos de OSI o TCP/IP (tema 2) y acceso al medio (tema 4). Se han mencionado otros muchos temas que, en su conjunto, conforman "la comunicación" efectiva, fiable y segura, que se estudiará a lo largo de este texto.

1.10. BIBLIOGRAFÍA

M. Castro y A. Colmenar. Guía multimedia: Sistemas básicos de comunicaciones. Ed. RA-MA, 1999.

P. Mariño. Las comunicaciones en la empresa: Normas, redes y servicios. Ed. RA-MA, 2003.

W. Stallings. Comunicaciones y redes de computadores. Ed. Prentice Hall, 1997.

J. García Tomas. Redes para proceso distribuido. Ed. RA-MA, 2004.

J.A. Carballar. El libro de las comunicaciones del PC. Técnica, programación y aplicaciones. Ed. RA-MA, 1996.

C. Marven y G. Ewers. A simple approach to digital signal processing. Ed. Texas Instruments, 1994.

1.11. EVALUACIÓN

1.11.1. Ejercicios teóricos

1. Una señal analógica sinusoidal esta caracterizada por:

   A) Su frecuencia y su amplitud

   B) El espectro de frecuencia que ocupa

   C) El ancho de banda que necesita para propagarse

   D) Tres parámetros independientes entre sí

   Solución: La respuesta D) es VERDADERA. Los tres parámetros independientes entre sí en general son amplitud, frecuencia y fase.

2. Una señal periódica no sinusoidal:

   A) No puede representarse en el dominio de la frecuencia

   B) Tiene una descomposición en armónicos simples única

   C) Necesita un medio físico específico para su transmisión

   D) Tiene una descomposición en armónicos simples finita

   Solución: La respuesta B) es VERDADERA ya que la descomposición de Fourier es única.

3. El ancho de banda de una señal:

   A) Es mayor si se utiliza fibra óptica en su transmisión

   B) Determina la velocidad de transmisión que se puede alcanzar la comunicación

   C) No es un parámetro determinante de la capacidad de comunicación

   D) Depende de la dimensiones del medio de transmisión, longitud y diámetro

   Solución: La respuesta B) es VERDADERA. Aunque el medio debe ser capaz de transmitirla.

4. La atenuación:

   A) Es un factor presente en todos los medios de propagación

   B) No se produce en la fibra óptica

   C) No se produce en el aire

   D) No tiene una solución que permita incrementar la longitud del medio

   Solución: La respuesta A) es VERDADERA.

5. La capacidad de transmisión del medio:

   A) Depende del ruido presente en el medio y del ancho de banda del canal

   B) No presenta ningún límite teórico aunque sí real

   C) Está limitada únicamente por su ancho de banda

   D) Se incrementa al utilizar codificación multinivel

   Solución: La respuesta A) es VERDADERA.

6. La topología en anillo:

   A) La avería de una estación paraliza la red, dificultando la localización de la avería

   B) Utiliza un canal de difusión común a todas las estaciones

   C) Pueden quedar tramas defectuosas circulando por la red, por lo que requiere algún mecanismo de control de tramas

   D) Se caracteriza por las elevadas longitudes de cableado que precisa

   Solución: La respuesta C) es VERDADERA. La respuesta A) es FALSA ya que la avería de una estación la paraliza pero en general tiene un diagnostico bueno. La B) también es FALSA ya que puede utilizarse un mecanismo de control de tramas pero hay otras opciones.

7. En la transmisión de información digital por canales analógicos:

   A) Toda la información que puede ser enviada es analógica

   B) No se puede establecer comunicación con estas condiciones

   C) Se utilizan únicamente técnicas de codificación en frecuencia

   D) No se puede utilizar codificación multinivel

   Solución: La respuesta A) es VERDADERA, y deberá ser codificada.

8. En la transmisión de señales analógicas por canales digitales:

   A) Es necesario muestrear la señal pero esto no es una tarea crítica de la transmisión

   B) Para la digitalización de la señal es necesario un filtrado previo de la señal

   C) El problema del *aliasing* se produce por la componente continua de la señal analógica

   D) Se realiza un proceso de cuantificación, consistente en asignar un código secuencial creciente a cada muestra

   Solución: La respuesta B) es VERDADERA, La A) es FALSA, ya que el muestreo de una señal es determinante en la velocidad de la comunicación. La C) es FALSA, ya que el *aliasing* se debe a las frecuencias altas que no caracteriza el muestreo Y la D) también es FALSA ya que al proceso de cuantificación se le asigna un código en función de la amplitud de la muestra.

9. Multiplexación por división en el tiempo:

   A) Requiere la modulación de las señales a transmitir

   B) Es la técnica que se utiliza para las comunicaciones radiofónicas

   C) Consiste en la transmisión de impulsos analógicos en distintos instantes de tiempo

   D) En cada instante de tiempo permite sólo una comunicación

   Solución: La respuesta D) es VERDADERA.

10. El control de acceso al medio aleatorio:

    A) Garantiza una alta eficiencia del acceso al estar libre de colisiones

    B) Requiere que todas las estaciones estén “escuchando” el canal

    C) Se utiliza en topologías en anillo de tipo Token Ring

    D) Permite el inicio de la transición de cualquier estación en cualquier momento

    Solución: La respuesta D) es VERDADERA. La A) es FALSA, produce colisiones. La B) es FALSA se puede realizar sin escucha. La C) es FALSA, el anillo utiliza paso por testigo.

1.11.2. Evaluación objetiva

1. La fase, cuando se habla de señales electromagnéticas:

    A) Es el punto alcanzado por la señal en el proceso de comunicación

    B) Es un parámetro relacionado con la frecuencia

    C) Representa el desplazamiento de la señal respecto al tiempo

    D) Únicamente es aplicable a señales analógicas

2. Una onda cuadrada:

    A) Presenta infinitos armónicos impares

    B) No tiene descomposición armónica

    C) Nunca tiene componente continua en el espectro

    D) Tiene un espectro acotado de frecuencias

3. La fibra óptica:

    A) Es inmune a cualquier tipo de perturbación

    B) No presenta atenuación apreciable

    C) Es inmune a la diafonía

    D) En ningún caso presenta distorsión de retardo de propagación

4. El ruido que se produce en el medio físico:

    A) Es únicamente debido a la agitación térmica de los electrones del medio

    B) Es mucho más importante el ruido térmico que la diafonía

    C) Presentan siempre una magnitud constante

    D) Son todos los tipos predecibles salvo el ruido impulsivo

5. La transmisión asíncrona:

   A) Sincroniza el receptor y el carácter transmitido con un bit de inicio

   B) No presenta ningún mecanismo de sincronización de datos

   C) Requiere una codificación de señal de tipo Manchester

   D) Se utiliza siempre cuando la comunicación es de tipo simplex

6. En la transmisión de información analógica por canales analógicos:

   A) Siempre es necesario modular las señales

   B) La modulación se realiza para que varias comunicaciones compartan el canal

   C) Se utilizan técnicas de PCM

   D) Se utiliza codificación PSK

7. En la transmisión de información digital por canales digitales:

   A) No se requiere modulación

   B) No se requiere codificación

   C) Requiere una modulación PCM

   D) No requiere procedimientos de sincronización

8. Multiplexación por división en frecuencia:

   A) Consiste en repartir el medio de comunicación en distintos instantes de tiempo ente los canales de comunicación

   B) Es la técnica que se utiliza para las comunicaciones por la atmósfera

   C) Requiere una digitalización de la señalas señales a transmitir

   D) Se realiza mediante “tramas” de datos

9. Los protocolos:

   A) Constituyen un conjunto de información redundante que se transmite por la red

   B) Son tramas específicas para la transmisión de comandos

   C) Se utilizan únicamente para el control de errores en la transmisión

   D) Son el conjunto de reglas que permite la comunicación

10. La topología en anillo:

    A) Puede estar constituida por una red en bus con un método de acceso por paso de testigo.

    B) Utilizan protocolos de acceso múltiple

    C) Puede presentar colisiones en la comunicación

    D) No utiliza protocolos ya que esta constituida por comunicaciones punto a punto entre terminales

TEMA 2

MODELO OSI. ESTÁNDARES. PROTOCOLOS. MODELO TCP/IP

2.1. Introducción y orientación en el estudio
2.2. Objetivos
2.3. Arquitectura de protocolos
2.4. Modelo OSI
2.5. Modelo TCP/IP
2.6. IP versión 6
2.7. Conclusiones
2.8. Bibliografía
2.9. Evaluación

2.1. INTRODUCCIÓN Y ORIENTACIÓN EN EL ESTUDIO

En estás últimas décadas el avance en las comunicaciones ha sido más que notable. La aparición de Internet y la unión entre redes de diferentes fabricantes provoco la necesidad de que fabricantes de diferentes redes buscarán una solución para crear un software de comunicaciones que permitiera comunicar diferentes redes independientemente del *hardware* utilizado.

El primer modelo o arquitectura de comunicaciones fue el modelo OSI. Dicho modelo es un modelo conceptual que divide la comunicación en 7 capas. Posteriormente apareció el modelo TCP/IP. Actualmente es el más utilizado en Internet y recibe su nombre de sus dos protocolos más importantes: el protocolo de Internet o IP y el protocolo de control de transporte o TCP.

2.2. OBJETIVOS

El objetivo de este capitulo es introducir al lector en el mundo de las comunicaciones. Para ello se definirán conceptos como: protocolo, jerarquías de protocolos y estándares utilizados.

En la actualidad el modelo TCP/IP es el modelo más utilizado en Internet. En este capitulo se describirá que esta formado en cinco capas. Cada una de estas capas tendrá unas funcionalidades determinada que será ofrecidas por los protocolos pertenecientes a dichas capas.

Por último se describirá una nueva versión del protocolo IP conocida como IPv6, que pretende resolver problemas que están empezando a aparecer en Internet.

2.3. ARQUITECTURA DE PROTOCOLOS

Las primeras redes se diseñaron teniendo en cuenta al *hardware* como elemento principal y dejando el software en un segundo plano. En la actualidad esta práctica no funciona, existen un gran número de redes distintas que deben interconectarse y comunicarse entre si. De ahí que el *software* haya adquirido una gran importancia a la hora de diseñar redes de comunicación.

Para reducir la complejidad de su diseño, la mayoría de las redes están organizadas como una pila de niveles o capas (Fig. 2.1). Cada una de ellas está construida a partir de la capa inferior a ella. El número de capas, así como el nombre, contenido y función de cada una de ellas difieren de red a red. El propósito de cada capa o nivel es ofrecer ciertos servicios a la capa superior. A dichas capas no se les muestran los detalles de implementación de los servicios ofrecidos por la capa inmediatamente inferior.

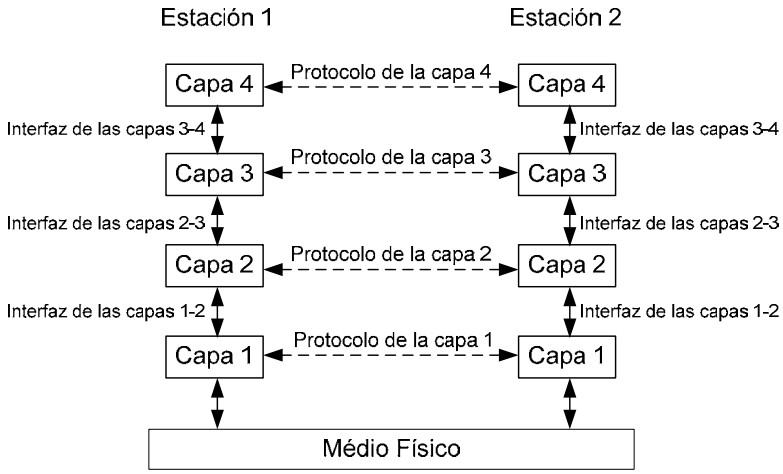


FIGURA 2.1. Capas, protocolos e interfaces de una red de 4 capas.

La capa n de una máquina mantiene una conversación con la capa n de otra máquina. Esta conversación sigue unas reglas y convenciones que se conoce como protocolo de capa n. Por tanto, se puede decir que un protocolo es un acuerdo entre las partes en comunicación sobre como se debe llevar a cabo la comunicación.

Entre cada capa adyacente existe una interfaz o punto de acceso a servicio (SAP, *Service Access Point*). Dicha interfaz define que operaciones y servicios primitivos pone la capa inferior a disposición de la capa superior. Cuando los diseñadores de redes deciden el número de capas a incluir en la red y cual es la función de cada una de ellas. También deben definir interfaces adecuadas entre las capas.

A continuación se va a describir brevemente que ocurre cuando una máquina 1 desea transmitir datos a otra máquina 2 de la red:

- Cada capa pasa los datos y la información de control a la capa inmediatamente inferior hasta llegar a la capa más baja.

- La capa más baja transmite la información a través del medio físico.

- La capa inferior de la máquina 2 recoge la información. Cada capa inferior pasa la información a la capa inmediatamente superior.

Es importante indicar que a medida que se pasa información e una capa superior a una inferior se añade un encabezado que contiene información de control. Cuando la máquina destinataria recoge la información, las capas inferiores comprueban el encabezado que le corresponde, lo elimina y pasa el resto de información a su capa inmediatamente superior (Fig. 2.2).

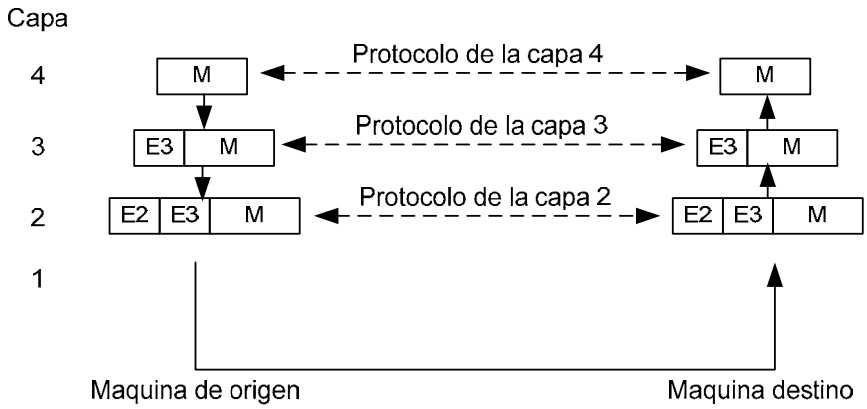


FIGURA 2.2. Ejemplo de flujo de información que soporta una comunicación en la capa 4.

La unión de los datos generados por la capa superior, junto con la información de control de la capa actual, se denomina unidad de datos del protocolo (PDU, *Protocol Data Unit*).

2.3.1. Arquitectura de protocolos normalizada

El aumento en el número de comunicaciones y de redes de distintos fabricantes ha provocado que el desarrollo de *software* de comunicaciones de propósito específico sea demasiado costoso para ser aceptable. La única alternativa para los fabricantes es adoptar e implementar un conjunto de

convenciones comunes o estándares. Estos estándares proporcionan las siguientes ventajas:

- Los fabricantes que desarrollan sus productos de acuerdo a los estándares tienen un mercado mayor.

- Los clientes pueden exigir a cualquier fabricante implemente los estándares.

Existen dos arquitecturas que han sido determinantes y básicas en el desarrollo de los estándares de comunicación: el conjunto de protocolos TCP/IP (*Transport Control Protocol/Internet Protocol*) y el modelo de referencia OSI (*Open System Interconnection*). Cada una de estas arquitecturas se tratará en los siguientes apartados.

2.4. MODELO OSI

Es un modelo basado en una propuesta desarrollada por la organización internacional de estándares (ISO, *International Organization for Standardization*). El modelo recibe el nombre de OSI ya que tiene que ver con la conexión de sistemas abiertos a la comunicación con otros sistemas.

El modelo OSI define siete capas (Fig. 2.3). Los principios que se aplicaron para llegar a dichas capas:

- Una capa se crea cuando se necesite una abstracción diferente.

- Cada capa debe realizar una función bien definida.

- La función de cada capa se debe elegir con la intención de definir protocolos estandarizados de internacionales.

- Los límites de las capas se deben elegir con el fin de minimizar el flujo de información a través de las interfaces.

- El número de capas debe ser suficientemente grande para no tener que agrupar funcionalidades distintas dentro de una misma capa y lo suficientemente pequeña para que la arquitectura no se vuelva inmanejable.

- Permitir que los cambios en las funciones o protocolos se puedan llevar a cabo sin afectar a otras capas.

- Crear posteriormente subagrupamientos y reestructurar las funciones formando subcapas dentro de una capa en aquellos casos en los que se necesite diferentes servicios de comunicación.

A continuación se van a describir cada una de las capas que forma el modelo OSI.

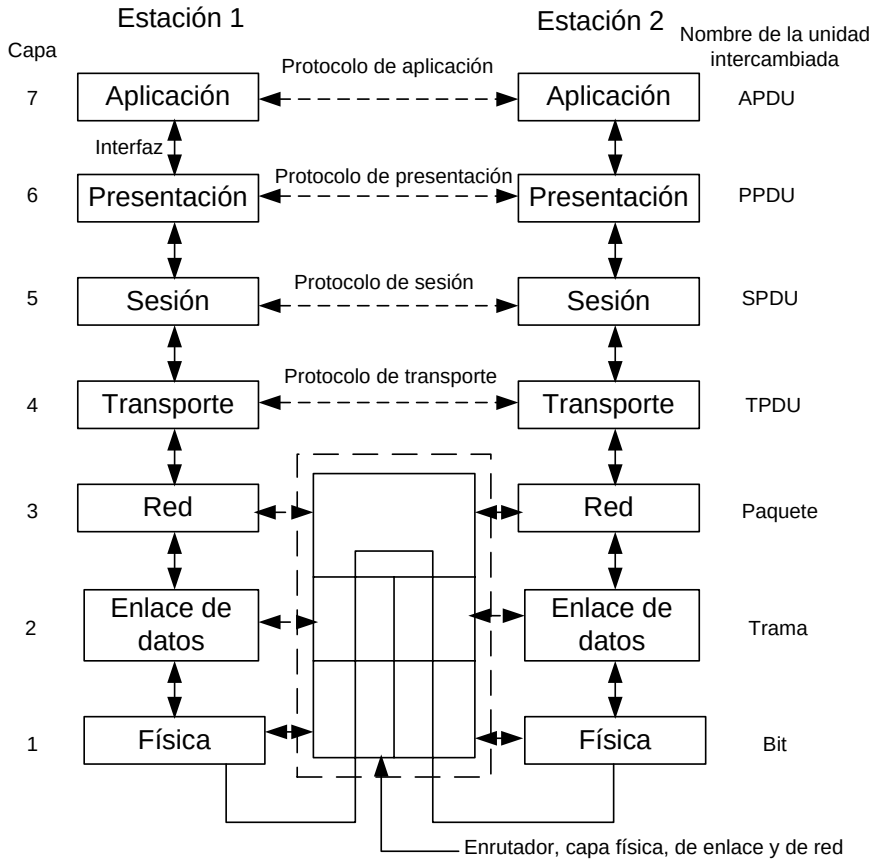


FIGURA 2.3. Capas y unidades de intercambio del modelo OSI.

2.4.1. Capa física

La capa física es la encargada de interactuar con el medio físico. Además, establece las reglas que rigen la transmisión de los bits. La capa física tiene cuatro características importantes:

- Mecánicas. Relacionadas con las propiedades físicas de la interfaz con el medio de transmisión. Dentro de estas características se incluye la especificación del conector que transmite las señales a través del conductor o circuitos.

- Eléctricas. Especifica como se representan los bits, por ejemplo en términos de niveles de tensión. Así como la velocidad de transmisión.

- Funcionalidades. Especifican las funciones que realiza cada uno de los circuitos de la interfaz física entre el sistema y el medio físico.

- De procedimiento. Especifican la secuencia de eventos que se llevan a cabo en el intercambio del flujo de bits a través del medio físico.

Un ejemplo de estándar de esta capa es el EIA-232_F que cubre cada una de las cuatro características mencionadas anteriormente:

- Mecánicas: ISO 2110.

- Eléctricas. V.28 de la UITT.

- Funcionales: V.24 de la UITT.

- De procedimiento: V.24 de la UITT.

2.4.2. Capa de enlace de datos

Mientras que la capa física proporciona exclusivamente un servicio de transmisión de datos, la capa de enlace de datos proporciona los medios para activar, mantener y desactivar un enlace físico fiable.

El principal servicio de está capa es el de detección y control de errores. Para lograr esto el emisor debe fragmentar los datos de entrada en tramas de datos y transmitirlas de manera secuencial. Si el servicio es confiable, el

receptor confirma la recepción de cada una de las tramas, mediante una trama de confirmación de recepción.

Otra cuestión que trata es como hacer que un transmisor rápido no sature de datos a un receptor lento. Por lo general se utiliza un mecanismo de regulación de tráfico que indique al transmisor cuánto espacio de buffer (memoria de datos) tiene el receptor en ese instante.

Las redes de difusión tienen un aspecto adicional en la capa de enlace de datos que es el como controlar el acceso al canal compartido.

Algunos ejemplos de estándares de la capa de enlace son los protocolos: HDDL (*High-Level Data-Link Control*) y LLC (*Logical Link Control*).

2.4.3. Capa de red

Esta capa controla las operaciones de la subred. Un aspecto clave es determinar como se erutan los paquetes desde su origen a su destino dentro de la red de comunicación. Las rutas pueden estar basadas en tablas estáticas (enrutamiento estático) codificadas en la red y que rara vez cambian o pueden ser un enrutamiento dinámico donde las rutas pueden cambiar para reflejar la topología o el estado de la red.

La capa de red puede controlar la congestión, aunque esta función también puede ser compartida por la capa de transmisión. La capa de red debe solucionar problemas tales como el direccionamiento de los equipos dentro de la red, el retardo de paquetes, el tiempo de transito, etc.

2.4.4. Capa de transporte

Esta capa tiene como función básica aceptar los datos de las capas superiores, dividirlos en unidades más pequeñas si es necesario, pasar estas unidades a la capa de red y asegurarse que todas las piezas llegan correctamente al otro extremo. Todo esto debe realizarse de forma que aisle las capas superiores de los cambios inevitables en la tecnología hardware.

La capa de transporte determina el tipo de servicio proporcionado a la capa de sesión y a los usuarios de la red. Los tipos de servicios que proporciona la capa de transporte son:

- Un canal punto a punto libre de errores, que entrega mensajes o bytes en el orden en que se enviaron.

- Otra opción en transportar mensajes aislados, que no garantizan el orden de entrega. Y la difusión de mensajes a múltiples destinos.

2.4.5. Capa de sesión

Permite que los usuarios de máquinas diferentes establezcan sesiones entre ellos. Los servicios proporcionados por esta capa son:

- Control de dialogo. Éste puede ser simultáneo en los dos sentidos (*Full Duplex*) o alternado en ambos sentidos (*Half Duplex*).

- Administración de *token*. Impide que las dos partes traten de realizar la misma operación crítica al mismo tiempo.

- Sincronización. Insertar puntos de referencia a transmisiones largas para permitirles continuar desde donde se encontraban después de una caída.

2.4.6. Capa de presentación

Define el formato de los datos a transmitir. También ofrece a los programas de aplicación un conjunto de servicios de transformación de datos.

Su fin es que ordenadores con diferentes representaciones de datos se puedan comunicar, las estructuras de datos que se intercambian se pueden definir de una manera abstracta, junto con una codificación estándar para su uso. La capa de presentación maneja estas estructuras y permite definir e intercambiar estructuras de datos de un nivel más alto.

Algunos ejemplos de servicios específicos que se pueden realizar en esta capa son los de compresión y el cifrado de datos.

2.4.7. Capa de aplicación

Esta capa administra y proporciona los mecanismos genéricos necesarios para la implementación de aplicaciones distribuidas. En esta capa también residen las aplicaciones de uso general como la transferencia de archivos, el

correo electrónico, la *World Wide Web*, el acceso de terminales a computadores remotos, etc.

A continuación se va a mencionar algunos de los protocolos que pertenecen a dicha capa:

- Protocolo de transferencia de hipertexto o HTTP (*Hypertext Transfer Protocol*).
- Protocolo de transferencia de ficheros o FTP (*File Transfer Protocol*).
- Protocolo simple de transferencia de correo electrónico o SMTP (*Simple Mail Transfer Protocol*).

2.5. MODELO TCP/IP

El modelo o arquitectura TCP/IP fue el modelo utilizado por ARPANET y es el modelo utilizado por la actual Internet. ARPANET fue desarrollada por el departamento de defensa de los Estados Unidos, con el paso del tiempo conectó cientos de universidades e instalaciones gubernamentales. Más tarde dicha red se liberó para permitir que diferentes redes de todo el mundo se pudieran conectar entre si. Dando lugar a lo que actualmente se conoce como Internet.

Es conveniente diferenciar que existen 2 tipos de estándares:

- Los estándares de jure. Son aquellos desarrollados y aprobados por organizaciones de ámbito nacional o Internacional como: IEEE, ISO o CEN/ISSS.
- Los estándares de facto. Son aquellas aplicaciones, patrones o normas que aunque no hayan sido aprobadas por las organizaciones de estandarización (IEE, ISO, etc.), han llegado a ser tan utilizadas que tanto los fabricantes como los usuarios solicitan que sus equipos o redes lo tengan. Un ejemplo claro de este tipo de estándar es el TCP/IP.

La arquitectura TCP/IP recibe este nombre debido a dos de sus principales protocolos: El protocolo de Internet o IP (*Internet Protocol*) y el protocolo de control de transmisión o TCP (*Transmisión Control Protocol*). TCP/IP está estructurada en 5 capas (Fig. 2.4):

- Capa física. Define la interfaz física entre el computador, terminal, etc. Y el medio de transmisión o red. Se ocupa de la especificación de las características del medio de transmisión, de la naturaleza de las señales, de la velocidad de datos y cuestiones similares.

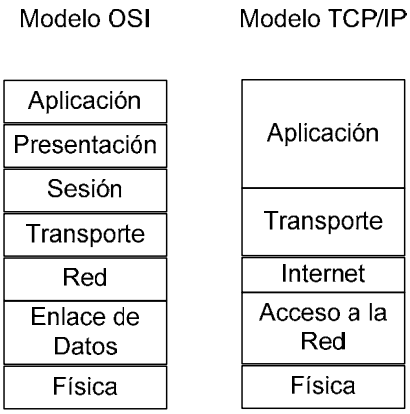


FIGURA 2.4. Comparación del modelo TCP/IP con el modelo OSI.

- Capa de acceso a la red. Es la responsable del intercambio de datos entre el sistema final (terminal, estación de trabajo, etc.) y la red a la cual está conectado. El emisor debe proporcionar a la red la dirección destino, de tal manera que esta pueda encaminar los datos hasta el destino apropiado. El tipo de *software* utilizado en esta capa dependerá del tipo de red que se disponga. Por lo que se han desarrollado diversos estándares para la conmutación de circuitos, la conmutación de paquetes y para las redes de área local (como por ejemplo Ethernet). Esto permite al software de comunicaciones de las capas superiores no tengan que ocuparse de los detalles específicos de la red a utilizar.

- Capa de Internet. Como ya se ha comentado, la capa de acceso a la red está relacionada con el acceso y el encaminamiento de los datos. Cuando los dos dispositivos que se quieren comunicar están en redes diferentes, se necesitarán una serie de procedimientos que los datos atraviesen las distintas redes interconectadas. Esta es la función de la capa Internet. Dicha capa utiliza el protocolo de Internet o IP.

- Capa de transporte. Está diseñada para permitir que las entidades iguales en el host de origen y destino. En dicha capa se han definido dos protocolos de transporte extremo a extremo.

- Protocolo de control de transmisión o TCP. Es un protocolo confiable, orientado a conexión, que permite que un flujo de bytes que se origina en una máquina se entregue sin errores en cualquier otra máquina de la interred.

- Protocolo de datagrama de usuario o UDP (*User Datagram Protocol*). Es un protocolo no confiable y no orientado a conexión.

- Capa de Aplicación. Contiene todos los protocolos de nivel superior. Estos protocolos son utilizados por aplicaciones como navegadores, correo electrónico, etc.

TCP/IP es por tanto una arquitectura de 5 capas y que recibe el nombre de sus dos principales protocolos. Cada una de esta dispone de diferentes protocolos de comunicación (Fig. 2.5). A continuación se van a describir de forma más detallada los protocolos IP y TCP.

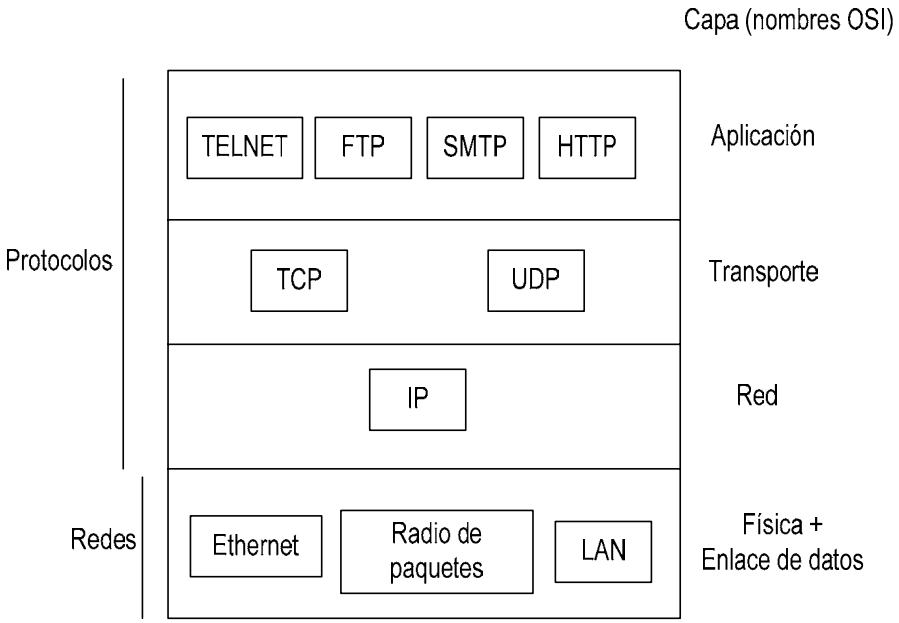


FIGURA 2.5. Algunos de los protocolos y redes utilizados en el modelo TCP/IP.

2.5.1. Internet y TCP/IP

Internet no es una red si no un conjunto de redes diferentes conectadas entre si. Una red de redes TCP/IP proporciona un conjunto de servicios (Fig. 2.6).

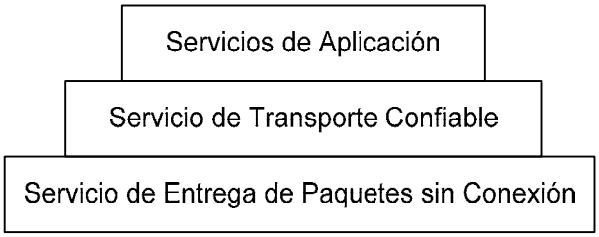


FIGURA 2.6. Capas conceptúales de los servicios de Internet.

Un servicio no confiable es aquel en que la entrega de los paquetes no está garantizada. Por tanto los paquetes se pueden perder, duplicar, retrasar o entregar sin orden, pero el servicio no detectará estas condiciones ni informará al emisor o al receptor.

Un servicio se denomina sin conexión o no orientado a conexión ya que cada paquete es tratado de manera independiente al anterior. Una secuencia de paquetes que se envían puede viajar por diferentes rutas.

2.5.2. Protocolo de Internet o IP

Es un protocolo no orientado a conexión y no confiable. El protocolo IP proporciona tres definiciones importantes:

1. Especifica el formato exacto de todos los datos que pasarán a través de una red de redes TCP/IP.

2. El software IP realiza la función de encaminamiento (ruteo), seleccionando la ruta por la que los datos serán enviados.

3. IP incluye un conjunto de reglas que le dan forma a la idea de entrega no confiable. Estas reglas caracterizan la forma en que los anfitriones y encaminadores deben procesar los paquetes, como y cuando se deben generar los mensajes de error y las condiciones bajo las cuales los paquetes pueden ser descartados.

La unidad de transferencia en el protocolo IP se conoce como datagrama. Este se divide en un área de encabezado y un área de datos (Fig. 2.7). La Figura 2.8 muestra de manera más detallada la estructura de un datagrama IP.

FIGURA 2.7. Forma general de un datagrama IP.

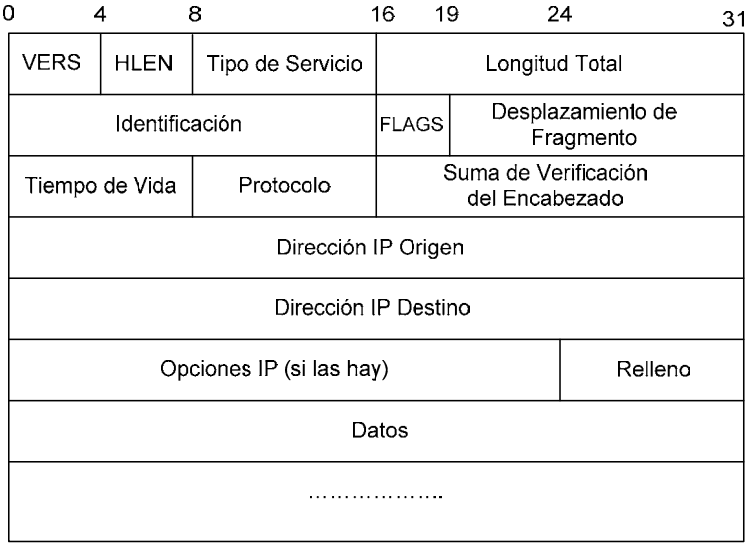


FIGURA 2.8. Formato de un datagrama IP.

A continuación se describirá cada uno de los campos que forma el datagrama:

- VERS. Contiene la versión del protocolo IP que se utilizo para crear el datagrama. El más común es la versión 4.

- HLEN o longitud del encabezado. Como se puede observar en la Figura 2.8, todos los campos del encabezado tienen una longitud fija excepto el campo opciones IP y relleno. El encabezado más común, no contiene opciones ni relleno tiene una longitud de 20 bytes u octetos.

- Tipo de servicio. Especifica como debe manejarse el datagrama. El campo está subdividido en 5 subcampos (Fig. 2.9).

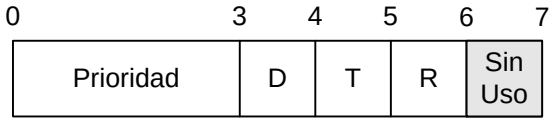


FIGURA 2.9. Estructura del campo tipo de servicio.

  - Prioridad. Indica la prioridad del datagama. Esta prioridad abarca desde el 0 (prioridad normal) hasta el 7 (Control de red).

  - El bit D (*Delay*) o retardo.

  - El bit T (*Throughput*) o velocidad real de transporte.

  - El bit R (*Reliability*) o confiabilidad.

  Estos tres bits permite a los encaminadores tomar decisiones del camino.

  - El último de los bits no tiene uso.

- Longitud total. Proporciona la longitud en bytes del datagrama, incluye los octetos del encabezado y de los datos. La longitud máxima del datagrama es de 65.535 bytes. Actualmente este tamaño de datagrama es aceptable, pero con las redes de alta velocidad se necesitarán datagramas mayores.

- Identificación. Es utilizado para que la máquina destino determine a que datagrama pertenece el fragmento recién llegado. Todos los fragmentos de un datagrama contienen el mismo valor de identificador.

  Como se ha visto en apartados anteriores, los datos a enviar van bajando por las diferentes capas del modelo y se encapsulan dentro de otra unidad como datos. Así los datagramas pertenecen al nivel de red por lo que al pasar al nivel de enlace son encapsuladas como datos dentro de una trama (Fig. 2.10).

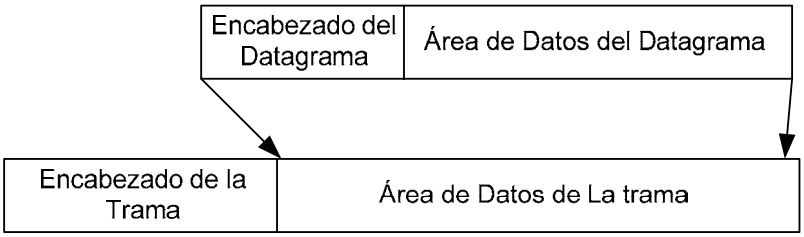


FIGURA 2.10. Encapsulación de un datagrama IP en una trama.

Cada tecnología de conmutación de paquetes establece un número superior fijo para la cantidad de datos que pueden transferirse en una trama física. Esto se conoce como unidad de transferencia máxima de una red o MTU (*Maximum Transfer Unit*). Por ejemplo en Ethernet la MTU es de 1500 bytes, mientras que en FDDI la MTU es de 4.470 bytes. Hay tecnologías *hardware* que solo permiten 128 bytes. Los datagramas pueden ser mayores que las MTU por lo TCP/IP selecciona desde el principio un tamaño de datagrama más conveniente y establece una forma para dividir el datagrama en fragmentos que puedan viajar por la red (Fig. 2.11).

- FLAGS. Esta formado por 3 bits:
  - El primer bit no tiene uso.
  - El segundo bit o DF (*Don´t Fragment*) significa no fragmentar y es una orden para los encaminadores de que no fragmente el datagrama, ya que el destino es incapaz de juntar las piezas de nuevo.
  - El tercer bit o MF (*more fragments*) indica si hay más fragmentos del datagrama o no.
- Desplazamiento de fragmentación. Indica en que parte del datagrama actual va este fragmento. Todos los fragmentos excepto el último deben tener un múltiplo de 8 bytes.
- Tiempo de vida. Es un contador que sirve para limitar el tiempo de vida del paquete dentro de la red de redes. Este tiempo se cuenta en segundos, máximo 255 segundos y debe disminuir en cada salto.
- Suma de verificación del encabezado. Asegura la integridad.

- La dirección IP origen y la dirección IP destino indican la red y máquina que genero el datagrama y la red y máquina destinataria del datagrama.

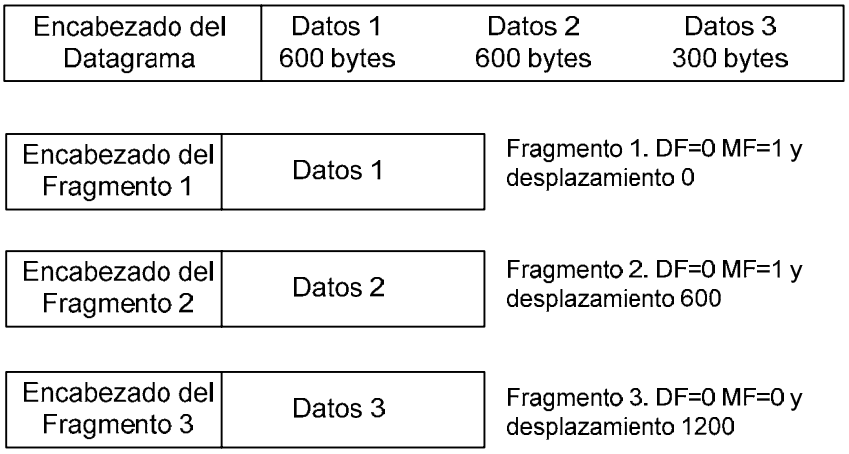


FIGURA 2.11. Fragmentación de un datagrama de 1500 bytes para una red con MTU de 620 bytes.

Las direcciones IP son de 32 bits e indican la red y el *host* o máquina. Existen 5 categorías (Fig. 2.12).

- Opciones IP. No es necesaria que este en todos los datagramas. Estas opciones suelen utilizarse para pruebas de red o depuración.

- Relleno. Como su propio nombre indica son bits de relleno.

- Datos. Son los datos de información recibidos de la capa superior.

2.5.3. Protocolo de control de transmisión o TCP

Como ya se ha comentado, el protocolo IP proporciona una entrega de paquetes no confiable. Los paquetes se pueden perder o destruir cuando los errores interfieren con los datos, cuando falla el hardware de red o cuando las redes se sobrecargan demasiado. El protocolo TCP se diseño para proporcionar un flujo de bytes confiable extremo a extremo a través de una red de redes no confiable.

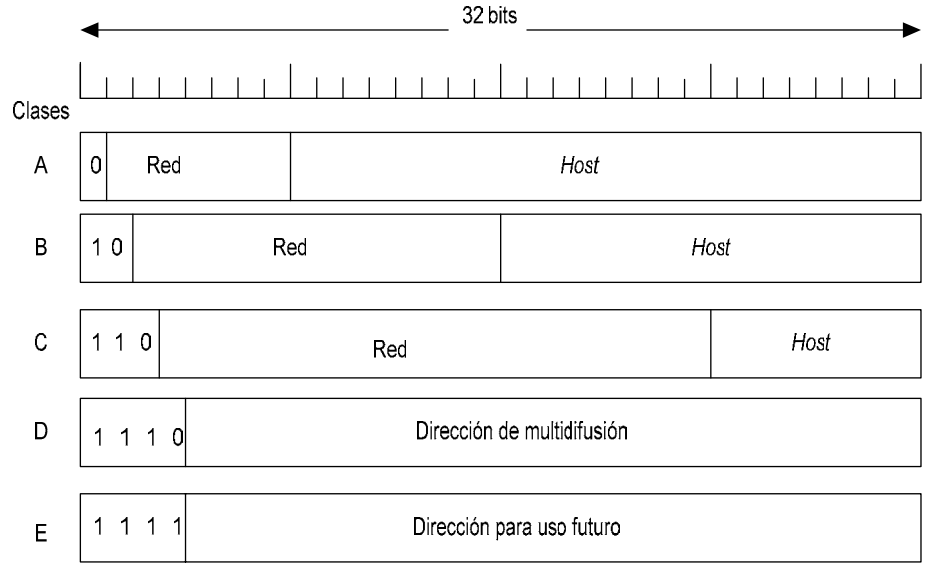


FIGURA 2.12. Formato de dirección IP.

El servicio TCP se obtiene al hacer que tanto el servidor como el cliente creen puntos terminales, llamados *sockets*. Cada *socket* tiene un número o dirección que consiste en:

- La dirección IP del *host*.
- Un número de 16 bits que es local a ese *host*, llamado puerto. Los números de puerto menores a 1024 se reservan para servicios estándares (Tabla 2.1).

Para obtener el servicio TCP se debe establecer de manera explicita entre un *socket* de la máquina emisora y uno de la máquina emisora. Dos o más conexiones pueden terminar en el mismo *socket*. Las conexiones se identifican mediante los identificadores de *socket* de los dos extremos (*socket1, socket2*).

La entidad TCP emisora y receptora intercambian datos en forma de segmentos. Un segmento esta formado por un encabezado TCP fijo de 20 bytes, más una parte opcional de cero o más bytes de datos (Fig. 2.13). A continuación se describirán los campos que forman dicho encabezado.

- Puerto origen y puerto destino. Identifican los puntos terminales locales de la conexión. La dirección de un puerto más la dirección IP de su *host* forman un punto de Terminal único de 48 bits. Los puntos terminales de origen y de destino en conjunto identifican una conexión.

TABLA 2.1. Algunos puertos reservados

Puerto	Protocolo	uso
21	FTP	Transferencia de ficheros
23	TELNET	Inicio remoto de sesión
25	SMTP	Correo electrónico
69	TFTP	Transferencia de ficheros trivial
80	HTTP	*World Wide Web*
110	POP-3	Acceso remoto al correo electrónico

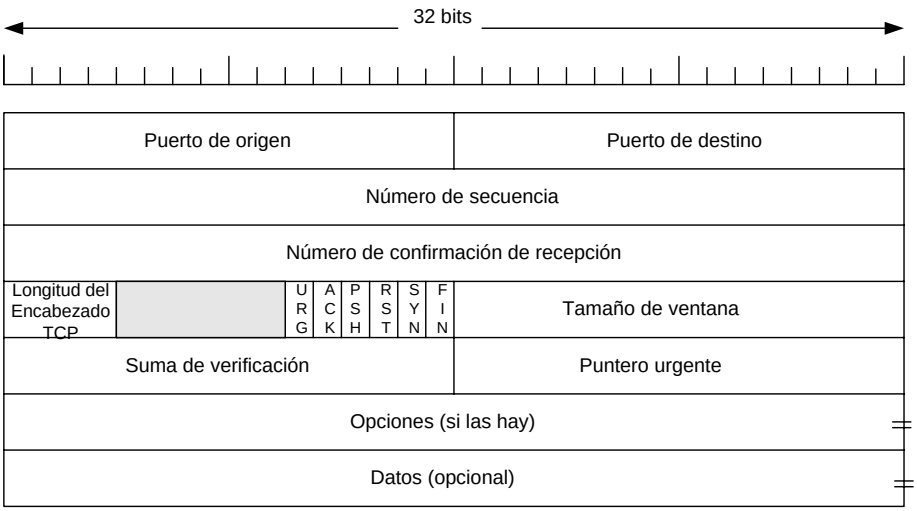


FIGURA 2.13. Formato de un segmento TCP.

- Número de secuencia. Identifica la posición de los datos del segmento en el flujo de datos del emisor.

- Número de confirmación de recepción. Identifica el número de octetos que el origen espera recibir después. TCP es un protocolo confiable

donde emisor y receptor establecen una conexión. El receptor debe confirmar el establecimiento de la conexión y los datos que recibe.

- Longitud del encabezado. Contiene un número entero que especifica la longitud del encabezado del segmento, medida en múltiplos de 32 bytes.

- Bit URG. Si el bit está puesto a uno, indica que el campo de puntero urgente es valido.

- Bit ACK. Si el bit está puesto a uno, indica que el número de confirmación de recepción es valido. Es decir que el segmento contiene una confirmación de recepción.

- Bit PSH. Si el bit está puesto a uno, indica datos que se tienen que transmitir de inmediato. Por este medio se solicita atentamente al receptor que entregue los datos a la aplicación a su llegada y no los almacene en búfer hasta la recepción de un buffer completo.

- Bit RST. Si el bit está puesto a uno, indica iniciación de la conexión.

- Bit SYN. Si el bit está puesto a uno, sincroniza números de secuencia.

- Bit FIN. Si el bit está puesto a uno, indica que el emisor ha llegado al final de su flujo de octetos.

- Tamaño de ventana. Indica la cantidad de bytes que pueden enviarse comenzando por el byte cuya recepción se ha confirmado. TCP utiliza una técnica de control de flujo denominada ventana deslizante (Fig. 2.14). Esta consiste en utilizar una “ventana” que divide la secuencia de paquetes en tres:

    - Los paquetes a la izquierda de la ventana se transmitieron, recibieron y se envío un acuse de recibo con éxito.

    - Los paquetes a la derecha de la ventana no se han transmitido.

    - Los paquetes que están dentro de la ventana están en proceso de transmisión y no se ha hecho acuse de recibo.

- Suma de verificación. Se utiliza para comprobar la integridad de los datos y del encabezado.

- Puntero urgente. Sirve para indicar un desplazamiento en bytes a partir del número actual de secuencia en el que se encuentran datos urgentes.

- Opciones. Es una forma de agregar características extras que no son cubiertas por el encabezado normal.

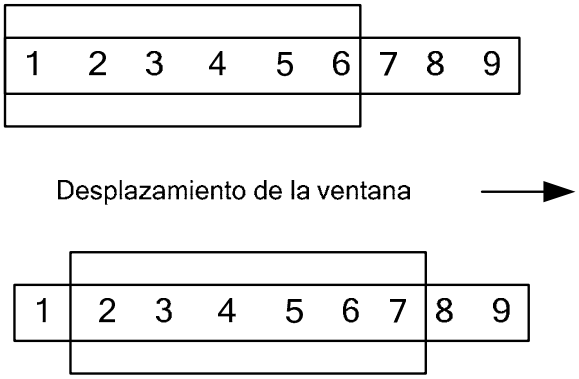


FIGURA 2.14. Ventana deslizante de 6 paquetes.

2.5.4. Otros protocolos de TCP/IP

En los dos últimos apartados se han descrito brevemente los protocolos principales de la arquitectura TCP/IP y las unidades de datos utilizadas. TCP/IP incluye otros muchos protocolos. A continuación se mencionarán algunos de estos:

- Protocolo de datagramas de usuario o UDP (*User Datagram Protocol*). Pertenece a la capa de transporte y a diferencia de TCP permite que las aplicaciones envíen datagramas IP encapsulados sin tener que establecer una conexión.

- Protocolo de transferencia de hipertexto o http. Pertenece a la capa de aplicación y permite a los navegadores mostrar contenido de la *World Wide Web*.

- Protocolo de sistema de nombres de dominio o DNS (*Domain Name System*). Pertenece a la capa de aplicación e implanta una jerarquía de nombres de máquina para las redes de redes TCP/IP.

- Protocolo TELNET. Pertenece a la capa de aplicación que establece un Terminal remoto a otra máquina.

- Protocolo de transferencia de ficheros o FTP. Pertenece a la capa de aplicación y permite el intercambio de ficheros de una máquina a otra.

2.6. IP VERSIÓN 6

La evolución de las tecnologías de comunicación, el aumento de redes conectadas y por tanto el aumento de direcciones IP de *host* conectados a Internet ha dado lugar a nuevas necesidades que la versión 4 del protocolo no puede resolver. Por estos motivos se comenzó a trabajar en una nueva versión del protocolo IP que cubriera los siguientes objetivos:

- Manejar miles de millones de *hosts*.

- Disminuir el tamaño de las tablas de enrutamiento.

- Simplificar el protocolo, para permitir a los encaminadores el procesado más rápido de los paquetes.

- Proporcionar mayor seguridad que el IP actual.

- Prestar mayor atención al tipo de servicio, especialmente con datos en tiempo real.

- Posibilitar que un *host* sea móvil sin cambiar su dirección.

- Permitir que el protocolo antiguo y el nuevo coexistan sin problemas.

IPv6 cumple estos objetivos bastante bien. Mantiene las características del IP, elimina y reduce las malas, y agrega nuevas características donde se necesitan. Los cambios introducidos por IPv6 son:

- Direcciones de mayor longitud. IPv6 utiliza direcciones de 128 bits, a diferencia de los 32 bits utilizados por IPv4. El espacio de direcciones es tan grande que no podrá agotarse en un futuro próximo.

- Formato de encabezados flexible. IPv6 utiliza un formato de datagrama incompatible con IPv4 (pero compatible con los protocolos TCP, UDP, ICMP, etc.) y completamente nuevo.

- Opciones mejoradas. IPv6 permite que un datagrama incluya información de control opcional.

- Soporte para asignación de recursos. IPv6 reemplaza la especificación tipo de servicio del IPv4 con un mecanismo que permita la preasignación de recursos de red. Por ejemplo las aplicaciones de video en tiempo real necesitan de una garantía de ancho de banda y de retardo.

- Provisión para extensión del protocolo. Tiene la capacidad de extensión.

IPv6 cambia por completo el formato del datagrama. Así un datagrama contiene un encabezado base de tamaño fijo, a continuación le siguen cero o más encabezados de extensión y por último los datos a enviar (Fig. 2.15).

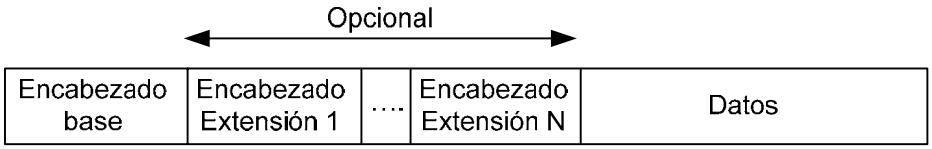


FIGURA 2.15. Formato general del datagrama IPv6.

2.6.1. Encabezado base

En este apartado se van a describir brevemente los campos que forman el encabezado base del datagrama (Fig. 2.16). El tamaño del encabezado base es de 40 bytes.

- Versión. En este caso siempre será 6. Este campo lo utilizan los encaminadores para saber el tipo de paquete que tienen que encaminar.

- Clase de tráfico. Se utiliza para distinguir entre los paquetes con requisitos diferentes de entrega en tiempo real.

- Etiqueta de flujo. Se utilizará para permitir a un origen y a un destino establecer una pseudoconexión con propiedades y requisitos particulares.

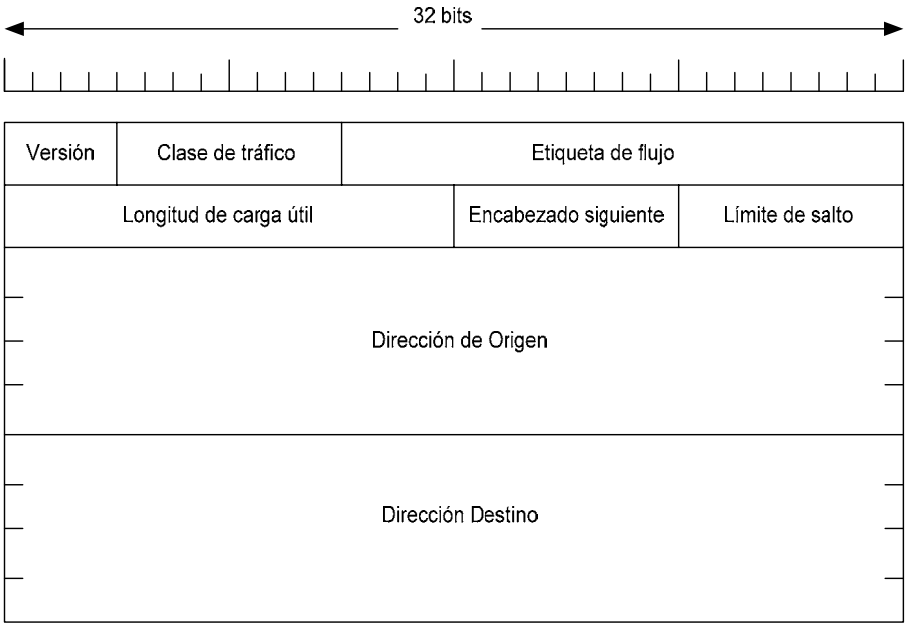


FIGURA 2.16. Formato del encabezado base.

- Longitud de carga útil. Indica cuantos bytes siguen al encabezado base.
- Encabezado siguiente. Indica que encabezado de extensión, de haberlo, sigue a este. Si este encabezado es el último encabezado de IP, indica el manejador de protocolo de transporte al que se entregará el paquete.
- Límite de salto. Disminuye cada vez que se produce un salto.
- Dirección de origen y destino. Indican quien genera el datagrama y a quien va destinado.

2.6.2. Encabezado de extensión

Los encabezados de extensión se utilizan para proporcionar información extra. Algunos encabezados tienen un formato fijo mientras que otros contienen un número variable de campos de longitud variable.

Algunos de los encabezados de extensión de IPv6 son:

- Opciones salto por salto. Contiene información diversa para los encaminadores.
- Opciones de destino. Contiene información adicional para el destino.
- Enrutamiento. Contiene la ruta total o parcial a seguir.
- Fragmentación. Manejo de fragmentos de datagramas.
- Autenticación. Verificación de la identidad del emisor.
- Carga útil de seguridad encriptada. Información sobre el contenido encriptado.

2.7. CONCLUSIONES

La utilización de una arquitectura de protocolos ha permitido que diferentes redes de diferentes fabricantes puedan comunicarse. Si tener que cambiar el *software* de comunicación. Se ha mostrado que principios se utilizan para crear está arquitectura y como se establecen los mecanismo de encapsulado e integridad de datos. Un ejemplo claro de arquitectura es el modelo utilizado en Internet que es el modelo TCP/IP. Este incluye un conjunto de servicios y protocolos que permiten a los usuarios la utilización de aplicaciones tales como: correo electrónico, transferencia de ficheros, etc. A través de un conjunto de redes de diferentes fabricantes.

2.8. BIBLIOGRAFÍA

W. Stallings. Comunicaciones y redes de computadores. Ed. Pearson, 2004.

A.S. Tanenbaum. Redes de ordenador. Ed. Prentice-Hall, 2003.

E.C. Douglas. Redes globales de información con Internet y TCP/IP: Principios básicos, protocolos y arquitectura. Ed. Prentice Hall, 1996.

P. Mariño. Las comunicaciones en la empresa: Normas, redes y servicios. Ed. RA-MA, 2003.

J. García Tomas. Redes para proceso distribuido: Área local, arquitecturas, rendimiento, banda ancha. Ed. RA-MA, 2001.

J. García Tomas. Alta velocidad y calidad de servicio en redes IP. Ed. RA-MA, 2002.

2.9. EVALUACIÓN

2.9.1. Evaluación objetiva

1. El punto de acceso al servicio o SAP es:

   A) Un programa que permite comunicar capas de diferentes modelos

   B) Es una interfaz que define las operaciones y servicios primitivos que pone la capa inferior a disposición de la capa superior.

   C) Es un punto de conexión al medio físico.

   D) Es un protocolo de acceso al medio.

2. ¿Qué capa del modelo OSI define el formato de los datos a transmitir?

   A) La capa de sesión

   B) La capa de aplicación

   C) La capa de presentación

   D) La capa de transporte

3. En el modelo TCP/IP desaparecen las capas

   A) La capa de sesión y la capa física

   B) La capa de enlace.

   C) La capa de enlace y la capa de red

   D) La capa de sesión y la capa de transporte

4. Indique a que protocolo está asociado el puerto 25:

   A) FTP

   B) SMTP

   C) TELNET

   D) HTTP

5. El protocolo IP es:
   A) Orientado a conexión y no confiable
   B) No orientado a conexión y no confiable
   C) No orientado a conexión y confiable
   D) Orientado a conexión y confiable

6. El protocolo TCP es:
   A) Orientado a conexión y no confiable
   B) No orientado a conexión y no confiable
   C) No orientado a conexión y confiable
   D) Orientado a conexión y confiable

7. Las conexiones TCP se identifican:
   A) Mediante los identificadores de *socket* de los dos extremos
   B) Mediante el *socket* destino
   C) Mediante el *socket* destino y el puerto origen
   D) Mediante el puerto destino

8. ¿Cuál de los siguientes protocolos no pertenece a la capa de aplicación del modelo TCP/IP?
   A) HTTP
   B) UDP
   C) DNS
   D) FTP

9. IPv4 e IPv6:
   A) Son incompatibles
   B) No pueden coexistir
   C) Tiene la misma estructura de datagramas
   D) Utilizan direcciones IP de 32 bits.

10. Indique a que protocolo está asociado el puerto 80:

    A) FTP

    B) SMTP

    C) TELNET

    D) HTTP

TEMA 3
MEDIOS Y MODOS DE TRANSMISIÓN

3.1. Introducción y orientación en el estudio
3.2. Objetivos
3.3. Conceptos básicos de los medios físicos
3.4. Los conductores metálicos
3.5. Las fibras ópticas
3.6. Los satélites espaciales de comunicaciones
3.7. La multiplexación y la modulación
3.8. Modulación en banda base y banda ancha
3.9. Aplicaciones de la multiplexación y modulación
3.10. Conclusiones
3.11. Bibliografía
3.12. Evaluación

3.1. INTRODUCCIÓN Y ORIENTACIÓN EN EL ESTUDIO

En este capítulo se analizan en detalle las características de los medios físicos, cables y sistemas inalámbricos, que transportan la información en las redes industriales. Así, se resumen los conceptos básicos que tratan aspectos como la canalización de un medio dentro de su ancho de banda y su relación con el espectro electromagnético de las señales que lo atraviesan. Dado que las redes industriales han de desplegarse de una forma organizada dentro de la ubicación geográfica que cubren, a continuación se estudian los cables metálicos, incluyendo el cableado estructurado y la certificación de los mismos, y posteriormente las fibras ópticas, así como se introducen las comunicaciones por satélite, su configuración, los tipos de órbitas y alguno de los servicios típicos, para finalmente estudiarse algunas aplicaciones de la multiplexación y la modulación en redes de operadores de cable y satélite, y se introducen los Buses de Campo. Se exponen ejemplos reales de redes de comunicación existentes en el mercado de los operadores y fabricantes, cuya comprensión permite asentar el conocimiento adquirido a lo largo del capítulo, y justificar su interés práctico.

3.2. OBJETIVOS

Se trata de familiarizar al alumno con los conceptos básicos que explican las limitaciones que imponen los medios físicos a las señales electromagnéticas que deben transportar, en el ámbito de las comunicaciones industriales. Estas limitaciones pueden ser cuantificadas para valorar adecuadamente lo que un determinado medio físico es capaz de permitir en términos de velocidad, retardo y errores de transmisión. La correcta cuantificación implica un profundo conocimiento de los fenómenos físicos en la interacción de las señales electromagnéticas con la materia, que se va presentando paso a paso. La experiencia práctica adquirida en la instalación de redes con todo tipo de materiales se manifiesta también en las normas existentes, y por ello se van incorporando al conocimiento a medida que se dominan los fundamentos de la transmisión.

3.3. CONCEPTOS BÁSICOS DE LOS MEDIOS FÍSICOS

El medio físico indica el tipo de soporte material que presenta una red para transmitir la información a través de sus rutas, y con los equipos terminales de los usuarios. Estos medios físicos son los siguientes:

- Inalámbricos (radiaciones electromagnéticas):
  - Antenas: emisoras y receptoras en los extremos de la red y en los terminales de usuario.
  - Radioenlaces terrestres: antenas y reemisores en las rutas terrestres de la red.
  - Satélites espaciales: radioenlaces a bordo de un vehículo en el espacio exterior.
- Cables:
  - Metálicos: de pares, cuadretes y coaxiales.
  - Fibra óptica: de sílice, plástico, fluoruros y otros materiales.
  - Híbridos: combinan cables coaxiales y fibras (son los de los "operadores del cable").
- Mixtos: mezcla de todos los medios físicos anteriores en una misma red.

El "ancho de banda" (BW, *Bandwidth*) puede definirse como:

- De una Señal: Porción del espectro electromagnético que ocupa en su propagación a través de un medio de transmisión (por ejemplo, cable, atmósfera, espacio exterior, etc.). Conjunto de frecuencias de su espectro que puede ser medido en un instrumento electrónico denominado Analizador de Espectros (por ejemplo, banda telefónica, banda de audio, banda de TV, etc.).
- De un Medio de Transmisión: Margen del espectro electromagnético que es capaz de transportar para la correcta propagación de las señales de información que lo atraviesan (sin distorsionarlas o deformarlas).

La "canalización" de señales en un medio físico consiste en la ubicación de sus respectivos anchos de banda dentro del ancho de banda del medio (por ejemplo, "plan de frecuencias" de una red de cable). La porción de espectro ocupado por cada señal dentro del margen total del correspondiente al medio

representa un "canal" de transmisión, y por ello hay tantos canales como señales en el medio (por ejemplo, canales de TV y radiodifusión).

Una señal en "banda base" (BB, *Baseband*) es aquella que se transmite dentro de la misma porción del espectro (ancho de banda) en que es generada o reconstruida (distinción con IF y RF en emisores y receptores).

Una señal de "espectro ampliado" (SS, *Spread Spectrum*) es aquella se transmite con mayor porción del espectro del que tiene cuando es generada o reconstruida, y por ello presenta un margen de frecuencias mayor que su ancho de banda original.

Una señal de "ultra banda ancha" (UWB, *Ultra Wide Band*) es la que se transmite con un ancho de banda superior al 20% de la frecuencia de su portadora (por ejemplo, con una portadora de 6 GHz su ancho de banda es 1,2 GHz), o con un ancho de banda superior a 500 MHz (lo que sea mayor).

Los servicios de comunicación en banda ancha son aquellos que proporcionan una velocidad de transferencia de información digital superior a los 2 Mbps. Son prestados a través de las redes multimedia denominadas popularmente "autopistas de la información".

El "factor de mérito ancho de banda x distancia" se define para los cables, sean metálicos o de fibra óptica, y en las redes digitales se expresa por el producto "velocidad (bps) x distancia (km)", que indica la velocidad máxima que puede alcanzar la señal digital a lo largo de 1 km de cable sin necesidad de regenerarla mediante repetidores. Cuanto mayor sea el producto, mayor es la calidad del cable para la transmisión de señales digitales.

La "ecuación de Shannon" relaciona la velocidad de la información digital en un medio (C en bps) con el ancho de banda del canal (BW en Hz) y la relación Señal/Ruido en el mismo (S/N como relación de potencias en Watios), según la expresión 3.1.

$$C = BW.\log_2 (1 + S/N) \qquad (3.1)$$

3.3.1. El espectro electromagnético

Una de las fuentes de energía de la naturaleza es la producida por ondas electromagnéticas, que se propagan a través de la materia con diferentes velocidades. Cuando la materia es el espacio libre, al fenómeno de propagación

de las ondas electromagnéticas por este medio se denomina radiación. Por esta razón, la energía que se propaga en el espacio libre mediante ondas electromagnéticas recibe el nombre de "radiación electromagnética".

Las emisiones de programas de radio y televisión (TV) son posibles, porque las radiaciones electromagnéticas que salen de sus antenas se propagan por el espacio libre hasta alcanzar a las antenas de los receptores de radio y televisión correspondientes. Por lo tanto, las radiaciones electromagnéticas de radio y TV "enlazan", de forma invisible, a un emisor y a un receptor a través de sus antenas, produciendo una transferencia de información sin cables o inalámbrica.

Las ondas electromagnéticas deben su denominación de "ondas", al hecho de que durante su avance modifican su forma de una manera semejante a las olas del mar, pues éstas se manifiestan ante el observador como ondulaciones de la superficie del agua.

El término electromagnético es debido a que la energía de esas ondas, a diferencia de las olas del mar, procede de cargas eléctricas estáticas y en movimiento, y se manifiesta en la presencia de "campos" eléctricos y magnéticos. La idea de "campo" sugiere precisamente la creación de un área de influencia de esa energía sobre un medio físico, de tal manera que las cargas eléctricas de ese medio físico perciben el efecto de esos "campos". Eso es lo que ocurre en una antena receptora de ondas electromagnéticas, pues sus propias cargas eléctricas perciben la acción a distancia de esos "campos" generados en la antena emisora, y a ese fenómeno de percepción se le denomina "recibir la onda".

Lo característico de una onda electromagnética es precisamente su capacidad para cambiar su forma en el transcurso del tiempo. Para su estudio teórico se considera a una onda electromagnética de forma senoidal que cambia con el tiempo periódicamente, y al número de veces que la onda senoidal periódica se repite a sí misma en el tiempo se denomina "frecuencia". Los valores que toma la señal senoidal en su evolución en el tiempo se denominan "amplitud" y, debido a la forma de la señal, la amplitud puede ser determinada matemáticamente mediante la expresión trigonométrica de la función seno. Precisamente por ser obtenida de esta función trigonométrica, esta onda recibe el nombre de senoidal.

Si la señal periódica senoidal realiza una repetición o ciclo cada segundo, se obtiene la unidad de frecuencia denominada Hercio (Hz), que representa un ciclo de cambio de la onda senoidal en un segundo y por ello equivale a un ciclo por segundo (Hz = ciclo/s). El "período" de la onda senoidal es el valor inverso

de la frecuencia, y si la frecuencia se expresa en Hz el período se mide en segundos. Por ejemplo, a una onda electromagnética senoidal de 100 Hz le corresponde un período de una centésima de segundo (0,01 s = 1/100 Hz).

Para establecer la relación entre la velocidad de propagación de la onda electromagnética a través de la materia y su frecuencia, se emplea el concepto de longitud de onda (λ). En el vacío las ondas electromagnéticas alcanzan su máxima velocidad (c), que es la denominada "velocidad de la luz" y que equivale a 300.000 kilómetros por segundo. La longitud de onda (λ) representa físicamente el camino recorrido por una onda electromagnética senoidal, en su propagación por el vacío, durante el cambio de un período. Esto equivale a considerar que λ es el cociente entre la velocidad de la luz (c) y la frecuencia (f). Teniendo en cuenta que todas las ondas electromagnéticas se propagan en el vacío a una velocidad constante, que es la velocidad de la luz, existe una relación inversa entre su longitud de onda y su frecuencia ($\lambda = c / f$).

La identificación de una onda electromagnética senoidal por su frecuencia y su longitud de onda asociada, permite establecer una clasificación de las radiaciones electromagnéticas existentes en la naturaleza, en grupos de frecuencias, o "bandas de frecuencia", que tienen una gran importancia en las comunicaciones, pues delimitan los tipos de radiaciones que pueden ser utilizadas en aplicaciones de comunicación concretas.

Para facilitar el cálculo de las longitudes de onda asociadas a las bandas de frecuencia, éstas se han delimitado en márgenes de 0,3 Hz a 3 Hz con un factor multiplicativo que es una potencia entera del número 10 (10 elevado a "n"), siendo el exponente entero "n" el número de la banda. Por ejemplo, la "banda 4" equivale al margen de frecuencias entre 3 kHz y 30 kHz, pues 0,3 Hz por 10.000 es igual a 3000 Hz (3 kHz), y 3 Hz por 10.000 resulta ser 30 kHz (30.000 Hz).

La presencia del número 3 está relacionada con el valor de la velocidad de la luz, ya que las longitudes de onda (λ) asociadas a cada banda, que resultan del cociente "c/f", son iguales a una unidad de longitud entera. En el ejemplo de la "banda 4", el margen de longitudes de onda que le corresponden está entre 100 km (c/3 kHz) y 10 km (c/30 kHz).

Los servicios de radiocomunicación hasta la "banda óptica" se clasifican en once (n = 11) bandas de frecuencia. A partir de la banda óptica las radiaciones electromagnéticas no tienen aplicación en las comunicaciones inalámbricas

comerciales, y su clasificación en bandas recibe otras denominaciones distintas a las anteriores.

Cuando la clasificación de las radiaciones electromagnéticas se hace por grupos de longitudes de onda, en vez de por grupos de frecuencias, se denomina "espectro electromagnético". En la terminología de las comunicaciones los términos de "banda" o de "espectro" se emplean como sinónimos, y por ello el conjunto de todas las bandas de frecuencia que existen en la naturaleza suele denominarse "espectro electromagnético".

El uso controlado del espectro electromagnético de la naturaleza tiene un carácter estratégico en la prestación de los servicios de radiocomunicación, en el sentido de que cuando una onda electromagnética viaja por el espacio libre, a una frecuencia dada, dentro de una determinada banda, cualquier otra onda que se propague a la misma frecuencia por el mismo espacio producirá una perturbación sobre aquélla cuyo efecto se denomina "interferencia". Este efecto interferente puede debilitar, alterar o hacer desaparecer completamente a la señal, con lo cual la comunicación es nula o ininteligible en el receptor.

Para eliminar el problema de la interferencia entre diferentes servicios de comunicación, se establecen asignaciones de frecuencias dentro de las bandas, de tal manera que una vez realizada la asignación de las frecuencias disponibles, no se pueden utilizar otras frecuencias. En los receptores de un determinado servicio de comunicación, al procedimiento de seleccionar la frecuencia de radiación de un determinado emisor o proveedor del servicio, se denomina "sintonización".

Efectivamente, cuando un usuario de un servicio de radiocomunicaciones desea recibir la prestación del servicio, actúa sobre el sintonizador de su receptor mediante un telemando o una botonera, para seleccionar la frecuencia de radiación del emisor del servicio, como ocurre en el caso de la audición de un programa de radio o de la contemplación de un programa de TV. El concepto de sintonía es precisamente el de poner de acuerdo a varias entidades para su relación amistosa. Así, una vez sintonizados emisor y receptor a la misma frecuencia es posible una comunicación correcta.

Para el conocimiento detallado de las bandas de comunicación existentes en el espectro electromagnético, se presentan a continuación las unidades de frecuencia y longitud de onda, así como la nomenclatura de las bandas de frecuencia.

3.3.1.1. Unidades y nomenclatura

Las unidades de frecuencia utilizadas son múltiplos del Hercio:

- Hercio (Hz): equivale a 1 ciclo por segundo (ciclo/s)
- Kilohercio (kHz): equivale a mil ciclos por segundo (1000 Hz)
- Megahercio (MHz): equivale a un millón de ciclos por segundo (1.000.000 Hz o 10^6 Hz)
- Gigahercio (GHz): equivale a mil millones de Hz (1.000.000.000 Hz o 10^9 Hz)
- Terahercio (THz): equivale a un billón de Hz (10^{12} Hz)
- Petahercio (PHz): equivale a mil billones de Hz (10^{15} Hz)

Las unidades de longitud de onda (λ)son submúltiplos del milímetro:

- milímetro (mm): milésima parte del metro (10^{-3} m)
- micra (μm): milésima parte del milímetro (10^{-6} m)
- nanómetro (nm): millonésima parte del milímetro (10^{-9} m)
- picometro (pm): mil millonésima parte del milímetro (10^{-12} m)
- femtometro (fm): billonésima parte del milímetro (10^{-15} m)

La nomenclatura de las bandas de frecuencia del espectro electromagnético hasta la "banda 11" procede de la terminología anglosajona, y es la siguiente:

- F (*frequency*): frecuencia
- L (*low*): baja
- M (*medium*): media
- H (*high*): alta
- V (*very*): muy
- U (*ultra*): ultra
- S (*super*): super
- E (extra): extra

3.3.2. Las bandas de comunicación y sus aplicaciones

En la presentación de las aplicaciones de las bandas de comunicación del espectro electromagnético, se utiliza una terminología que emplea agrupaciones por márgenes de longitud de onda (λ) en vez de márgenes de frecuencia. Dentro de cada margen de las λ (pronunciado lambda), se mencionan las bandas de frecuencia que abarca. Los márgenes de las λ establecidos en comunicaciones son los siguientes: la Radiofrecuencia, las Microondas y la Banda Óptica.

Seguidamente se analizan sus particularidades y aplicaciones.

3.3.2.1. Radiofrecuencia

El espectro de radiofrecuencia, también denominado de ondas de radio o de RF, comprende un margen de λ desde los 10.000 km a los 10 cm. Las bandas de frecuencia involucradas (desde la 2 hasta la 9) y sus nomenclaturas son:

- ELF (Extremadamente bajas frecuencias): 10.000 km a 1000 km (30 Hz a 300 Hz)
- VF (Frecuencias de voz, banda vocal): 1000 km a 100 km (300 Hz a 3000 Hz)
- VLF (Muy bajas frecuencias): 100 km a 10 km (3 kHz a 30 kHz)
- LF (Bajas frecuencias): 10 km a 1 km (30 kHz a 300 kHz)
- MF(Frecuencias medias): 1000 m a 100 m (300 kHz a 3000 kHz)
- HF (Altas frecuencias): 100 m a 10 m (3 a 30 MHz)
- VHF (Muy altas frecuencias): 10 m a 1 m (30 MHz a 300 MHz)
- UHF (Ultra altas frecuencias): 10 dm a 1 dm (300 MHz a 3000 MHz)

Seguidamente se describen las aplicaciones más destacadas de estas bandas de frecuencia (Tabla 3.1).

- La banda 1 (3 Hz a 30 Hz) no se considera dentro de la RF, y por ello no recibe una nomenclatura propia. No obstante las señales eléctricas que se utilizan en los circuitos electrónicos pueden trabajar a estas frecuencias, y de hecho se emplean en aplicaciones de laboratorio. En la navegación submarina, como señales eléctricas, se utilizan en los

equipos electrónicos de Sónar para generar infrasonidos con fines de vigilancia, determinando la posición de objetos bajo el agua y calculando su distancia. En la naturaleza se producen espontáneamente ondas elásticas o de presión en esta banda, por la actividad de los volcanes y terremotos en la superficie de la tierra.

- La ELF (banda 2) se utiliza en comunicaciones para navegación de superficie. En esta banda están también las señales eléctricas de los circuitos electrónicos. Una señal eléctrica muy conocida en esta banda, es la de suministro de energía eléctrica, que en los países europeos tiene una frecuencia de 50 Hz mientras que en Estados Unidos es de 60 Hz. Esta señal eléctrica es la que se aplica a la fuente de alimentación de equipos electrónicos como el PC.
- El nombre de la VF (banda 3), que también se denomina "banda vocal", proviene de que esta banda se utiliza también en las señales eléctricas de telefonía, es decir, las que circulan por los cables telefónicos.
- La VLF (banda 4) se utiliza en ondas largas de radio, cuya propagación se realiza sobre la superficie terrestre. No obstante, a partir de los 20 kHz puede ser utilizada en señales eléctricas, para generar ultrasonidos con diversos fines, como: la detección de intrusos, la limpieza de piezas mecánicas, aplicaciones en medicina, etc.

  Las señales eléctricas de sonido, o "audio", están en el margen entre 20 Hz y 20 kHz. Su nombre se debe al hecho de que es en este margen donde las ondas elásticas, o de presión, son captadas por el oído humano, percibiéndolas como sonidos o estímulos auditivos.

Dado que las señales eléctricas de los circuitos electrónicos de equipos de sonido, o de audio (sintonizadores, amplificadores, magnetofones, lectores de discos compactos, mesas mezcladoras de sonidos, etc.), pertenecen a dicho margen, tanto la VF como la VLF y ELF soportan estas aplicaciones.

Es importante enfatizar que todas las ondas de presión que están fuera del margen de audio no se consideran propiamente sonidos porque son inaudibles para el oído humano. De ahí la denominación de infrasonidos para las ondas elásticas que están por debajo de los 20 Hz (banda 1), y de ultrasonidos para las que superan los 20 kHz (VLF).

- La LF (banda 5) tiene las mismas aplicaciones en radiocomunicaciones y con señales eléctricas que la VLF, salvo las aplicaciones de audio. También se aplican en sistemas de ayuda a la navegación y en

radiofaros. Hasta la banda LF las señales eléctricas pueden ser transportadas por cables metálicos de pares y de cuadretes. Los cables de pares constituyen el medio de transmisión guiado más económico. Los cables de cuadretes son conjuntos de cables de pares agrupados de dos en dos para transmitir un elevado número de señales eléctricas a grandes distancias.

- La MF (banda 6) se utiliza en radiodifusión por modulación de amplitud (AM), en defensa civil y en onda pesquera. A las radiaciones en este margen de frecuencias se les suele denominar ondas medias de radio, de ahí el nombre de la banda. Esta banda es utilizada en forma de señales eléctricas, en equipos electrónicos de tecnologías de acceso a servicios interactivos mediante la red telefónica conmutada (RTC), denominadas de "lazo digital de abonado" o "xDSL", la más conocida de las cuales es ADSL (desde 25 kHz en VLF a 1,1 MHz en MF).

- La HF (banda 7) utiliza las ondas cortas de radio. Se aplica a la radiotelefonía móvil, y debido a su capacidad de penetración a través de la atmósfera pueden alcanzar las capas de la "ionosfera".

  La ionosfera es la región que envuelve a la Tierra a distancias de la superficie terrestre entre 60 km y 1000 km de altitud. La ionosfera recibe su nombre del efecto ionizante que producen las radiaciones procedentes del sol, fundamentalmente los rayos X (margen de λ entre 10 nm y 10 pm) y las radiaciones ultravioleta (margen de λ entre 380 nm y 10 nm), sobre las moléculas y átomos de las capas altas de la atmósfera. El efecto ionizante convierte a cada molécula y a cada átomo de la atmósfera, que son eléctricamente neutros, en dos tipos de partículas eléctricas: una partícula de carga positiva o ion positivo, y una partícula de carga negativa o electrón libre.

  Las reflexiones de las ondas cortas sobre las capas de la ionosfera (denominadas D, E y F de menor a mayor altitud) permiten enlazar puntos sobre la superficie terrestre muy distantes entre sí, incluso los situados en las antípodas, entre los que no hay visión directa. Debido a esta capacidad de alcanzar grandes distancias sobre la superficie de la tierra propagándose por reflexión, las ondas cortas de radio son empleadas por los radioaficionados y en radiocomunicaciones militares.

  Esta banda es utilizada en forma de señales eléctricas, para generar ultrasonidos en equipos de electromedicina de ecografía, en aplicaciones de ginecología, obstetricia y ecocardiografía.

TABLA 3.1. Bandas de frecuencias y aplicaciones

Nº	Nombre	Margen	Medios	Aplicaciones
3	VF Banda vocal	300 Hz – 3 kHz	• Cables de pares	• Telefonía y audio • Datos (ETD y ETCD) • Navegación submarina (Sónar) • Navegación de superficie
4	VLF Muy baja frecuencia	3 – 30 kHz	• Cables de pares • Radio: Onda larga	
5	LF Baja frecuencia	30 – 300 kHz	• Cables de pares • Radio: Onda larga	• Ultrasonidos • Navegación: Ayudas y radiofaros • Com. industriales
6	MF Frecuencia media	300 kHz – 3 MHz	• Cable coaxial • Radio: Onda media	• Radiodifusión (AM) • Defensa civil • Onda pesquera
7	HF Alta frecuencia	3 – 30 MHz	• Cable coaxial • Radio: Onda corta	• Radiotelefonía móvil • Radioaficionados • Comunicaciones militares (ionosfera)
8	VHF Muy alta frecuencia	30 – 300 MHz	• Cable coaxial • Radio: Ondas métricas	• Televisión en VHF • Radiodifusión (FM) • Control de tráfico aéreo • Com. móviles
9	UHF Ultra alta frecuencia	300 MHz – 3 GHz	• Guía de onda • Radio: Ondas decimétricas	• Televisión en UHF • Radiolocalización (Radar) • Telemetría espacial • Telefonía móvil aut.
10	SHF Super alta frecuencia	3 – 30 GHz	• Guía de onda • Microondas	• Comunicaciones espaciales (satélites) • Radioenlaces de microondas • Radiolocalización (Radar)
11	EHF Extremadamente alta frecuencia	30 – 300 GHz	• Guía de onda • Microondas	• Radioastronomía • Com. ferroviarias • Sist. Experimentales
12 a 16	Infrarrojo Luz visible Ultravioleta	$0,3 – 4.10^{14}$ Hz $4 – 7,7.10^{14}$ Hz $7,7.10^{14} – 3.10^{16}$ Hz	• Fibras ópticas • Rayos láser	• Comunicaciones ópticas

- En la VHF (banda 8), también denominada de ondas métricas de radio, a diferencia de la HF sus radiaciones no alcanzan la ionosfera, y por ello necesitan la visión directa entre el emisor y el receptor. Se emplea en comunicaciones de corta distancia como: radiodifusión en frecuencia modulada (FM), canales bajos y altos de TV, control de tráfico aéreo y comunicaciones móviles.

  Entre la VHF y la MF, ambas incluidas, las señales electromagnéticas guiadas son transportadas por cables metálicos coaxiales. Su denominación procede de la forma constructiva de estos cables, debido a que están formados por un cilindro metálico en cuyo interior se coloca un conductor en posición axial, y ambos están separados por un aislante.

- La UHF (banda 9), también denominada de ondas decimétricas de radio, presenta características de propagación similares a las de VHF. Sus aplicaciones típicas son: canales altos de TV, radioenlaces terrestres en UHF, telefonía móvil celular (NMT, GSM y DCS), radiolocalización (RADAR), telemetría espacial como en servicios satelitales de navegación global (GNSS), comunicaciones terrestres mediante redes fijas de acceso inalámbrico en banda ancha MMDS y redes de área local inalámbricas (WLAN, *Wireless Local Area Network*) en la banda ISM (Industria, Ciencia y Medicina) sin licencia.

En las bandas de VHF y UHF la presencia de diferentes servicios introduce las siguientes subdivisiones:

- Banda I (40 MHz a 68 MHz): canales bajos de TV (1 a 4) en VHF
- Banda II (80 MHz a 108 MHz): radiodifusión en FM (VHF)
- Banda III (162 MHz a 230 MHz): canales altos de TV (5 a 11) en VHF
- Banda IV (470 MHz a 590 MHz): canales 21 a 35 de TV (UHF)
- Banda V (590 MHz a 960 MHz): canales 36 a 69 de TV, telefonía móvil y redes locales WLAN
- Banda L (1 GHz a 2 GHz): radioenlaces, telemetría (GNSS, *Global Navigation Satellite Services*) y telefonía móvil celular DCS

Hay que señalar que no todos los servicios mencionados en UHF utilizan en su ámbito de aplicación concreto las denominaciones especificadas para cada subbanda, no obstante el interés de relacionarlos reside en destacar las semejanzas de los márgenes de frecuencia utilizados por cada uno de ellos.

En general los “radioenlaces” son sistemas de transmisión que propagan las radiaciones electromagnéticas entre tramos de visión directa, en comunicación punto a punto, cuando hay que enviarlas a distancias largas, y están compuestos por antenas y reemisores.

La denominación de comunicación “punto a punto” utilizada en los radioenlaces, enfatiza su diferencia respecto de la radiodifusión debido a que ésta es una comunicación del tipo “punto a multipunto”, es decir, hay una antena emisora y múltiples antenas receptoras. En los radioenlaces hay solamente una antena emisora y una única antena receptora en cada tramo de visión directa, de ahí la expresión “punto a punto”.

Los servicios satelitales de navegación global (GNSS) utilizan una constelación de 2 a 3 decenas de satélites en órbitas medias (23.200 km de altitud) situadas en tres planos orbitales, realizando cálculos de telemetría con radiaciones electromagnéticas en la Banda L. La palabra “navegación” proviene que estos servicios permiten determinar la posición (longitud, latitud y altitud), hora y velocidad de vehículos terrestres o de naves aéreas y marítimas, en cualquier punto del planeta.

Para recibir este servicio los vehículos y naves están dotados de un equipo electrónico emisor-receptor GNSS, que se conecta directamente con dos o tres satélites de la constelación. Un sistema GNSS muy conocido es el GPS. Las aplicaciones de los GNSS están destinadas a: sistemas de transporte por tierra, mar y aire, sistemas de comunicaciones móviles, servicios de emergencias para salvamento y otros.

En telefonía móvil celular, dentro de UHF, se emplearon frecuencias próximas a los 450 MHz en los primeros sistemas analógicos (NMT), mientras que en los sistemas digitales se emplean otras próximas a 900 MHz (GSM) y 1800 MHz (DCS).

Cabe destacar que la capacidad de las radiaciones electromagnéticas próximas a los 900 MHz, de producir calentamiento en los tejidos biológicos y todo tipo de alimentos, ha llevado a la fabricación de los populares “hornos de microondas”.

Este fenómeno físico desató la polémica de si los teléfonos móviles celulares, que trabajan en frecuencias próximas, no provocarían un efecto similar peligrosamente inadvertido por los usuarios de los teléfonos móviles, que a la larga pudiesen afectar a tejidos próximos a los oídos causando mutaciones genéticas capaces de desarrollar metástasis cancerosas.

Los estudios de carcinogénesis, tanto epidemiológicos sobre estadísticas de enfermos, como las pruebas normalizadas en experimentos de laboratorio con animales, no han sido concluyentes. La Comisión Federal de Comunicaciones (FCC) en USA recomienda un límite de exposición máximo de 1,6 Watios de potencia por kg de masa biológica. Teóricamente los teléfonos móviles no superan potencias de radiación superiores a: 600 milésimas de Watio (600 mW) en los analógicos, y 125 mW en los digitales.

Las redes fijas de acceso inalámbrico en banda ancha, para sistemas de distribución multicanal multipunto (MMDS) utilizan radiofrecuencias en el margen de 2,5 a 2,69 GHz (UHF). Las redes de área local inalámbricas (WLAN) en la banda ISM, por sus características técnicas de realización y sus condiciones de funcionamiento, no necesitan licencia para operar. La banda ISM fue liberada por la FCC en 1985, y está formada por los siguientes márgenes de frecuencia: 902 a 928 MHz (UHF), 2,4 a 2,483 GHz (UHF) que es el más utilizado, y 5,725 a 5,875 GHz (SHF) en la banda de las microondas.

Es interesante señalar que las redes WLAN permiten la movilidad de los usuarios, dentro de un área de cobertura de unos 300 m, y sus velocidades de transmisión están en el límite de los 10 Mbps. Las redes fijas de acceso inalámbrico en banda ancha MMDS y LMDS (sistemas de distribución local multipunto), no permiten la movilidad de los usuarios, aunque presentan mayores áreas de cobertura que las WLAN y sus velocidades están por encima de los 20 Mbps, de ahí su denominación de “banda ancha”.

3.3.2.2 Microondas

El espectro de las microondas comprende un margen de λ desde los 10 cm a 1 mm. Las bandas de frecuencia involucradas (10 y 11) y sus nomenclaturas son:

- SHF (Super altas frecuencias): 10 cm a 1 cm (3 GHz a 30 GHz)
- EHF (Extremadamente altas frecuencias): 10 mm a 1 mm (30 GHz a 300 GHz)

Las aplicaciones más destacadas de estas bandas de frecuencia son:

- La SHF (banda 10), también denominada de ondas centimétricas, presenta las siguientes aplicaciones: investigación del espacio exterior, comunicaciones espaciales por satélite de servicios fijos (FSS) y móviles (MSS), radioenlaces terrestres de microondas, radiodifusión por

satélite (BSS), servicios de meteorología, teledetección espacial para investigación de recursos terrestres, control de tráfico en carreteras (cinamómetros basados en RADAR), alarmas contra intrusos en edificios, comunicaciones terrestres mediante redes fijas de acceso inalámbrico en banda ancha LMDS, y redes de área local inalámbricas (WLAN) en la banda ISM sin licencia.

- La EHF (banda 11), también denominada de ondas milimétricas, presenta las siguientes aplicaciones: comunicaciones espaciales por satélite, radioastronomía, comunicaciones ferroviarias, radioenlaces terrestres de microondas, y sistemas experimentales.

En las bandas de SHF y EHF la presencia de los diferentes servicios mencionados, tales como los satelitales, de radioenlaces terrestres de microondas, de redes inalámbricas y de RADAR introduce las siguientes subdivisiones:

- Banda S (2 GHz a 4 GHz): investigación del espacio exterior, redes MMDS y WLAN en UHF
- Banda C (4 GHz a 8 GHz): redes locales inalámbricas (WLAN), radioenlaces terrestres y FSS
- Banda X (8 GHz a 12 GHz): teledetección, meteorología y control en carreteras (RADAR)
- Banda Ku (12 GHz a 18 GHz): radiodifusión por satélite (BSS), radioenlaces terrestres y FSS
- Banda K (18 GHz a 27 GHz): radioenlaces terrestres de microondas, BSS y FSS
- Banda Ka (27 GHz a 40 GHz): radioenlaces terrestres, acceso inalámbrico LMDS y FSS
- Banda V (40 GHz a 50 GHz): experimental

También conviene señalar aquí que no todos los servicios mencionados utilizan en su ámbito de aplicación concreto las denominaciones especificadas para cada subbanda y, como se mencionó en la banda de UHF de radiofrecuencia (RF), el interés de relacionarlos reside principalmente en destacar las semejanzas en los márgenes de frecuencias utilizados por cada uno de ellos.

Las comunicaciones terrestres mediante redes fijas de acceso inalámbrico en banda ancha para LMDS, utilizan las microondas en el margen de 27,5 a 29,5 GHz. Entre las microondas y la UHF las señales electromagnéticas guiadas son transportadas por tubos metálicos de sección rectangular o circular denominados "guías de onda". Se utilizan en equipos de radioenlaces para conectar los equipos emisores y receptores a la antena correspondiente.

3.3.2.3. Banda óptica

El espectro de la banda óptica tiene un margen de λ desde los 10 nm a 1 mm (las 5 "décadas" de la óptica). A su vez, los submárgenes espectrales en que se divide son los siguientes:

- Infrarrojos: 1 mm a 760 nm
- Luz visible: 760 nm a 380 nm (la "octava" de la visión)
- Ultravioleta: 380 nm a 10 nm

La denominación de banda óptica se realiza al encontrase en ese margen espectral todas las radiaciones procedentes del Sol que tienen interés en la óptica. El espectro visible recibe su denominación por el hecho de el ojo humano es sensible a sus radiaciones, siendo ciego para las radiaciones luminosas del infrarrojo y del ultravioleta.

Los nombres de los submárgenes parten del color más próximo del espectro visible. Dado que el espectro visible está comprendido, de mayor a menor λ, entre la λ del color rojo (760 nm) y la λ del color violeta (380 nm), a todas las radiaciones cuyas frecuencias son inferiores a la del rojo (λ mayor que 760 nm hasta 1 mm) se denominan infrarrojos, mientras que las frecuencias que están por encima de la del violeta (λ menor que 380 nm hasta 10 nm) reciben el nombre de ultravioleta. Para recodar con facilidad el margen espectral de la banda óptica, se indica que ocupa 5 décadas porque la relación entre las λ del margen (1 mm/0,00001 mm) es una potencia entera del número 10 (década), y el valor del exponente es el número 5 (10 elevado a 5).

De la misma forma, para memorizar fácilmente el margen espectral de la luz visible, se indica que ocupa 1 octava porque la relación entre las λ del margen (760 nm/380 nm) es una potencia entera del número 2 (octava), y el valor del exponente es el número 1 (2 elevado a 1).

El nombre de “octava” para referirse a las potencias enteras del número 2, procede de la relación de frecuencias de las ondas sonoras producidas por los instrumentos de música, con objeto de facilitar las comparaciones entre notas musicales.

Seguidamente se analizan las particularidades y aplicaciones de estos márgenes:

- Las radiaciones infrarrojas son muy utilizadas en comunicaciones ópticas, debido a que en este margen las redes de fibra óptica (apartado 3.5) permiten las velocidades más elevadas de transferencia de información digital, superiores al Terabit por segundo (Tbps que equivale a un billón de bits por segundo), a lo largo de grandes distancias (miles de km). Una fibra óptica se comporta como una tubería de luz, y su forma es la de un cilindro de material dieléctrico (sílice, plástico, fluoruros, etc.), aproximadamente del espesor de un cabello humano en las más finas, con una cubierta.

  Una red de fibra óptica consta de un dispositivo óptico emisor, que es un diodo electroluminiscente tipo láser (LD) que se conecta en un extremo de la fibra, y el receptor óptico es un dispositivo detector de radiación tipo fotodiodo que se conecta en el extremo opuesto. Para amplificar la luz entre los extremos emisor y receptor de la fibra, se utilizan amplificadores ópticos de fibra contaminada con Erbio (EDFA).

  El diodo láser introduce impulsos de radiación infrarroja, bajo la acción de una corriente eléctrica, en el extremo de entrada a la fibra, y el fotodiodo los convierte a su vez a corriente eléctrica en el extremo de salida de la fibra. Estos impulsos de radiación infrarroja constituyen los bits de información digital, que son transportados a través de la red de fibra.

  Existen aplicaciones de comunicaciones de infrarrojos a través del espacio libre para dar lugar a redes de área local inalámbricas (WLAN).

  También se pueden utilizar como enlaces de comunicación punto a punto, de la misma forma que los radioenlaces de RF y microondas, aunque presentan el inconveniente de las perturbaciones del medio. Son empleadas en el espacio exterior para la comunicación entre satélites, con enlaces punto a punto.

Una gran ventaja de los infrarrojos es que existen dispositivos ópticos emisores más económicos que los láseres (LD), que son los diodos electroluminiscentes tipo LED, y fotodiodos con muy baja potencia de consumo. Por esta razón son muy populares los telemandos de infrarrojos, para controlar a distancias cortas (5 m) equipos de electrónica de consumo tales como: televisores, vídeos, cadenas de alta fidelidad, lectores de DVD, etc. Todos los dispositivos ópticos, ya sean electroluminiscentes (LD y LED) o fotodiodos se realizan con materiales optoelectrónicos semiconductores.

Asimismo, los PC llevan un "puerto de infrarrojos", para conectar periféricos sin cables tales como el teclado y el ratón. En estas aplicaciones el puerto de infrarrojos es un fotodiodo situado en el PC, que capta la radiación infrarroja emitida por el dispositivo electroluminiscente situado en el periférico concreto. Lo mismo sucede en el caso del telemando, que es quien lleva el dispositivo electroluminiscente, mientras que el equipo controlado lleva el fotodiodo. Desde el punto de vista de los peligros biológicos de las radiaciones infrarrojas, se considera que los diodos láser (LD) pueden dañar los ojos para valores de λ por debajo de los 1400 nm.

Las ventajas de los diodos láser (LD) frente a los diodos LED son: mayor concentración de la radiación formando un haz muy fino, pudiendo recorrer grandes distancias sin perder el enfoque, gran estabilidad en la λ de la radiación, y mayor potencia de radiación que le permite ser utilizada como herramienta de corte de materiales de gran dureza.

La diferencia principal entre el diodo láser (LD) y el diodo LED es que el primero es un dispositivo de emisión "estimulada" por tener un elemento de bombeo de radiación, de ahí el nombre de láser, mientras el segundo es de emisión "espontánea", es decir, no es estimulado por ningún elemento de bombeo de radiación.

Además de las aplicaciones de los infrarrojos en comunicaciones ópticas, también son utilizados en equipos electrónicos de "visión nocturna". Las "cámaras de infrarrojos" realizadas con estos equipos electrónicos, detectan las radiaciones infrarrojas que emiten los cuerpos según su temperatura, formando áreas de igual temperatura que convenientemente coloreadas producen imágenes en el espectro visible, que serían imposibles de percibir directamente por el ojo humano.

- Dentro del espectro visible, las comunicaciones ópticas se emplean en redes muy económicas y a distancias cortas, a través de fibras de materiales plásticos que son más gruesas que las de sílice (del tamaño de un hilo telefónico).

  En aplicaciones de medicina los diodos láser (LD) en el espectro visible permiten enfocar un haz muy fino de luz de elevada potencia en regiones muy pequeñas del cuerpo humano, lo que posibilita intervenciones quirúrgicas incruentas y muy precisas, y por ello se utilizan en la cirugía oftálmica y otras.

- Las radiaciones ultravioleta no se utilizan en comunicaciones. No obstante, son empleadas en situaciones donde se necesite detectar cuerpos de tamaño próximo al margen de las λ, de la misma manera que en las aplicaciones de RADAR de RF y microondas. Debido a la utilización de radiaciones en la banda óptica estos equipos se denominan LIDAR, y se emplean entre otros usos en medidas de densidad para aplicaciones meteorológicas tales como: nubes, lluvia, niebla, turbulencias, etc.

  Desde el punto de vista biológico las radiaciones ultravioleta presentan efectos ionizantes que pueden afectar severamente a la integridad de los tejidos humanos, y por ello su manejo en el espacio libre tiene dificultades de operación para que no entren en contacto con las personas.

  Las aplicaciones mencionadas de RADAR, LIDAR, SÓNAR y ecografía no son propiamente de comunicaciones, pero tienen interés desde el punto de vista de las tecnologías de la información, en el sentido de que permiten obtener una representación gráfica de la disposición de objetos o cuerpos en un área del espacio determinada, e incluso su velocidad.

  En todas ellas el principio de funcionamiento es el mismo: unas ondas, ya sean electromagnéticas o elásticas, en un determinado margen de frecuencias, se emiten como impulsos en un medio material, y se mide el tiempo que tardan en regresar esos impulsos (ecos) al receptor, después de rebotar sobre los objetos o cuerpos que existen en ese medio, ya sean estáticos o móviles.

La medida del tiempo que tarda en regresar cada eco permite determinar la distancia del obstáculo. Si además los impulsos se orientan en todas las direcciones del espacio, se consigue determinar la posición relativa del obstáculo respecto del emisor de impulsos. Cuando el obstáculo se mueve hay un cambio en la frecuencia del eco respecto de la del impulso (efecto Doppler), y es también posible determinar la velocidad del obstáculo calculando la diferencia de frecuencias entre el impulso del emisor y su correspondiente eco en el receptor. El adecuado tratamiento gráfico de los cálculos de distancia, posición y velocidad permiten representar a los objetos sobre una pantalla para su visualización.

En los casos del RADAR y del LIDAR los impulsos son radiaciones electromagnéticas, mientras que en el SÓNAR y los equipos de ecografía los impulsos son ondas elásticas o de presión.

3.4. LOS CONDUCTORES METÁLICOS

3.4.1. Características

Un "cable de pares" está formado por una pareja de conductores metálicos separados por material aislante. Para eliminar la interferencia entre ellos (diafonía) se retuercen ambos conductores uno sobre otro formando un par trenzado (*Twisted Pair*) (Fig. 3.1). En ambientes con fuertes perturbaciones electromagnéticas (interferencias) se utiliza el cable de pares apantallado, que se obtiene agregando al cable de pares una malla metálica envolvente. Este tipo de cables presenta el menor ancho de banda de los conductores metálicos. Por ejemplo, en el caso de los pares telefónicos para servicios xDSL (apartado 3.9) su ancho de banda es de 1,1 MHz (AWG26, *American Wire Gauge*, desde la central hasta el abonado con 0,5 mm de diámetro), mientras que los pares de Categorías 5E y 6 para redes Gigabit Ethernet alcanzan un ancho de banda de 250 MHz. Los cables de pares constituyen el medio físico de comunicación más económico y su factor de mérito es de 16 Mbps.km (1 Mbps = 10^6 bps).

Los "cables de cuadretes" son conjuntos de cables de pares agrupados de dos en dos (cuadrete: cuatro hilos) para transmitir un elevado número de señales eléctricas a grandes distancias (Fig. 3.1).

Un "cable coaxial" está formado por un tubo metálico en cuyo interior se coloca un conductor en posición axial, separados ambos por un material aislante (Fig. 3.1). El ancho de banda típico de estos cables alcanza las bandas 6 a 9 del espectro electromagnético (apartado 3.3.2), por lo que un coaxial de CATV (operador de TV por cable) ofrece hasta 2050 MHz, permitiendo la canalización simultánea de servicios de telefonía, datos y TV (plan de frecuencias). Por ello su factor de mérito es de 500 Mbps.km. Se pueden agrupar varios cables coaxiales de pequeña sección o utilizar uno solo de mayor diámetro (cables submarinos) para llevar un gran número de señales eléctricas a distancias largas.

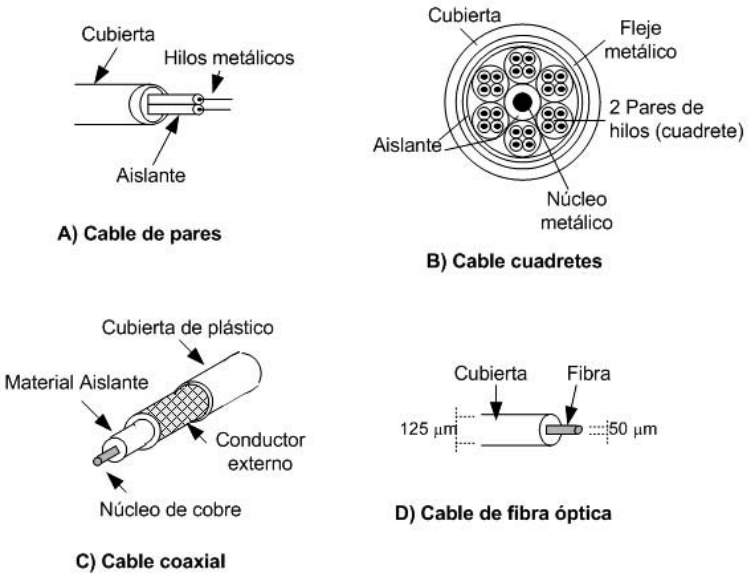


FIGURA 3.1. Tipos de cableado.

Una "guiaonda" es un tubo metálico de sección rectangular o circular que está destinado a soportar señales electromagnéticas de elevada frecuencia denominadas microondas (bandas 9 a 11 del espectro electromagnético, entre 3 GHz a 300 GHz). Se utilizan en equipos de radioenlaces para conectar los emisores y receptores a la antena correspondiente.

Los radioenlaces terrestres utilizan el medio atmosférico como soporte de las señales electromagnéticas (microondas), y están constituidos por antenas (parabólicas) colocadas en los puntos de emisión y recepción (fijos o móviles) de las señales (Fig. 3.2). Deben tener visibilidad óptica entre ellas (LOS, *Line-Of-Sight*) y su distancia máxima está determinada por la curvatura de la tierra (línea del horizonte).

Los radioenlaces, además de salvar obstáculos inaccesibles a los cables en comunicaciones fijas, fueron históricamente el único medio capaz de soportar comunicaciones móviles intercontinentales antes de la aparición de los satélites artificiales. Los telepuertos son conjuntos de radioenlaces terrestres ubicados en un mismo lugar, que se conectan con los satélites de comunicaciones para permitir el acceso a redes y servicios internacionales.

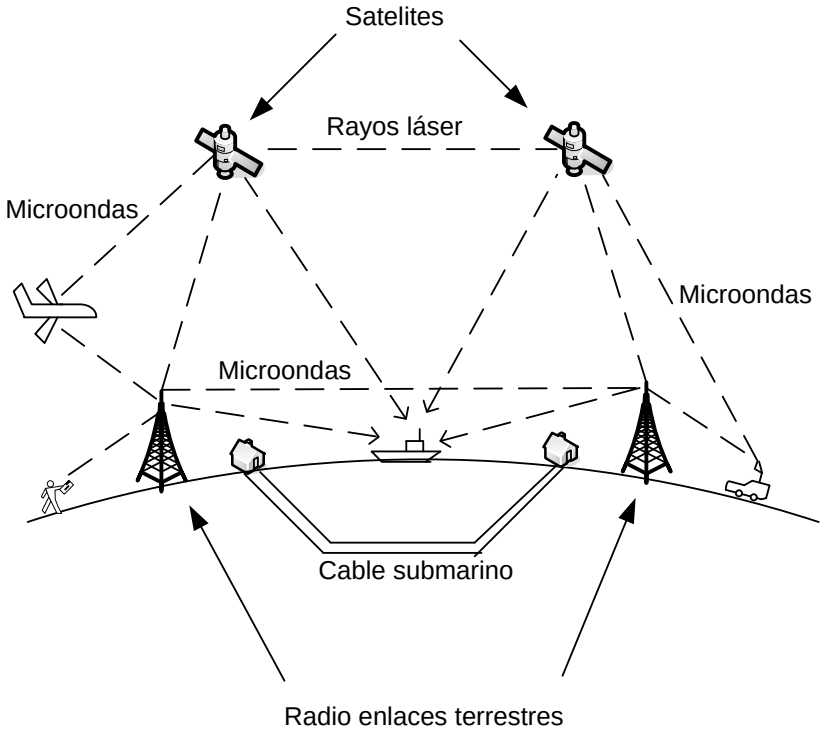


FIGURA 3.2. Radioenlaces terrestres y satélites espaciales de comunicación.

En las instalaciones de las redes de comunicación en el ámbito de los centros de trabajo, se establece una planificación del tendido de los cables sean éstos metálicos o de fibra óptica, que permite no solamente tener un conocimiento preciso de todos los equipos de interconexión presentes y los equipos terminales conectados, sino también hacer una previsión de las futuras ampliaciones a nuevos usuarios, y facilitar reconfiguraciones que permitan cambios de conexiones, debidas por ejemplo a traslados de usuarios, sin necesidad de perturbar la continuidad de los servicios en la red ni afectar a la instalación precableada.

Esta forma de implantación de una red en los edificios de los usuarios que contemple sus necesidades actuales y futuras de comunicación, sin provocar interrupciones en los servicios, ni obligar a efectuar complejas reconfiguraciones durante su período útil, es el objetivo del cableado estructurado.

Por otra parte la necesidad de asegurar la calidad de los cables tendidos desde el punto de vista de la operación fiable de la red de comunicación una vez instalada, ha llevado a los fabricantes de cables a "certificar" sus productos dentro de unas normas de calidad. Esta certificación implica la garantía de que los cables de la red cumplen unas especificaciones técnicas de transmisión determinadas.

Con objeto de presentar los aspectos más importantes que caracterizan a la instalación de las redes de comunicación en la práctica, se abordará primeramente el Cableado estructurado y a continuación la Certificación de cables.

Seguidamente se analizan estos aspectos de la instalación de redes de comunicación.

3.4.2. Cableado estructurado

Los cables son los elementos que configuran el medio físico de la red de comunicación, y su despliegue en un edificio se realiza de acuerdo con la topología de la red. El plan de la instalación mantiene la topología de la red en todos los locales ocupados por los terminales de los usuarios planta por planta.

Desde abajo hacia arriba, los elementos de interconexión del cableado estructurado son los siguientes:

- Cables troncales
- Distribuidor general
- Distribuidor de edificio
- Cables verticales
- Distribuidor de planta
- Cables horizontales

- Puntos de concentración (opcionales)
- Tomas de equipos terminales (rosetas)

A continuación se describe cada uno de estos elementos del cableado estructurado.

- Los “cables troncales” son los que llegan desde la calle a la planta baja del edificio procedentes de redes LAN, MAN o WAN externas. Serán los que permitirá unir a todos los equipos terminales del edificio a las redes que les proporcionarán todos los servicios de las tecnologías de la información y las comunicaciones, y por ello tienen una importancia primordial en todos aquellos servicios que se presten fuera de la propia red del edificio. Es evidente pues, que los fallos de funcionamiento en estos cables interrumpirán todos o parte de los servicios que se presten externamente.
- El “distribuidor general” es un armario que conecta los cables troncales con el distribuidor de edificio. En este armario están los equipos multiplexores que asignan los canales de los cables troncales a los cables verticales del edificio, y las pasarelas que adaptan los 7 niveles del modelo OSI entre la red del edificio y las redes externas.
- El “distribuidor de edificio” es un armario que conecta los cables verticales del edificio y al distribuidor de la planta baja con el distribuidor general. En este armario se colocan los equipos encaminadores que orientarán los paquetes enviados por los terminales hacia el exterior del edifico o hacia otras rutas en el interior del mismo.
- Los “cables verticales” proceden de los diferentes distribuidores de planta a partir de la planta baja, y se conectan al distribuidor de edificio.
- El “distribuidor de planta” es un armario que conecta los cables horizontales de una determinada planta a su correspondiente cable vertical. En este armario están los repetidores, concentradores, puentes y conmutadores de la planta, para crear redes virtuales dentro del edificio.
- Los “cables horizontales” proceden de las tomas de los equipos terminales de usuario u opcionalmente de los puntos de concentración, y se conectan al distribuidor de planta.
- Los “puntos de concentración” en una planta son elementos de interconexión opcionales, que conectan las tomas de los equipos terminales a los cables horizontales. Pueden ser simplemente equipos

concentradores situados en despachos o salas del personal del edificio, para agrupar tomas de equipos terminales y reducir el número de cables horizontales de la planta. Son opcionales y su función puede hacerse en el propio distribuidor de planta.

- Las "tomas de equipos terminales" son conectores empotrados en la pared que conectan los terminales de los usuarios con los cables horizontales y opcionalmente con los puntos de concentración. Habitualmente son conectores hembra RJ-45 del tipo de roseta telefónica con 8 hilos de cobre.

En el cableado estructurado generalmente los armarios de interconexión son metálicos, y llevan un conjunto de raíles laterales para crear estanterías de diferentes alturas adaptables, según su función en la instalación, a las dimensiones de los equipos de interconexión de redes mencionados. En general dentro de un armario de interconexión de redes se presentan los siguientes elementos:

- Regletas: para la conexión de cables entrantes y salientes del armario
- Paneles repartidores: que facilitan el cambio de conexiones entre equipos internos y regletas
- Latiguillos: cables cortos para hacer conexiones removibles entre los elementos del armario

3.4.3. Certificación de cables

Para la certificación de instalaciones de cableado estructurado, la Asociación de la Industria de Telecomunicación en USA (TIA) y la Organización Internacional de Normas (ISO), han establecido unas especificaciones de transmisión para cables de pares, que están basadas en el factor de mérito representado por el producto velocidad-distancia.

La ISO utiliza el término "Clase" y una letra mayúscula en orden alfabético creciente (A, B, C, D, E y F) para señalar el valor de velocidad que presenta un determinado cable para una aplicación de un ancho de banda dado, en una distancia fija. La entidad TIA emplea la palabra "Categoría" y un número en orden creciente (3, 5, 5E y 6), en cada una de las cuales agrupa diferentes distancias y velocidades.

Tanto ISO (norma 11801) como TIA (norma 568A) definen esas clases y categorías respectivamente a partir de unas especificaciones técnicas calculadas para cada frecuencia de la banda, que en cables de 4 pares (8 hilos de cobre) con una longitud dada son las siguientes:

- Atenuación: debilitamiento de la señal a lo largo del cable (medidas S/N de operación segura)
- Pérdidas de retorno: relación "señal reflejada/señal transmitida" en un par
- Diferencia de retardos entre pares (nanosegundos)
- Diafonía: perturbaciones entre pares adyacentes
- Relación "atenuación/diafonía" (ACR, *Attenuation-to-Crosstalk Ratio*)
- Suma de la ACR de todos los pares (ACR-PS, *ACR-Power Sum*)
- NEXT (*Near-End Crosstalk*): Diafonía en el extremo de entrada entre dos pares cualesquiera (extremo cercano)
- NEXT-PS (*NEXT-Power Sum*): Diafonía en un par inducida por los tres restantes en la entrada
- FEXT (*Far-End Crosstalk*): Diafonía en el extremo de salida entre dos pares cualesquiera (extremo lejano)
- FEXT-PS (*FEXT-Power Sum*): Diafonía inducida en un par por los tres restantes en la salida
- ELFEXT (*Equal Level FEXT*): Diafonía en el extremo de salida entre dos pares cualesquiera sin atenuación
- ELFEXT-PS (*ELFEXT-Power Sum*): Diafonía inducida en un par por los tres restantes sin atenuación a la salida

Otras especificaciones de los cables de pares se refieren a los materiales de las cubiertas aislantes, que deben estar libres de halógenos (FH, *Free Halogen*) para no producir humos tóxicos en caso de incendio, y además deben ser retardantes a la llama (FR, *Flame Retardant*) en el sentido de descomponerse sin crear llamas.

Desde el punto de vista de los tipos de apantallamiento eléctrico que se incorporan a los cables de pares para mantener su inmunidad electromagnética, éstos se clasifican en:

- UTP (*Unshielded Twisted Pair*): la cubierta es pirorretardante con baja emisión de humos (sin pantalla)
- FTP (*Foiled UTP*): bajo la cubierta plástica hay sólo un apantallamiento envolvente
- STP (*Shielded Twisted Pair*): bajo el apantallamiento envolvente además cada par tiene su propio apantallado

3.5. LAS FIBRAS ÓPTICAS

3.5.1. Características

Son cilindros de material aislante como las fibras de vidrio o materiales plásticos con un diámetro del orden de 50 micras y una cubierta protectora de 125 micras (1 micra = 10^{-3} mm), capaces de transmitir las señales electromagnéticas de la banda óptica (apartado 3.3.2), ya sean éstas visibles o invisibles (ultravioleta e infrarrojo) (Fig. 3.1). Los cables de fibras ópticas agrupan conjuntos de fibras para soportar mayor número de señales, o bien una sola fibra mediante técnicas de multiplexación. Existen cables con 8, 16, 120, 200 fibras ópticas que pueden incluir cables planos (*Optical Fibres Ribbon*) de 4, 8 o más fibras hasta alcanzar conjuntos de 400 fibras. La fibra óptica es el medio físico de comunicación que permite mayores velocidades de transmisión que los cables metálicos debido a su elevadísimo ancho de banda (10 PHz =10^{16} Hz), y por ello su factor de mérito alcanza valores superiores al Gbps.km (1 Gbps = 10^{9} bps).

Para comprender el comportamiento de las fibras ópticas se mencionan los fenómenos básicos de la propagación de la luz (banda óptica) a su través:

- Absorción
- Reflexión
- Refracción
- Dispersión

La **absorción** consiste en la pérdida de energía o atenuación (en dB) como consecuencia de la acción de frenado de los átomos de la fibra sobre los fotones que los atraviesan, la señal óptica se debilita a medida que avanza a lo largo del material. El cálculo de la atenuación se calcula según la ecuación 3.2,

$$\text{Atenuación (dB)} = 10.\log (Pe/Ps) \quad (3.2)$$

de tal manera que si a la salida de la fibra la potencia de la luz es la mitad de la que se introdujo a la entrada (Pe = 2.Ps) la atenuación es de 3 dB (caída de potencia al 50%).

La **reflexión** es una desviación del haz de fotones en sentido contrario a su propagación cuando alcanza la frontera entre dos medios físicos, de tal manera que el ángulo de incidencia (θ) es igual al ángulo de desviación en sentido opuesto, respecto al eje perpendicular a la frontera entre los dos medios (el de transmisión y el de rechazo). El efecto es como si los fotones "rebotasen" contra la frontera.

En las fibras ópticas el ángulo (θ) en el que se produce el rebote, desde el medio exterior a su extremo de entrada, se denomina "apertura numérica" de la fibra y depende de los índices de refracción del núcleo (Nn) y la cubierta (Nc) de la fibra según la ecuación 3.3.

$$\text{Sen } \theta = (Nn^2 - Nc^2)^{½} \quad (3.3)$$

El **índice de refracción** (N) de un material representa físicamente la inversa de la densidad de sus átomos, y se define como la relación entre la velocidad de la luz en el vacío (Cv = 300.000 km/s) y la de la luz en el medio (Cm) según al ecuación 3.4.

$$N = Cv/Cm \quad (3.4)$$

En consecuencia, los medios muy poco densos (próximos al vacío) presentarán un índice de refracción próximo a la unidad, mientras que los muy densos tendrán un índice de refracción muy superior a la unidad ($N >> 1$). Ocurre a la inversa que con las ondas sonoras o elásticas (mecánicas) que no se propagan en el vacío (por ausencia de materia), mientras que lo hacen con mucha facilidad en medios muy densos (por ejemplo el agua).

La "apertura numérica" tiene gran importancia en el acoplamiento de la fuente de luz (diodo LED, Láser, etc.) externa, situada al extremo de entrada a la fibra, para garantizar la mayor eficiencia en la transmisión de luz desde el fotoemisor (fuente) al interior de la fibra. Para ello hay que asegurar que el ángulo de emisión de la fuente tenga un valor inferior a θ.

Otro aspecto fundamental en la conducción de la luz por una fibra óptica es que el índice de refracción de la cubierta sea mucho menor que el índice de refracción del núcleo (Nc << Nn) para asegurar la reflexión de la luz hacia el interior de la fibra, como si rebotase en la cara interior de la cubierta, y no pase la luz a través de la cubierta fuera de la fibra. De esta forma la fibra se comporta como una guía de luz, confinando al haz de fotones en su interior durante la propagación de un extremo a otro (Fig. 3.3).

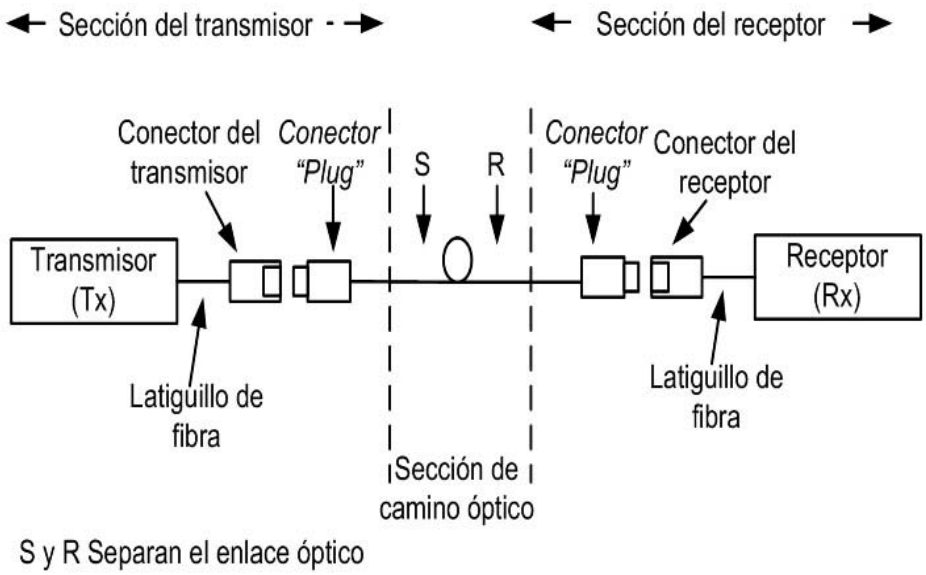


FIGURA 3.3. Interfaces de un sistema óptico.

La refracción es una desviación de la dirección del haz de fotones cuando se produce un cambio en el índice de refracción de un mismo medio (Ni) o se atraviesa la frontera entre dos medios. Este cambio de trayectoria puede producir fluctuaciones en la potencia que llega al extremo de la fibra, dando lugar a distorsiones (deformaciones) en la señal que entrega el fotorreceptor (fotodiodo) a lo largo del tiempo.

La **dispersión** (*scattering*) es un fenómeno relacionado con el cambio del índice de refracción de un medio que depende de la longitud de onda (λ) de la radiación luminosa. Esta dispersión se denomina de "Rayleigh" si el tamaño de las partículas en un medio es menor que λ (si λ aumenta entonces N disminuye), en caso contrario se denomina "Mie".

Un experimento muy conocido al respecto es la dispersión de la luz en un prisma, descubierto por Newton (descomposición de la luz blanca en los colores del arco iris), y su conversión de nuevo en luz blanca al incidir en otro prisma.

Como ventajas importantes de las fibras ópticas frente a los cables metálicos cabe señalar:

- Elevado ancho de banda que permite altas velocidades binarias (superiores al Gbps)
- Baja atenuación que asegura la correcta transmisión a lo largo de grandes distancias
- Ausencia de emisiones electromagnéticas lo que impide efectos de diafonía fuera del cable
- Pequeños volumen y peso que permiten una gran capacidad para el transporte de datos
- El material es dieléctrico y por lo tanto es intrínsecamente seguro (ausencia de chispas)
- Elevada tensión de aislamiento lo que evita lazos de masa

3.5.2. Tipos de modos

En la fabricación de las fibras ópticas prácticas, la propagación de la luz está condicionada por la dispersión que el propio material de la fibra ejerce sobre el haz de fotones. Por esta razón en la transmisión digital se estima el efecto de la dispersión en términos de retardo de la señal digital en nanosegundos por kilómetro (10^{-9} s/km). Con objeto de mitigar este efecto de retardo de un extremo a otro, se han diseñado fibras ópticas cuyo núcleo presenta cambios en el índice de refracción en función de la distancia al centro, es decir, su N es una función del radio (r). Según sea esta relación la fibra óptica se denomina Multimodo de "Salto de Índice" (*step-index*), donde Ni es una función discreta de r (Ni disminuye a saltos desde el eje de la fibra hacia la cubierta), o Multimodo de "Índice Gradual" (*graded-index*) en la que Ni es una función continua de r (N disminuye continuamente desde el eje de la fibra hacia la cubierta).

Las fibras Multimodo suelen tener un diámetro de núcleo de 50 μm (1 micra = 10^{-6} m) y su diámetro de cubierta de 125 μm. Las de Salto Índice presentan diferentes trayectorias (modos) rectilíneas en zigzag dentro de la fibra, lo que

les limita su factor de mérito a 30 Mbps.km, con un retardo de 60 ns/km. Las de Índice Gradual tienen modos en forma senoidal que se ponen en fase a cada semiperíodo, permitiéndoles un factor de mérito de 600 Mbps.km, con un retardo de 0,5 ns/km. Estos efectos se denominan "dispersión modal".

Para alcanzar factores de mérito más elevados se fabrican las fibras Monomodo que, como su nombre indica, concentran el haz de fotones en una sola trayectoria, que idealmente se propaga en la dirección del eje del núcleo, eliminando por tanto la "dispersión modal". Estas fibras son más caras que las Multimodo presentando un núcleo con un diámetro de 5 μm, y un factor de mérito de 100 Gbps.km. Su empleo típico es en las grandes redes troncales de los operadores internacionales de redes de comunicación, cables submarinos, operadores de CATV (apartado 3.9), y en todos aquellos casos en que se necesiten grandes velocidades binarias para dar servicios de elevado ancho de banda. En este tipo de fibras también se produce un retardo que es debido a las dispersiones denominadas "cromática" y de "modo de polarización". La "dispersión cromática" es el resultado de la combinación de la "dispersión material" y la "dispersión de guiaonda", que físicamente representan las limitaciones que el material de la fibra y su forma de fabricación ejercen sobre la propagación del haz de fotones. La dispersión material es el efecto del cambio del índice de refracción del material de la fibra con el valor de λ (Ni es función de λ_i) y la dispersión de guiaonda corresponde al efecto en la fibra de la relación de N con r. La dispersión de "modo de polarización" se produce por el hecho de que los vectores de campo eléctrico y campo magnético de la onda se desfasan como consecuencia del cambio de Nx y Ny del material con la distancia (eje Z), y tiene importancia para velocidades superiores a 10 Gbps.

3.5.3. Nomenclatura de las portadoras ópticas

El material más utilizado para la realización de fibras ópticas en el infrarrojo es la sílice (dióxido de Silicio). Las λ utilizadas en las redes de fibra óptica de sílice son: 850 nm, 1310 nm y 1550 nm que, en el lenguaje de estas redes, se denominan respectivamente "primera ventana de transmisión", "segunda ventana de transmisión" y "tercera ventana de transmisión". La velocidad de la red es tanto mayor cuanto lo sea el número de la ventana.

Un ejemplo de red de fibra óptica de alta velocidad y gran distancia de transmisión, es el cable submarino FLAG que cubre una distancia de 27.300 km desde el Reino Unido al Japón, con una velocidad de 10,6 Gbps (10.600 millones de bits por segundo).

La norma ITU-T G692 para portadoras ópticas en la 3ª ventana establece una separación de 0,8 nm para anchos de banda de 100 GHz, mientras que para anchos de banda de 50 GHz es de 0,4nm. Esto permite su aplicación en sistemas de multiplexación densa por división de longitud de onda (DWDM, *Dense Wavelength División Multiplexing).* El término "densa" proviene de que el canal de guarda entre portadoras ópticas es del orden de 1nm.

Las velocidades en ATM (*Asynchronous Transfer Mode*) se toman de la especificación de la "jerarquía digital síncrona" (SDH, *Synchronous Digital Hierarchy*). Esta jerarquía define un formato de multiplexación mundial que permite transportar tráfico síncrono (igual velocidad) e isócrono (distinta velocidad) por las redes digitales (TDM síncrona). El CCITT designa la "jerarquía digital síncrona" por el acrónimo SDH, mientras que ANSI utiliza el acrónimo SONET (*Synchronous Optical Network*). El valor de referencia para la velocidad se toma de RDSI-BA (Red Digital de Servicios Integrados en Banda Ancha) y es de 155,52 Mbps.

Esta velocidad de referencia se multiplica por números enteros para obtener los valores de velocidad normalizados. En SONET estos valores se designan de forma distinta para redes ópticas (OC, *Optical Carrier)* que en el caso de redes con otros medios físicos (STS, *Synchronous Transport Signal*). En SDH se designan como STM (*Synchronous Transfer Module*) con un número a la derecha que indica el factor multiplicativo. En ATM las velocidades están comprendidas entre 25 Mbps y 622,08 Mbps (STM-4).

A continuación se indican algunas velocidades y sus denominaciones correspondientes:

- 51,84 Mbps: STS-1/OC-1
- 155,52 Mbps: STS-3/OC-3, STM-1
- 466,56 Mbps: STS-9/OC-9, STM-3
- 622,08 Mbps: STS-12/OC-12, STM-4
- 933,12 Mbps: STS-18/OC-18, STM-6
- 1244,16 Mbps: STS-24/OC-24, STM-8
- 1866,24 Mbps: STS-36/OC-36, STM-12
- 2488,32 Mbps: STS-48/OC-48, STM-16

3.6. LOS SATÉLITES ESPACIALES DE COMUNICACIONES

3.6.1. Conceptos básicos

Los servicios de satélites, como su nombre indica, se basan en la utilización de satélites espaciales de comunicaciones situados en diferentes órbitas alrededor de la tierra, que permiten a las señales radioeléctricas enlazar puntos cualesquiera, ya sean de la superficie terrestre por muy inaccesibles que resulten en mar o en tierra, de la atmósfera en la navegación aérea, o del espacio exterior en las naves espaciales y estaciones espaciales. Para su estudio pormenorizado se presentan primeramente los conceptos fundamentales, y a continuación se describen los servicios basados en satélites.

Si bien la vida operativa de un satélite de comunicaciones es aproximadamente de 10 a 15 años, que resulta corta comparada con la vida útil de un cable submarino, estimada en 25 años, la importancia de las comunicaciones fijas y móviles soportadas por los satélites es de tal magnitud, que resultan imprescindibles debido a que solamente ellos permiten una cobertura mundial con una única "red" de tres satélites geoestacionarios.

En efecto, ya en 1945 Arthur C. Clarke propuso la cobertura mundial del planeta Tierra con una red de tres satélites de comunicaciones separados por tres arcos de 120 grados, sobre una órbita circular en el plano ecuatorial, que distase aproximadamente 36.000 km de su superficie. A esta particular órbita se la denomina "geoestacionaria", y es utilizada por la mayoría de los sistemas de satélites existentes en el mundo, ya sean de carácter internacional como Inmarsat e Intelsat, o de tipo nacional como es Hispasat en España.

Esta cobertura mundial no es posible con una única red terrestre, cualquiera que sea el medio que utilice, en cuanto que cada terminal de usuario necesita una conexión directa a un cable o a una red inalámbrica. Por el contrario, una "estación terrestre" de satélite, ya sea para un único usuario o para varios usuarios, no necesita conexión en tierra a ninguna red terrestre para recibir un servicio de satélite. Esto es más evidente en las "estaciones de satélite" a bordo de aeronaves y en la navegación marítima. Por esta razón las redes de satélites representan una alternativa rápida y eficaz, en los aspectos técnicos y económicos, cuando se trata de dar servicios de comunicación a países en desarrollo o con infraestructuras terrestres muy escasas.

Los servicios móviles terrestres despliegan sus recursos de transmisión, cables, radioenlaces y antenas, sobre la superficie terrestre, y por ello están orientados fundamentalmente a los usuarios que se desplazan por ella, aunque también dan cobertura a navíos próximos a las costas ya las aeronaves en las cercanías de los aeropuertos.

Existen servicios de comunicación que combinan redes de satélites con redes terrestres, dado que los servicios móviles terrestres pueden apoyarse en redes de satélites en algunos casos, y también los servicios móviles de satélites pueden apoyarse en redes terrestres.

La adecuada combinación de los recursos en el espacio y en tierra para configurar una red de satélites requiere el conocimiento de: la forma en que los satélites se enlazan con las "estaciones terrestres" o los terminales de los usuarios en su caso, las órbitas utilizadas por los satélites y la posición de los satélites dentro de sus órbitas.

Desde el punto de vista del enlace se distinguen los siguientes conceptos:

- Segmento espacial
  - Satélite
  - Enlace ascendente
  - Enlace descendente
  - Estaciones terrestres de telecontrol
- Segmento terrestre
  - Fuente de información
  - Estación terrestre emisora
  - Estación terrestre receptora
  - Terminal de usuario

En el "segmento espacial" de una red de satélites se incluyen los propios satélites que actúan como equipos electrónicos repetidores del nivel físico del modelo OSI, recibiendo las señales emitidas por las estaciones transmisoras en unas determinadas frecuencias, lo que se denomina enlace ascendente, amplificando dichas señales, y cambiándolas a otras frecuencias de valor inferior, para reenviarlas a las estaciones terrestres receptoras, lo que recibe el nombre de enlace descendente. Las estaciones terrestres de telecontrol realizan

las funciones de seguimiento, telemetría y telemando de la red de satélites, y no participan directamente en la comunicación entre las estaciones terrestres emisoras y receptoras a través de los satélites de la red.

El segmento espacial es el que encarece la red de satélites, puesto que debe garantizar la correcta emisión y recepción de las señales de los enlaces ascendente y descendente de cada satélite.

Debido a que las pérdidas de propagación de la señal, en términos de relación entre la potencia recibida y la potencia transmitida, son directamente proporcionales a la frecuencia de la señal y al cuadrado de la distancia, con el agravante de que para frecuencias superiores a 10 GHz la lluvia añade mayores pérdidas, las potencias de emisión de las estaciones terrestres y de los satélites deben estar compensadas para garantizar el correcto funcionamiento de los enlaces ascendente y descendente.

Por esta razón se prefiere utilizar una frecuencia mayor en el enlace ascendente ya que es posible compensar las pérdidas con estaciones emisoras terrestres de gran potencia, mientras que la limitación de potencia que impone el satélite para poder alargar su vida útil, aconseja utilizar la frecuencia menor en el enlace descendente. Otra razón importante para cambiar las frecuencias de emisión y recepción en el satélite es reducir las interferencias entre sus antenas de a bordo.

En el "segmento terrestre" de una red de satélites se consideran todos los equipos electrónicos encargados de la comunicación a través del satélite, tanto si están ubicados en una posición fija sobre la superficie terrestre, como si están a bordo de navíos, o en vehículos terrestres, espaciales, o aeronáuticos. Si bien en un caso general se distingue entre equipos que generan información, los que la transmiten, los que la reciben y los que finalmente la presentan en un terminal de usuario, existen aplicaciones en que un terminal de usuario realiza todas esas funciones como en la telefonía móvil celular por satélite. Asimismo en los servicios fijos de satélites (FSS, *Fixed Satellite Services*) para aplicaciones de difusión de señales de televisión (DBS, *Direct Broadcast Satellite*) los terminales de los usuarios solamente reciben información desde los satélites, mientras que en aplicaciones de envío de datos y órdenes, como telemetría, noticias, etc., los terminales solamente emiten información hacia los satélites.

3.6.2. Tipos de órbitas

Desde el punto de vista del tipo de las órbitas, considerando las altitudes típicas, el período orbital, el retardo, el número mínimo de satélites para dar total cobertura a la Tierra, y la posición de los planos orbitales, se clasifican en:

- LEO (*Low Earth Orbit*): menor de 1000 km, 1,5 h/día, 20 ms a 60 ms, 50 satélites, varios planos polares o inclinados (Fig. 3.4).
- MEO (*Medium Earth Orbit*): 10.000 km a 20.000 km, 6 h a 12 h/día, 80 ms a 120 ms, 10 a 15 satélites, varios planos polares o inclinados
- GEO (*Geostationary Orbit*): 36.000 km, 24 h/día, 280 ms, 3 satélites, una órbita ecuatorial (Latitud 0°)
- HEO (*Highly inclined Elliptical Orbit*): 1000 km a 40.000 km, 12 h a 24 h/día, 200 ms a 300 ms, 5 a 12 satélites, varias órbitas de elevada elipticidad y gran inclinación respecto al plano ecuatorial (próxima a un plano meridiano)

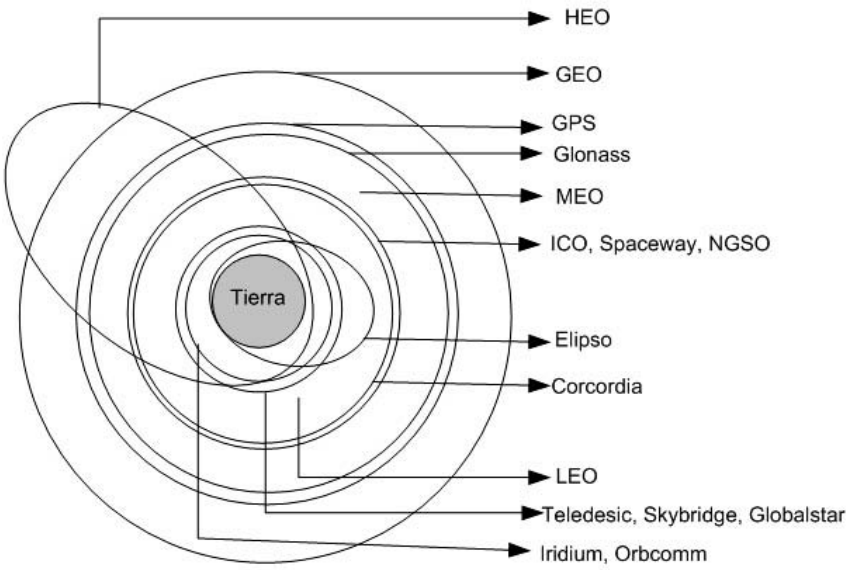


FIGURA 3.4. Órbitas utilizadas en las comunicaciones.

Las órbitas LEO son las más próximas a la tierra y por ello hay que mantener muchos satélites en órbita simultáneamente para dar cobertura a toda la superficie terrestre. Su velocidad de desplazamiento respecto a la tierra es de

10 km/s, y por ello en una posición terrestre un satélite es visto solamente durante 1,5 horas al día. Asimismo, su vida es más limitada porque en estas órbitas los satélites van decayendo, debido a la atracción terrestre, y consumen mucho combustible en su recolocación. Otro inconveniente es la aparición del efecto Doppler, es decir, el cambio en la frecuencia de la señal que llega al receptor, debido a su velocidad relativa de desplazamiento. Sus ventajas residen en que por su menor distancia tanto los retardos de las señales en volver a Tierra como las potencias de emisión son más reducidos que en MEO y GEO, y que los costes de lanzamiento son menores. Desde el punto de vista de las estaciones terrestres, éstas deben incorporar a sus antenas un mecanismo de seguimiento para mantener permanentemente dirigida la trayectoria del haz de radiación de la antena hacia el satélite, durante su período orbital. En el caso de las constelaciones de satélites LEO para aplicaciones de telefonía celular, se presenta el inconveniente del aumento de la tasa de traspaso entre satélites (*handoff*) como consecuencia de su desplazamiento.

En el caso de las órbitas MEO las ventajas y los inconvenientes son muy parecidas a las LEO, aunque los costes de lanzamiento, el retardo y las potencias de emisión de los satélites son mayores. Como en las LEO también las estaciones terrestres necesitan un mecanismo de seguimiento incorporado a sus antenas durante el período orbital. La velocidad de desplazamiento es de 4 km/s, y el satélite es visto durante 6 horas al día.

Las órbitas polares LEO y MEO reciben su denominación por estar situadas en planos que se cortan en el eje de rotación terrestre, siguiendo los meridianos, y por ello todas se cruzan en los polos. Estas órbitas tienen la ventaja de dar la mejor cobertura en esas zonas del planeta. Las órbitas inclinadas reciben su nombre por situarse en planos inclinados, con ángulos diferentes respecto al plano ecuatorial terrestre, que se cortan en el centro de la Tierra.

La órbita GEO recibe su nombre de que los satélites en ella situados se mueven en sincronismo con el giro de la Tierra, siguiendo para ello una trayectoria circular en el plano ecuatorial terrestre, por tanto de latitud 0º, llamada “geosíncrona” o “geoestacionaria”. Debido a este sincronismo, los satélites en esta órbita acompañan a la tierra en su giro completo de 24 horas, y por ello aparecen a las estaciones terrestres como si estuviesen situados en un punto fijo del espacio caracterizado por su longitud que puede ser Este (E) u Oeste (W). Por ejemplo los satélites Hispasat están en la posición de longitud 30º W en la órbita geosíncrona. Esto tiene las ventajas de que las estaciones terrestres son: más simples en el caso de los satélites GEO porque su período orbital es de 24 h, no necesitan mecanismos de seguimiento en sus antenas, y

que en las redes de satélites prácticas se utilizan solamente 4 satélites para dar cobertura mundial (con algunos problemas en los polos). Otra ventaja es que, por su posición estacionaria respecto a las estaciones terrestres, el efecto Doppler es nulo. No obstante, debido a su gran distancia a la tierra presentan elevados retardos y grandes potencias de emisión frente a los satélites LEO y MEO.

El interés de los satélites HEO reside en que para cada órbita existe una zona geográfica de elevada latitud (mayor de 70° N y 70° S), como es el caso de las zonas polares terrestres, que siempre está visible en las proximidades del apogeo de la órbita, durante una fracción del período orbital superior a la mitad del mismo. De tal forma que con dos satélites situados en la misma órbita HEO se obtiene la cobertura permanente de casi un hemisferio (de ahí también el nombre de órbitas "heliosíncronas"), y con una constelación de 5 satélites en órbitas HEO adecuadas se consigue la cobertura total del planeta. Frente a los satélites LEO y MEO, los HEO presentan similares ventajas e inconvenientes que los GEO. No obstante el mayor ángulo de elevación de los HEO, superior a 45°, mejora notablemente las prestaciones de las estaciones terrestres, lo que supone una ventaja frente a los GEO, aunque tienen el inconveniente de necesitar un mayor número de satélites que éstos para alcanzar la cobertura global.

3.6.3. Servicios VSAT

Son servicios de comunicación fijos destinados a las empresas, como el fax, la videoconferencia, la transmisión de datos, etc., que se prestan a través de redes de satélites (FSS), cuyas estaciones terrestres tienen la particularidad de presentar antenas parabólicas de pequeño diámetro (VSAT, *Very Small Aperture Terminal*), inferior a 0,5 m, de las que reciben su nombre. Estos servicios aparecieron a principios de los años 80 del siglo XX como consecuencia del insuficiente desarrollo de las redes terrestres.

Además de los terminales VSAT y del satélite, existe una estación terrestre central, denominada "*Hub*" en la terminología anglosajona, que constituye en núcleo central de la red VSAT. Esta estación permite la conexión vía satélite de todos los terminales VSAT, y tiene una antena parabólica de 5 m a 12 m de diámetro.

Las topologías de las redes VSAT son las siguientes:

- Estrella
  - Punto-multipunto unidireccional
  - Multipunto-punto unidireccional
  - Punto-multipunto bidireccional
  - Punto-punto bidireccional
- Malla

En la topología en estrella cada terminal VSAT se conecta con la estación terrestre central (*hub*) mediante un enlace directo, presentando diferentes configuraciones. En las configuraciones unidireccionales la información circula en un solo sentido: bien desde la estación central hacia los terminales VSAT a través del satélite (punto-multipunto) en una modalidad de "difusión" o viceversa (multipunto-punto) en una modalidad de "concentración". En la modalidad de difusión se destacan las aplicaciones de VSAT para distribución de noticias para agencias periodísticas, de canales musicales para emisoras de radio, información de bolsa y financiera, así como la impresión remota de periódicos. En la modalidad de concentración las aplicaciones son de telecontrol tales como la telemetría, para transmitir datos de supervisión de instalaciones en las que predomina la seguridad, o para la recogida de datos de campo, como meteorología, niveles de ríos y embalses, etc., y su envío al centro de control de operaciones.

En las configuraciones bidireccionales los terminales VSAT transmiten y reciben información, permitiendo que la transferencia sea en los dos sentidos: *hub*-VSAT y VSAT-*hub*. En la configuración punto-multipunto los terminales VSAT no se comunican entre sí, y se utiliza para enlazar una sede central con puntos dispersos geográficamente, que no necesiten conectarse entre ellos, como es el caso de la conexión entre la sede central de una empresa y sus sucursales. En la configuración punto-punto los terminales VSAT se comunican entre ellos a través de la estación central mediante un doble salto de satélite, de tal manera que la información circula desde el terminal VSAT emisor a través del satélite al *hub*, y desde éste de nuevo a través del satélite hacia el terminal VSAT receptor.

La topología en malla permite la comunicación directa entre los terminales VSAT sin pasar por la estación central, y por ello sólo necesitan un único salto de satélite. Esto elimina los retardos debidos al doble salto en la configuración de estrella punto-punto bidireccional, y mejora la calidad de los servicios interactivos como la telefonía. Para aumentar la eficiencia en el uso de los

canales del satélite, en la topología en malla, se utilizan los métodos de acceso al medio FDMA y TDMA con demanda adaptativa que se denominan DAMA.

Las aplicaciones de las configuraciones bidireccionales permiten la realización de servicios fijos conocidos de las redes terrestres como los de: consulta, transacción, audiovisuales y de teletratamiento.

3.7. LA MULTIPLEXACIÓN Y LA MODULACIÓN

3.7.1. Las señales analógicas y digitales

Independientemente de si una señal de información es de tipo eléctrico circulando por un cable metálico, o si es una radiación electromagnética transmitida por una guía de onda (microondas), una fibra óptica (infrarrojo) o el espacio libre (RF y microondas), esta señal puede tener un formato analógico o un formato digital (apartado 3.3.2).

Debido a que en su transmisión por un medio físico la señal de información es afectada de distinto modo según su formato, desde el punto de vista de los fenómenos que en ella se producen como consecuencia de su paso a través de cada uno de los medios mencionados, es necesario presentar los conceptos básicos que las distinguen, y los tipos de modulación que permiten modificar algunas de sus características para optimizar su transmisión por un determinado medio físico, y la posibilidad de conversión de una señal digital en analógica y viceversa.

A continuación se analizan los conceptos básicos y los tipos de modulación. Para su estudio detallado se presentan primeramente las señales analógicas y en segundo lugar las digitales:

- Una señal analógica presenta un conjunto infinito de valores en un intervalo dado. Por ejemplo, si su amplitud varía entre los valores 2 y 3, puede tomar cualquier valor entre ellos: 2,1-2,1003-2,1148-2,203-2,2401, etc., dando lugar a una serie infinita de valores.

  El concepto “analógico” procede de la idea de que las magnitudes físicas de la naturaleza cambian gradualmente con el tiempo tomando infinitos valores. Un ejemplo típico es la temperatura ambiente, que cambia de forma gradual a lo largo del día, entre una temperatura

máxima y una mínima. Las señales eléctricas y electromagnéticas que cambian de forma "análoga" a como lo hace una magnitud de la naturaleza, se denominan "señales analógicas".

Una señal analógica simple es la onda senoidal, que para su caracterización completa, a su amplitud, su frecuencia y su longitud de onda (λ) hay que añadir su "fase". La fase de una señal senoidal representa el valor del tiempo que transcurre antes de que la señal empiece a tomar valores distintos de cero. En una representación gráfica amplitud-tiempo, la fase representa la distancia del comienzo de la señal al origen de la gráfica sobre el eje horizontal de tiempo. En general la fase de una señal de cualquier tipo, se define como su posición de comienzo respecto al origen en una gráfica amplitud-tiempo. Asimismo, en aplicaciones de modulación, la fase suele expresarse en unidades de grados angulares en vez de unidades de tiempo.

Para simplificar el estudio de las señales analógicas, se considera que en el caso más general las señales analógicas de interés en comunicaciones son todas señales analógicas periódicas de forma irregular. Este tipo de señales pueden ser vistas evolucionando en el tiempo, como una representación gráfica amplitud-tiempo en el instrumento de medida electrónico denominado "osciloscopio". Se demuestra matemáticamente que una señal periódica irregular está compuesta por un número infinito de señales senoidales puras, cuyas frecuencias son múltiplos enteros de la frecuencia de la señal periódica irregular.

Las frecuencias de las señales senoidales puras se denominan "componentes espectrales" o "armónicos" de la señal analógica periódica, y pueden ser representadas en un instrumento de medida electrónico denominado "analizador de espectro", si se conecta la señal analógica periódica a su entrada. El "ancho de banda" de la señal analógica se define como el conjunto de todas las frecuencias de su espectro.

Si la señal analógica en su transmisión a través de un medio sufre la pérdida de alguno o algunos de sus componentes espectrales o armónicos, es decir si se recorta su ancho de banda (apartado 3.3.1), la señal que se obtiene en el extremo del receptor resulta "distorsionada".

El término distorsión proviene del hecho de que la señal analógica pierde la forma original que tenía en la entrada, presentando una forma

distinta en el receptor. Este cambio de forma en el receptor modifica la información que transporta la señal analógica original, y por ello el receptor no recibe correctamente la información enviada por el emisor.

Otra causa de distorsión de la señal analógica es la modificación de su forma por la acción de señales interferentes en el medio. Al conjunto de estas señales interferentes se denomina "ruido" (en inglés "*noise*"), y por ello se considera un parámetro importante en la transmisión de una señal analógica la relación entre la potencia de la señal original (S) y la potencia del ruido (N) presente en el medio, expresada por la notación S/N (leída "relación señal-ruido"). Así, cuanto mayor sea S/N mejor será la calidad de la señal en el receptor (apartado 3.3).

Señales analógicas típicas y su ancho de banda son las siguientes:

- Las señales telefónicas o vocales (0 Hz a 4 kHz)
- Las señales de audio en alta fidelidad (20 Hz a 20 kHz)
- Las señales de TV (0 Hz a 6 MHz)

Para poder transmitir adecuadamente estas señales por un medio físico, es decir sin distorsión apreciable, tienen que darse dos circunstancias en ese medio: que tenga un ancho de banda igual o mayor que el de la señal a transportar, y que esté libre de ruido o que la relación S/N en la transmisión sea muy elevada.

Es evidente que utilizar un mismo medio físico para transportar las señales analógicas mencionadas con elevada calidad, representaría un desperdicio en el caso de las señales de audio y todavía mayor en el caso de las señales telefónicas, frente a las señales de TV, puesto que las diferencias en ancho de banda son muy importantes.

Por esta razón se utilizan diferentes medios de transmisión en el transporte de las señales analógicas en función de su ancho de banda. Así los cables telefónicos son pares metálicos muy finos (ancho de banda en VF y VLF) y por ello más económicos, frente a los cables de las cadenas de audio (ancho de banda en VLF) que son más gruesos (apartado 3.4.1). Los cables para TV son distintos a los anteriores pues tienen una disposición coaxial, y por ello son los más caros de los tres tipos de señal, para presentar un mayor ancho de banda.

Hay que tener en cuenta que habitualmente por un mismo medio físico se transportan varias señales analógicas a la vez, surgiendo así el concepto de "canalización". En el caso de un cable coaxial que transporte varios canales de TV simultáneamente, éste tiene que garantizar que caben tantos anchos de banda como señales analógicas de TV distintas haya.

Si se considera que por un cable coaxial han de transportarse 7 canales de TV, éste debe tener un ancho de banda mínimo de 42 MHz (7 veces 6 MHz). Se comprueba que este ancho de banda cabe en la Banda III de VHF que está destinada a 7 "canales altos de TV" (5 a 11), ya que su diferencia es de 68 MHz (230 MHz menos 162 MHz). Por lo tanto ese cable coaxial con un ancho de banda de 68 MHz serviría para transportar los 7 canales de la Banda III de VHF.

- Una señal digital presenta un conjunto finito de valores en un intervalo dado, que en los circuitos electrónicos digitales es solamente de dos. La denominación de "digital" remite a la idea de que los valores intermedios del intervalo son tan escasos, que pueden ser contados con los dedos de las manos.

  Considerando el ejemplo analógico anterior, si la amplitud de la señal digital varía entre los valores 2 y 3, ésta debe tomar solamente dos valores: el 2 o el 3. En "lógica positiva" se elige llamarle "0 lógico" al valor inferior (2) y "1 lógico" al valor superior (3), representando el "0" o el "1" un bit de información digital. En "lógica negativa" la asignación sería al revés, el valor inferior (2) sería el "1" y el valor superior (3) sería el "0". El nombre de "bit" procede de la unión de letras de los términos ingleses "*binary digit*".

  Una señal digital simple es la onda "cuadrada" que, como la onda senoidal, también se caracteriza por su amplitud, su frecuencia y su fase, y puede ser vista en el osciloscopio evolucionando en el tiempo, como una representación gráfica amplitud-tiempo. En algunas aplicaciones de modulación con señales digitales, en vez de la frecuencia de la onda "cuadrada" se utiliza la "anchura" que tiene esta onda en el valor alto ("1 lógico").

  Cualquier información en formato digital se expresa como combinaciones de ceros y unos. Precisamente esta peculiar forma de representar la información obliga a caracterizar de forma distinta para

su transmisión, a las señales digitales frente a las señales analógicas. El las aplicaciones en las que es necesario ver un gran número de señales digitales simultáneamente, se utiliza el instrumento electrónico denominado "analizador lógico".

En la caracterización de las redes digitales, el concepto de ancho de banda de un medio se sustituye por el de velocidad de la señal digital que lo atraviesa en número de bits transmitidos en un segundo (bps), y el de relación señal-ruido (S/N) por el de "tasa de error de bit" (BER, *Bit Error Rate*) que introduce el medio para esa velocidad.

De hecho en las redes digitales es muy común decir que una red de elevada velocidad binaria (por ejemplo de 20 Mbps) es una red digital de "banda ancha", estableciendo como sinónimos los conceptos de "ancho de banda" y "velocidad binaria".

Esta equivalencia procede de la teoría de la comunicación que, mediante la ecuación de Shannon (apartado 3.3), permite determinar la máxima velocidad binaria alcanzable en el medio (bps), conocidos su ancho de banda (Hz) y su relación señal ruido (S/N).

La tasa de error de bit se expresa como una potencia entera negativa del número 10. Por ejemplo, un valor típico de BER en redes de área local es de 10^{-9}, lo que indica la posibilidad de que se produzca un error en la transmisión de un bit de información a través del medio físico, por cada mil millones de bits transmitidos (10^9).

3.7.2. Multiplexación

La creación de "canales" dentro del ancho de banda del medio físico de una red permite que pueda ser compartido por varias señales simultáneamente, y aumentar así la eficacia en el uso de los recursos de la red (apartado 3.3.1). Los procedimientos de asignación de canales en un medio físico se denominan de multiplexación:

- Por división del espacio (SDM, *Space Division Multiplexing*)
- Por división de tiempo (TDM, *Time Division Multiplexing*)
- Por división de frecuencia (FDM, *Frequency Division Multiplexing*)
- Por división de frecuencia ortogonal (OFDM, *Orthogonal FDM)*

- Densa por división de longitud de onda (DWDM, *Dense Wavelength Division Multiplexing*)
- Por división de código (CDM, *Code Division Multiplexing*)
- Por división de código en banda ancha (WCDM, *Wideband* CDM)

Los equipos electrónicos de multiplexación permiten combinar dos funciones: la ubicación del ancho de banda de las señales (espectro) dentro del ancho de banda del medio físico, y el tiempo durante el cual esas señales están presentes en el medio.

Seguidamente se describe cada uno de los procedimientos de multiplexación.

- El procedimiento de "multiplexación por división en el espacio" (SDM) asigna a cada señal digital un cable de salida distinto del de entrada para cambiar su ruta física, en el que dispone de todo su ancho de banda. Esto permite direccionar en el espacio todas las señales que llegan a un punto geográfico hacia otros destinos distintos. Un equipo electrónico multiplexor SDM es como una central de conmutación telefónica que encamina la señal de un abonado hacia otro cuyo número ha sido marcado.
- El procedimiento de "multiplexación por división en el tiempo" (TDM) asigna a cada señal digital un tiempo limitado de transmisión (canal TDM) en el medio físico, durante el cual dispone de todo su ancho de banda. Esto permite agrupar a un elevado número de señales lentas en el mismo medio físico. Un equipo electrónico multiplexor TDM consta de un conmutador, que va muestreando a alta velocidad los bits de las señales de baja velocidad almacenados en sus correspondientes memorias intermedias (*buffers*), y los introduce en una memoria de salida conectada a la línea de alta velocidad. Dependiendo de que los canales, o memorias intermedias de las señales digitales, se recorran siempre en el mismo orden, o en un orden estadístico basado en la existencia o no de señales en ellos, la multiplexación por división en el tiempo se divide respectivamente en síncrona (STDM, *Synchronous TDM*) o asíncrona (ASTDM, *Asynchronous TDM*) (Fig. 3.5).
- La "multiplexación por división de frecuencia" (FDM) se basa en el fraccionamiento del ancho de banda del medio físico en partes, denominadas canales (canal FDM), de tal forma que a cada canal se le asigna una señal de información distinta durante todo el tiempo. Esto es posible gracias a los procedimientos de modulación que permiten

asignar a cada señal de información una portadora analógica, cuyo ancho de banda se sitúa en un determinado canal del medio físico. Por lo tanto un equipo electrónico multiplexor FDM constará de tantos moduladores como canales tenga el medio físico, para crear las portadora analógicas correspondientes de la señales de información.

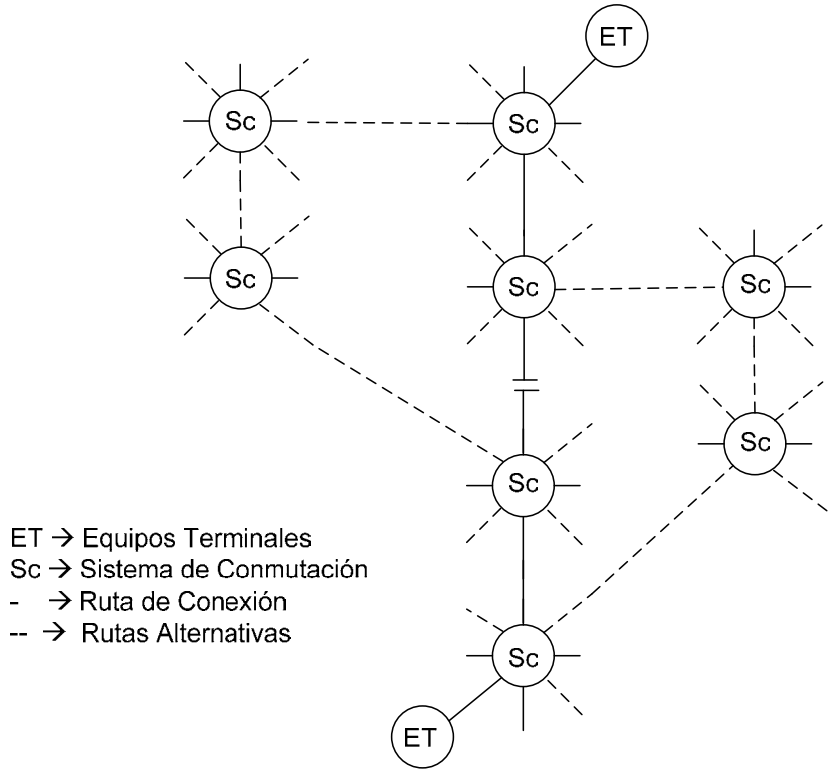


FIGURA 3.5. Ruta de conexión entre dos equipos terminales a través de una red conmutada.

Ejemplos cotidianos de "multiplexación por división de frecuencia" (FDM) son la radiodifusión y la televisión, en los que las señales de las diferentes emisoras viajan a través de un mismo medio físico, el espacio libre, por diferentes canales representados por las bandas de frecuencia de los sintonizadores (AM y FM en radiodifusión, y VHF y UHF en televisión), ubicados en los correspondientes receptores de radio y televisión de los usuarios.

Con estos ejemplos se puede comprobar la existencia simultánea de tantas señales de información en el medio físico como canales tenga éste, sin más que

utilizar un receptor para cada canal, con lo que un solo usuario podría recibir simultáneamente todas las señales presentes en el medio.

- Un caso particular de FDM es la "multiplexación por división de frecuencia ortogonal" (OFDM, *Orthogonal FDM*), en la que cada canal se divide en varias subportadoras cada una de ellas operando a una frecuencia más baja, transmitiendo sobre ellas datos en paralelo. El objetivo de la multiplexación OFDM es eliminar la superposición de los bits en el receptor (denominada "interferencia intersímbolo"), que se produce como consecuencia del retardo temporal variable debido a los múltiples caminos (multitrayecto) que la señal recorre entre el emisor y el receptor, para conseguir altas velocidades (Mbps) con buena relación señal-ruido (S/N). Según las condiciones variables del medio pueden cambiarse en cada subportadora tanto la secuencia u orden de emisión de los datos, como el tipo de modulación. Este tipo de multiplexación se aplica en redes de área local inalámbricas (WLAN) para aumentar su velocidad.
- La "multiplexación densa por división de longitud de onda" (DWDM) es un caso particular de FDM para redes de fibra óptica (apartado 3.5.3). La única diferencia es que en vez de referirse a las frecuencias de las señales electromagnéticas se prefiere utilizar sus longitudes de onda (λ), porque permiten manejar con más facilidad el espectro electromagnético en la banda óptica. El término "densa" proviene de que para canalizar las fibras ópticas las portadoras ópticas se separan entre sí mediante "canales de guarda" muy estrechos, que no transportan información, del orden de 1nm (millonésima parte del milímetro), lo que sugiere la idea de un fuerte empaquetamiento de canales en el medio físico. En portadoras ópticas situadas en la "tercera ventana de transmisión" (λ = 1552,52 nm) la separación para anchos de banda de 100 GHz es de 0,8 nm, mientras que para anchos de banda de 50 GHz es de 0,4 nm. Este tipo de multiplexación permite alcanzar elevadas velocidades de transmisión en las redes de fibra óptica. Por ejemplo, para un sistema de multiplexación DWDM con 64 portadoras, y una velocidad de 40 Gbps por portadora, se obtiene una velocidad total de 2,56 Tbps (64 veces 40 Gbps). Un equipo optoelectrónico multiplexor DWDM constará de tantos moduladores optoelectrónicos como canales tenga la fibra, para crear las portadoras correspondientes de la señales de información fotónicas (infrarrojo).
- En la "multiplexación por división de código" (CDM) se combinan los procedimientos de multiplexación TDM y FDM anteriores, junto con

las técnicas de transmisión de "espectro ampliado" (SS, *Spread Spectrum*). Los procedimientos de transmisión de espectro ampliado, como su nombre indica, expanden el ancho de banda de la señal a transmitir, ocupando todo el ancho de banda del medio físico, muy por encima del necesario para su correcta recepción, de ahí el nombre de "espectro ampliado" (apartado 3.3).

La ampliación del ancho de banda de la señal a transmitir (espectro), se realiza con otra señal independiente de la información. El exceso en el uso de ancho de banda del medio físico mejora la relación señal-ruido (S/N) en condiciones de baja potencia de emisión, aún compartiendo el medio señales potencialmente interferentes (sin licencia).

Los procedimientos de transmisión de espectro ampliado se utilizaron inicialmente en comunicaciones militares, para superar los problemas de interferencia y detección espúreas provocados por sistemas de contramedida hostiles. Su interés en las comunicaciones civiles se basa en la utilización eficiente de un medio físico, como el espacio libre, cuyo ancho de banda está muy congestionado por señales radioeléctricas distintas. La multiplexación CDM es muy utilizada en redes de área local inalámbricas (WLAN) que trabajan en la banda ISM sin licencia.

En la multiplexación CDM el receptor recupera la información correlacionando la señal recibida con una réplica de la señal utilizada para ampliar la información. La multiplexación se consigue haciendo que la señal de ampliación se genere mediante una secuencia aleatoria (código) compartida por el emisor y el receptor, de tal forma que cada usuario del medio físico posee una secuencia aleatoria distinta, y de aquí proviene el nombre de división de código.

Existen dos alternativas para generar el código de la secuencia aleatoria que crea la señal de ampliación: el salto de frecuencia (FH) y la secuencia directa (DS).

En FH, la secuencia aleatoria representa los valores de frecuencia de la portadora (FDM) que se van sucediendo en el tiempo (TDM) para comunicar emisor y receptor. El procedimiento de salto de frecuencia es comparable en recepción al "*zapping*" entre canales de TV, de tal manera que el usuario receptor va sintonizando la información que realmente quiere en distintos canales a lo largo del tiempo, pulsando a intervalos de tiempo con su telemando, como si tuviese un código que le indicase el canal al que tiene que "saltar" en un tiempo determinado.

Esta alternativa puede presentar una alta interferencia durante una pequeña parte del tiempo, en el sentido de que coincidan en un tiempo dado sobre un mismo canal (salto de frecuencia) varias señales que comparten el medio. Habitualmente en una configuración de red con multiplexación CDMA se utilizan métodos de codificación denominada "ortogonal", para que esta situación nunca pueda darse entre los terminales que la comparten. Esto hace muy útil a FH en redes inalámbricas con una alta densidad de usuarios.

Un tipo de modulación digital para FH es la de "desplazamiento gausiano de frecuencia" (GFSK, *Gaussian Frequency Shift Keying*), que es del tipo de moduladora digital y portadora analógica por desplazamiento de frecuencia (FSK), y que tiene una versión "bit a bit" y otra multibit (2 bits).

La Comisión Federal de Comunicaciones (FCC) del gobierno estadounidense, para liberar a las redes de área local inalámbricas (WLAN) en la banda ISM de la obligación de solicitar una licencia para operar, impone la condición de utilizar la multiplexación CDM con potencias de emisión inferiores a 1 W. Además, sus condiciones para la modalidad de salto de frecuencia (FH) son las siguientes:

- Ancho de banda de los canales: 1 MHz
- Tiempo de permanencia en un canal:
  - Máximo de 400 ms en 20 s (Margen de 900 MHz)
  - Máximo de 400 ms en 30 s (Margen de 2,4 GHz)
- Número de canales de salto:
  - Mínimo de 50 canales (Margen de 900 MHz)
  - Mínimo de 75 canales (Margen de 2,4 GHz)

En DS, la secuencia aleatoria representa valores de amplitud de la portadora que se van sucediendo en el tiempo para comunicar emisor y receptor. Estos valores de amplitud pueden ser discretos y son proporcionados por una combinación binaria denominada "*chip*", o infinitos dentro de un intervalo, en cuyo caso son suministrados por una señal analógica y la secuencia se denomina de "modulación de ruido".

En la opción digital de DS el emisor reemplaza cada bit de información por una secuencia de bits ("*chip*") que ocupa el mismo intervalo de tiempo. El factor por el cual el ancho de banda de la señal de información es ampliado en la transmisión se llama "ganancia de procesamiento". Se suele hacer referencia

a la "longitud de la secuencia de bits" como sinónimo de "ganancia de procesamiento", identificando ambas con el "número de bits por *chip*".

Como ejemplo de valores de la "ganancia de procesamiento", la FCC en las bandas ISM impone un mínimo de 10 bits/*chip*, y algunos sistemas comerciales utilizan 20 bits/*chip*, mientras que en sistemas militares este número puede alcanzar elevados valores de potencias enteras de 10. Cuanto más elevado es el número de bits por *chip* mayor es la inmunidad del sistema contra interferencias, y más grande la complejidad y el coste de los equipos electrónicos.

La DS puede presentar una pequeña interferencia durante todo el tiempo, debida a una señal espúrea en el medio que introduce una frecuencia dentro de la banda. No obstante su efecto en el receptor es despreciable debido a que la mayor parte de la energía de la señal de información está repartida en todo el ancho de banda del medio, como consecuencia del ensanchamiento de su espectro efectuado por la señal de ampliación. Esta característica de DS la hace muy útil si la potencia de la señal interferente es baja.

Algunos tipos de modulación digital para DS son los de moduladora digital y portadora analógica siguientes: el de "desplazamiento de fase binaria" (BPSK) y el de "desplazamiento de fase en cuadratura" (QPSK).

La multiplexación por división de código (CDM) ha introducido las siguientes ventajas en los sistemas inalámbricos digitales:

- Mayor tolerancia a las interferencias
- Reducción de los efectos perniciosos del multitrayecto como la interferencia intersímbolo
- Capacidad de compartir un ancho de banda común con otras señales

Estas ventajas permitieron multiplicar la densidad de usuarios en las redes analógicas de telefonía móvil celular (teléfonos móviles) de primera generación, por factores de 3 a 10.

La compartición del ancho de banda se consigue asignando a cada usuario de una zona de cobertura (célula) un código de ampliación distinto dentro de un conjunto "ortogonal", en el sentido de que los códigos pertenecientes al conjunto no son coincidentes, de ahí el nombre de división de código para esta técnica de multiplexación.

Históricamente la multiplexación CDM fue desarrollada por la firma estadounidense Qualcomm en 1989, y se incorporó en 1993 a la norma americana IS-95 propuesta por las asociaciones de telefonía digital CTIA/TIA (Asociación de la Industria de Telecomunicación Celular/Asociación de la Industria de Telecomunicación), cuando desde 1988 ambas asociaciones trataban de encontrar una alternativa digital a la red de datos móvil analógica AMPS (Sistema Avanzado de Teléfonos Móviles) en su proyecto DCDS (Servicio de Datos Celular Digital), dentro del cual surgió primeramente en 1990 la norma digital (segunda generación) denominada IS-54.

- Un caso particular de CDM es la "multiplexación por división de código en banda ancha" (WCDM, *Wideband CDM*), en la que las asignaciones de ancho de banda dentro de cada canal pueden ser cambiadas de forma independiente. Este tipo de multiplexación permite a los operadores de las redes reasignar anchos de banda dentro de un canal en función de las necesidades de los usuarios. Se utiliza en las normas europeas y americanas para redes de telefonía móvil digital de tercera generación. En el caso europeo esta norma de tercera generación para GSM se denomina UTRA (*UMTS Terrestrial Radio Access*).
- En el caso estadounidense existe una variante de WCDM denominada CDMA2000, que también recibe el nombre técnico de "CDMA en modo multiportadora" (MCCDMA), cuyo objetivo es implantar servicios de telefonía móvil de tercera generación sobre las redes de segunda generación IS-95 e IS-136. Esta alternativa lleva tres portadoras CDM.

3.7.3. El concepto de modulación

La modulación consiste en la modificación sistemática de alguna de las características de la señal (analógica o digital) en el medio físico, llamada "portadora", por la señal que se desea transmitir, denominada "moduladora". La palabra "modulación" procede del efecto de modificación de las características de una señal por la acción controlada de otra.

Las características mencionadas que se pueden modificar en una portadora, sea ésta una señal analógica o una señal digital, son: la amplitud, la frecuencia y la fase. Como ya se ha mencionado, en la modulación de portadoras digitales, en vez de la "frecuencia" de la onda "cuadrada" se utiliza la "anchura" que tiene esta onda en el valor alto ("1 lógico"). Cada cambio en una característica o características de la portadora se denomina "estado de modulación".

El equipo electrónico que modula a la señal portadora en el emisor se denomina "modulador". En el receptor a la señal portadora se le extrae la señal moduladora original mediante otro equipo electrónico llamado "demodulador". El equipo electrónico llamado módem, de la unión de letras de las palabras modulador y demodulador, realiza ambas funciones y de ahí su nombre.

Se distingue entre modulación en "banda base" respecto de la modulación en "banda ancha", debido a que en la primera las señales en el medio físico tienen idénticas frecuencias que las correspondientes a la información que se desea transmitir, mientras que en la segunda las señales en el medio físico tienen una frecuencia mayor. La modulación en "banda base" se utiliza en transmisión digital cuando las distancias entre los terminales de una red son cortas, y su número y velocidad de transmisión no son elevados.

Las ventajas que presenta la modulación son las siguientes:

- Reducción del tamaño de las antenas emisoras y receptoras de las radiaciones electromagnéticas
- Compartir un medio físico por varias señales de la misma frecuencia (canalización)
- Aumentar la eficiencia en el uso del ancho de banda de un medio físico
- Mejorar la relación señal-ruido (S/N) del medio físico y proporcionar a la señal en el medio físico mayor robustez frente a interferencias
- Permitir usar las redes analógicas para transmitir señales digitales (RTC)

3.8. MODULACIÓN EN BANDA BASE Y BANDA ANCHA

3.8.1. Modulación en banda base

La modulación en banda base consiste en secuencias de información binaria, y presenta las siguientes alternativas entre otras:

- Polar sin retorno a cero (NZR, *Non Zero Return*)

  El uno y el cero se representan siempre por un nivel de valor "A" y "-A", elegido de un juego de señales S1(t) y S2(t). Su inconveniente reside en la aparición de una fuerte componente de baja frecuencia.

- Manchester y Manchester Diferencial

  En la transmisión Manchester la representación del uno y el cero se realiza siempre mediante la misma transición de niveles (A y –A o –A y A), obtenida del juego de señales S'1(t) y S'2(t). En Manchester Diferencial el uno se representa sin transición inicial y el cero con transición inicial, de tal forma que la transición de niveles depende de la combinación de unos y ceros de la información a transmitir. La ventaja de estos métodos consiste en la reducción de la componente de baja frecuencia.

- Inversión de Marca Alternativa (AMI, *Alternative Mark Inversion*)

  Este tipo de modulación presenta un juego de tres niveles, dos para el uno (Marca) y uno para el cero (espacio). Por ejemplo el cero puede tener un nivel de cero Voltios [S3(t)], y el uno representa alternativamente el nivel alto [S1(t)] y el nivel bajo [S2(t)]. Su ventaja es idéntica a los tipos Manchester.

- Duobinario (*Duobinary*)

  Como la modulación AMI presenta un juego de tres niveles, dos para el uno [S1(t) y S2(t)] y uno para el cero [S3(t)]. La codificación del "uno" en la combinación binaria se realiza según la paridad par o impar de los "ceros" habidos desde el último "uno" codificado. De tal forma que si el número de ceros es impar el nivel del uno es opuesto al anterior, y si es par el nivel del uno es idéntico al anterior. Presenta el mismo inconveniente que el NZR.

3.8.2. La modulación en banda ancha

Los métodos de modulación en banda ancha dependen de que las señales portadora y moduladora sean analógicas o digitales, dando lugar a cuatro combinaciones. En las aplicaciones prácticas para modulación en banda ancha se emplean solamente tres, dos para moduladora analógica con portadora analógica o digital, y una para moduladora digital con portadora analógica.

Antes de realizar el estudio detallado de los métodos de modulación en banda ancha, se realiza su clasificación (Fig. 3.6):

- Moduladora analógica
  - Portadora analógica (AM, FM y PM)
  - Portadora digital
    - Parámetro continuo (PAM, PPM y PDM)
    - Parámetro discreto (PCM)
- Moduladora digital
  - Portadora analógica
    - bit a bit (ASK, FSK, PSK, BPSK)
    - Multibit (QPSK, QAM16/64/256)

Seguidamente se analizan los casos particulares de los métodos generales de modulación en banda ancha mencionados.

FIGURA 3.6. Señales del módem analógico (moduladora y portadora).

3.8.2.1. Moduladora analógica

Los tipos de modulación de moduladora analógica con portadora analógica son los más conocidos, debido a que fueron históricamente los primeros utilizados en la transmisión de radiaciones electromagnéticas por el espacio libre, y que los servicios de difusión para radio y televisión están en la experiencia cotidiana de la mayoría de la humanidad, desde la aparición en 1920 de la primera emisora de radiodifusión comercial, la KDKA de Pittsburgh, en Pennsylvania, USA.

La modulación de amplitud (AM), la modulación de frecuencia (FM) y la modulación de fase (PM), reciben su denominación del hecho de que las variaciones de amplitud de la señal moduladora analógica afecten respectivamente, en la portadora analógica, al cambio de su amplitud, su frecuencia y su fase.

- Modulación de amplitud (AM, *Amplitude Modulation*)

  La amplitud de la señal moduladora analógica cambia la amplitud de la portadora que también es analógica. Se utiliza en la radiodifusión comercial y su inconveniente es que duplica el ancho de banda de la señal a transmitir haciendo por ello un uso ineficaz del espectro radioeléctrico.

  Existen tres casos particulares de modulación en AM que tratan de reducir el inconveniente mencionado. La "doble banda lateral" (DBL) presenta mayor alcance debido a que no se transmite la frecuencia de la portadora. En la "banda lateral única" (BLU) se suprime una de las dos bandas de AM para reducir el ancho de banda, y se utiliza en telefonía. La "banda lateral vestigial" (BLV) realiza un filtrado de DBL para reducir el ancho de banda de la señal en el medio, pero sin suprimir totalmente una de las dos bandas laterales con objeto de no complicar excesivamente la reconstrucción de la moduladora en el receptor (aplicación en televisión).

- Modulación de frecuencia (FM, *Frequency Modulation*)

  La amplitud de la señal moduladora analógica cambia la frecuencia de la portadora analógica. Presenta mayor ancho de banda que AM, pero su ventaja es la protección contra el ruido por lo que se utiliza en radiodifusión y radioenlaces (apartado 3.3.2).

  La ventaja que representó la modulación en FM frente a la modulación en AM, en el ámbito de la radiodifusión sonora, fue la mejora de calidad en recepción. Esto es debido a que las interferencias en el espacio libre afectan más a la amplitud de la portadora que a su frecuencia.

  De esta forma, si toda la información se traslada mediante el modulador de FM en el emisor a las frecuencias de la portadora, cualquier perturbación en la amplitud la portadora durante su viaje por el espacio libre no afecta a la información transportada, y por ello el demodulador de FM en el receptor no percibe la interferencia.

- Modulación de fase (PM, *Phase Modulation*)

  La amplitud de la señal moduladora analógica cambia la fase de la portadora analógica. Sus características son similares a FM y se utiliza en la transmisión de datos a alta velocidad.

La alternativa de modulación de moduladora analógica y portadora digital presenta dos opciones, en cuanto a que los cambios en los parámetros, o características, de la portadora digital tomen valores infinitos (parámetro continuo) o finitos (parámetro discreto), en el margen de variación de amplitud de la señal moduladora analógica.

La opción de "parámetro continuo" tiene a su vez tres tipos de modulación: la modulación de impulsos en amplitud (PAM, *Pulse Amplitude Modulation*), la modulación impulsos en anchura (PDM, *Pulse Duration Modulation*) y la modulación de impulsos en fase (PPM, *Pulse Phase Modulation*), que reciben su denominación del hecho de que las variaciones de amplitud de la señal moduladora analógica afecten respectivamente, en la portadora digital, al cambio de su amplitud, su anchura y su fase (posición en el tiempo).

La modulación por impulsos codificados (PCM, *Pulse Code Modulation)* representa la opción de "parámetro discreto", y su denominación de "impulsos codificados" se debe a que cada valor de amplitud de la señal analógica moduladora está representado en la portadora digital por una combinación de bits (código) que siempre tiene un tamaño fijo, que habitualmente es de 8 bits. Con 8 bits se pueden tener hasta 256 combinaciones distintas (2 elevado a 8) en el modulador, y por esta razón la modulación se denomina de parámetro discreto, porque en el modulador solamente se pueden trasladar a la portadora digital las infinitas variaciones de la amplitud de la moduladora analógica en 256 valores (Fig. 3.7).

FIGURA 3.7. Codificación de una muestra.

Una aplicación típica de PCM es el transporte de voz digitalizada a través de redes de datos, de tal manera que un canal de voz, de ancho de banda de 4 kHz, se transmite a 64 kbps.

En PCM la señal moduladora se muestrea a una frecuencia doble de su ancho de banda para poder reconstruirla en el demodulador (Teorema de Nyquist), y a cada muestra se le asigna un nivel (cuantificación) discreto dentro del margen de valores. Como se ha dicho el número de niveles es una potencia entera de dos (2^n), de tal forma que a cada muestra se le asigna una combinación de "n" bits. Este tipo de modulación presenta la ventaja de permitir la

integración de señales distintas (voz, datos, imágenes, gráficos, fotos, etc.) en medios de transmisión de alta capacidad como las fibras ópticas. Debido a la limitación en el número de niveles de cuantificación aparece un "ruido de cuantificación", como consecuencia del error que se comete al asignar una combinación binaria a una muestra que está en medio de un intervalo entre dos valores.

Las características principales de una transmisión PCM son la velocidad (bps), la anchura máxima de los impulsos (segundos) y el ancho de banda del medio (B).

3.8.2.2. Moduladora digital

La alternativa de modulación de moduladora digital y portadora analógica presenta dos opciones, en cuanto a que la acción de la moduladora digital sobre la portadora analógica se realice "bit a bit" o con grupos de varios bits (multibit). Es decir, en el primer caso a cada "estado de modulación" de la portadora le corresponde un bit de información, mientras que en el segundo a cada "estado de modulación" de la portadora le corresponden varios bits (apartado 3.7.3).

La opción de "bit a bit" presenta dos valores de una determinada característica de la portadora analógica, uno de los cuales se asigna al "0" y otro al "1". Según la característica modificada en la portadora, se tienen a su vez tres tipos de modulación: la modulación por "desplazamiento de amplitud" (ASK, *Amplitude Shift Keying*), la modulación por "desplazamiento de frecuencia" (FSK, *Frequency Shift Keying*) y la modulación por "desplazamiento de fase" (PSK, *Phase Shift Keying*), que reciben su denominación del hecho de que los bits de la señal moduladora digital afecten respectivamente, en la portadora analógica, al cambio de su amplitud, su frecuencia y su fase. A continuación se detalla su funcionamiento y se explican algunas variantes.

- Desplazamiento de amplitud (ASK)

  En el "0" lógico la portadora presenta amplitud nula y en el "1" lógico la portadora tiene una amplitud máxima constante diferente de cero a una frecuencia fija (señal senoidal de frecuencia constante).

- Deslazamiento de frecuencia (FSK)

  Existe un juego de frecuencias de la portadora senoidal [S1(t) y S2(t)] para cada nivel lógico de la moduladora digital.

- Deslazamiento de fase (PSK)

  Para cada nivel de la moduladora digital existe un juego de fases (0° y 180°) en la portadora analógica de frecuencia y amplitud máxima constantes [S1(t) y S2(t)]. Por ejemplo el "1" lógico desfasa 0° y el "0" desfasa 180°.

- Desplazamiento de fase binaria (BPSK, *Binary Phase Shift Keying*)

  Es una variante de PSK, que no asigna cambio de fase al "0" en la portadora analógica, mientras que al "1" lo va alternando con valores de fase "0º" y "180º" en la portadora. La modulación BPSK es muy utilizada en comunicaciones por satélite por varias razones: la portadora analógica tiene una amplitud constante, puede ser transmitida a la máxima potencia, las variaciones espúreas de amplitud (por ruido, atenuación, etc.) son eliminadas en el demodulador, y realiza un uso muy eficiente del ancho de banda.

  La opción multibit presenta "N" valores de las características de la portadora analógica (amplitud, frecuencia y fase) que se asignan a grupos o combinaciones de "n" bits (siendo n mayor que 1) de la moduladora digital, de tal manera que hay una relación de una potencia entera de 2 entre los "N" valores y los "n" bits ("N" equivale a 2 elevado a "n").

  La ventaja de este tipo de modulación es aumentar la velocidad de transmisión de la portadora en el medio (bps) sin consumir ancho de banda, pero tiene como contrapartida el aumento de la complejidad de los moduladores y demoduladores en los emisores y receptores respectivamente.

La modulación multibit por "desplazamiento de fase en cuadratura" (QPSK, *Quadrature PSK*) asigna combinaciones de dos bits de la moduladora digital ("n" igual a 2), a uno de cuatro valores de fase de la portadora analógica (0º, 90º, 180º y 270º). El nombre de cuadratura proviene de que los ángulos de fase cambian 90º (valor del ángulo recto) de una combinación a otra. A los cuatro ángulos rectos del cuadrado, se añade la circunstancia de que con dos bits solamente hay cuatro posibles combinaciones, lo cual también aporta una nueva justificación a la denominación de este tipo de modulación. Las ventajas de la modulación QPSK (multibit) en las redes de satélites (apartado 3.9.3) son similares a las mencionadas para la modulación BPSK (bit a bit).

La modulación multibit por modulación de "amplitud en cuadratura" (QAM, *Quadrature AM*) asigna a las combinaciones de 4, 6 y 8 bits de la moduladora digital, valores de fase y amplitud de la portadora analógica, en las versiones QAM16, QAM64 y QAM256 respectivamente. El hecho de utilizar simultáneamente valores de amplitud y de fase de la portadora analógica para ser asignados a cada combinación de la moduladora digital, proporciona el nombre a este tipo de modulación multibit.

En las redes de operadores de cable, también llamadas "redes de cable" (apartado 3.9.1), las modulaciones multibit QAM64 y QAM256 se utilizan para la transmisión en sentido cabecera-usuario, mientras que las modulaciones multibit QPSK y QAM16 se utilizan en sentido inverso usuario-cabecera. Los "módem de cable" son los encargados de controlar estos tipos de modulación. Tanto en las redes de cable como en las de satélites, QAM presenta las mismas ventajas que las señaladas para la modulación BPSK (bit a bit) en las redes de satélites.

La necesidad de convertir las señales analógicas en digitales y viceversa, surge cuando se tienen que comunicar equipos electrónicos que manejan señales digitales a través de redes que transportan solamente señales analógicas.

Es el caso que se presenta en un PC, que genera información digital, cuando debe conectarse con una red digital, como es Internet, pero el usuario solamente dispone en su domicilio de la red telefónica conmutada (RTC) que solamente transporta señales analógicas. Para ello se utiliza un módem conectado entre el PC y la RTC, que modula las señales digitales del PC que van hacia la RTC en forma analógica, y viceversa, demodula las señales analógicas que proceden de la RTC a la forma digital que entiende el PC. El módem puede conectarse al PC mediante una tarjeta electrónica conectada internamente en una de sus ranuras de expansión, o externamente a través de un controlador de entrada/salida.

3.9. APLICACIONES DE LA MULTIPLEXACIÓN Y LA MODULACIÓN

3.9.1. Redes de cable CATV y PLC

Son redes fijas de comunicación de uso público con una cobertura superior a la de las MAN (radio entre 10 km a 100 km), y están destinadas a competir en el

mercado empresarial, profesional y doméstico con las redes de acceso de los operadores de telefonía fija (RDSI y xDSL), y servicios de distribución inalámbricos en banda ancha (MMDS y LMDS).

Si bien su mercado fundamental procede de la difusión de contenidos multimedia, como la televisión por cable CATV (apartado 3.4.1), compiten también con servicios de: acceso a Internet, transmisión de datos en banda ancha y telefonía fija. Debido a su situación competitiva es difícil su integración con otras redes de telefonía, datos y de otros operadores de CATV.

Para prestar sus servicios se basan en el "plan de frecuencias" que canaliza directamente el cable coaxial de TV dentro de su ancho de banda haciendo una partición para el canal de retorno (sentido usuario-red) de 5 MHz a 50 MHz, y en el canal de envío (sentido red-usuario) de 55 MHz a 850 MHz.

Un "módem de cable" utiliza las modulaciones QPSK (*Quadrature Phase Shift Keying*) y QAM16 en el canal de retorno (1 Mbps a 20 Mbps) para superar los problemas debidos al alto nivel de ruido e interferencias, mientras que en el canal de envío (27 Mbps a 55 Mbps) utiliza las modulaciones QAM64 y QAM256 por el bajo nivel de interferencias y la buena relación señal-ruido (S/N). Esto es debido a que el canal de envío es utilizado solamente por el operador para difundir la información hacia los usuarios, mientras el canal de retorno es compartido por todos los usuarios (apartado 3.8.2.2).

Las entidades de normalización de redes de cable son el Consorcio del Sistema de Red de Cable Multimedia (MCNS, *Multimedia Cable Network System partners*) de USA, el Consejo de Vídeo y Audio Digital y el Grupo de Difusión de Vídeo Digital (DAVIC/DVB, *Digital Audio and Vídeo Council/Digital Vídeo Broadcasting Group*) de los operadores europeos de redes de cable, y el IEEE. Estas instituciones han presentado respectivamente las siguientes normas para módems de cable: DOCSIS (*Data Over Cable Service Interface Specification),* DAVIC/DVB e IEEE 802.14.

Los "módems de cable" están diseñados para la conexión de los PC en un entorno doméstico. Un módem para recibir servicios de datos en un televisor a través de una red de cable se denomina STB (*Set-Top-Box*).

Las redes de cable suelen tener una topología en estrella, en cuyo centro está la "cabecera de la red CATV" y de ella parten los cables hacia las instalaciones de los usuarios. Su encaminamiento es del tipo de difusión. Normalmente el medio físico de transmisión utilizado hasta las instalaciones es una fibra óptica (red troncal), mientras que dentro de las instalaciones es de tipo coaxial, como

el cable de TV doméstico. Esta mezcla de ambos tipos de medios físicos recibe el nombre de red híbrida fibra-coaxial (HFC, *Hybrid Fibre-Coaxial cable).*

Como caso particular de este tipo de redes, desde el punto de vista de su configuración y tipo de servicios, cabe señalar a las basadas en las líneas de distribución de energía eléctrica (PLC, *Power Line Communications*), que utilizan multiplexación OFDM y modulación DQPSK (*Digital Quadrature Shift Keying*) para obtener velocidades de 14 Mbps (Fig. 3.8).

3.9.2. Redes de cable xDSL

Como su nombre es una red fija de uso público para telefonía analógica con una cobertura internacional (WAN) y, además de prestar el servicio telefónico convencional (RTC), está destinada a competir en el mercado del acceso a Internet y otros servicios de comunicación de datos, a través de los cables de pares telefónicos de los usuarios particulares.

El "lazo digital de abonado" (xDSL, *Digital Subscriber Line*) apareció en competencia con las redes de operadores de cable (CATV) y por ello su orientación inicial fue hacia el usuario particular.

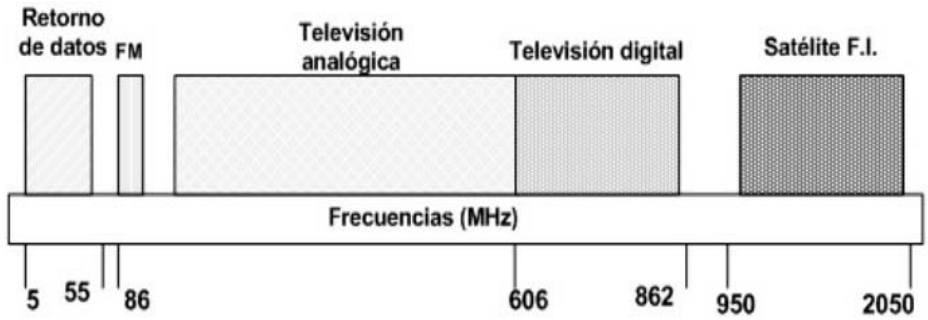


FIGURA 3.8. Propuesta de distribución de canales a través del cable.

La xDSL presenta un planteamiento distinto del par telefónico del abonado convencional, para hacer un enfoque próximo al de los operadores de las redes de cable, en el sentido de canalizar directamente el par de cobre dentro de su ancho de banda (1,1 MHz) mediante un "plan de frecuencias". En este plan se hace una partición para el canal de voz (4 kHz), otra para la comunicación usuario-red (por ejemplo de 25 kHz a 138 kHz) y el resto para la de red-usuario (por ejemplo de 200 kHz a 1,1 MHz).

Mediante el módem adecuado el usuario puede alcanzar velocidades cada vez mayores dentro del ancho de banda asignado. Estas particiones del ancho de banda del par telefónico son orientativas de una modalidad asimétrica, ya que en las modalidades simétricas de xDSL las particiones serían iguales. Las modalidades asimétricas limitan la velocidad en el sentido usuario-red frente a la velocidad red-usuario y en consecuencia son más económicas que las simétricas, ya que éstas permiten la igualdad de ambas velocidades (Fig. 3.9).

Existen diferentes modalidades de xDSL con distinto grado de implantación, aunque todas están limitadas a distancias inferiores a los 5 km, entre la instalación del abonado y la central telefónica más próxima del operador, para garantizar que el cable mantiene estables sus características de transmisión.

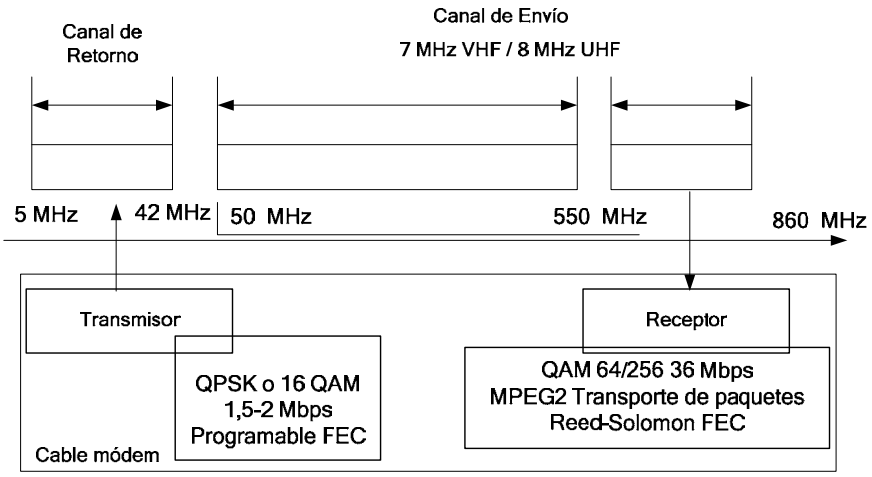


FIGURA 3.9. Funciones básicas del módem de cable.

Cada modalidad se distingue por la simetría o asimetría entre las velocidades usuario-red y red-usuario, y sus aplicaciones. A continuación se resumen sus características por este orden:

- HDSL (*High data rate DSL*): Simétrico, 1,544 Mbps a 2,048 Mbps, Acceso a servicios T1 o E1, LAN y WAN
- SDSL (*Single pair DSL*): Simétrico, 192 kbps a 2,32 Mbps, Acceso a LAN y WAN
- ADSL (*Asymmetric DSL*): Asimétrico, 640 kbps y 9 Mbps, Internet, Vídeo a la demanda, LAN, multimedia

- VDSL (*Very high data rate DSL*): Simétrico o asimétrico, 26 Mbps a 52 Mbps, como ADSL y TV alta definición

El "vídeo a la demanda" (VoD, *Video on Demand*) es un servicio que permite seleccionar un programa y controlarlo como a una cinta de vídeo en un magnetoscopio (VCR).

En xDSL existen tres opciones de acceso al lazo de abonado por otros operadores distintos del dominante (*incumbent*) denominadas respectivamente de: "acceso indirecto" si el operador entrante accede al lazo de abonado a través de la red de transporte del dominante, "acceso desagregado" (*unbundled*) cuando el operador entrante conecta su central directamente al repartidor del dominante, y de "acceso compartido" en caso de que el multiplexor del entrante esté directamente conectado al repartidor del dominante (apartado 3.4.2).

3.9.3. Redes de satélites

Una red de satélites, como cualquier otra red de comunicación, presenta unas características generales comunes a todas las redes, en cuanto a las señales transportadas y los medios físicos utilizados. Estas características se resumen a continuación:

- Topología. En una red de satélites la topología presenta los siguientes tipos generales:
  - Punto a punto
  - Multipunto
    - Estrella
    - Malla

    Debido a las variantes que pueden darse en las aplicaciones prácticas sobre estos dos tipos básicos, se han detallado otras opciones en los servicios fijos VSAT (apartado 3.6.3).
- Cobertura. Dado que una red de satélites normalmente se diseña para dar una cobertura mundial (WAN), cada satélite de la red cubre una zona de la superficie terrestre denominada "huella de iluminación", que junto a las del resto de los satélites de la red ocupan la totalidad del globo terráqueo. En las comunicaciones por satélite la antena "parabólica" es la más utilizada, tanto en recepción como en emisión, por sus especiales características geométricas. En recepción refleja las

radiaciones entrantes sobre un único punto de la antena, denominado foco, lo que permite aprovechar toda la potencia de la radiación entrante en el receptor. En emisión la radiación procedente del foco es reflejada en la antena y concentrada en un haz muy fino hacia el satélite siguiendo el eje de la parábola, lo que impide la dispersión de la potencia emitida fuera de la antena del satélite.

El diámetro de la antena parabólica en recepción, viene determinado por la "potencia isotrópica radiada equivalente" (PIRE) que presenta la huella de iluminación del satélite en la zona de cobertura. La PIRE se mide en "decibelios watio" (dBW), de tal manera que a mayor valor de dBW menor tiene que ser el diámetro de la antena parabólica, y a la inversa, cuanto menor sea la potencia de la radiación que llega a la antena más grande tiene que ser ésta para captar adecuadamente la señal enviada por el satélite.

Los dos tipos de antenas parabólicas más utilizadas son:

- Foco primario: paraboloide completo centrado sobre su eje donde está el foco
- Offset: corte lateral del paraboloide con el foco desplazado

La orientación del eje de la antena parabólica con objeto de apuntar correctamente hacia el satélite requiere el conocimiento de los siguientes ángulos:

- Azimut: entre el Norte y el eje dirigido hacia la posición en longitud del satélite. El azimut toma valores entre 0º y 360º, siguiendo el sentido de las agujas del reloj en un giro completo desde el Norte geográfico, de tal forma que en una posición geográfica dada, 90º representan el Este, 180º el Sur y 270º el Oeste. Por ejemplo, dada la posición de los satélites de la red Hispasat (1A, 1B y 1C, lanzados respectivamente en 1992, 1993 y 2000) en la órbita geoestacionaria (longitud 30º W), una antena situada en Barcelona debe tener un azimut de 223º, que representa una orientación Suroeste.
- Elevación: formado entre el plano del horizonte y el eje dirigido hacia el satélite. El ángulo de elevación toma valores desde 0º a 90º desde los polos al ecuador terrestre para una órbita GEO (apartado 3.6.2). Precisamente el bajo ángulo de elevación en las estaciones terrestres situadas en los polos (en latitudes superiores a 70º N y 70º

S), con que se ven los satélites geoestacionarios (que admiten ángulos de elevación entre 10° y 90°) introduce más porción de atmósfera entre las estaciones terrestres y los satélites, lo que atenúa más la señal y en consecuencia hay que elevar la potencia de emisión hacia las órbitas GEO.

- Encaminamiento. En las redes de satélites el encaminamiento es del tipo de difusión, de tal forma que las radiaciones electromagnéticas de un satélite llegan a todos los puntos de la superficie terrestre comprendidos dentro de su huella de iluminación, con la potencia definida por el valor de la PIRE (dBW). Las estaciones terrestres según el tipo de órbita, el valor de la PIRE, y la posición orbital del satélite tendrán que disponer de unas antenas con: el mecanismo de seguimiento adecuado, un tamaño adaptado a la PIRE, y unos ángulos de azimut y elevación que garanticen un correcto apuntamiento hacia el satélite. En el caso de constelaciones de satélites en aplicaciones de telefonía celular, pueden establecerse conexiones punto a punto entre satélites (ISL, *Inter Satellite Links*) con visión directa (LOS), para reducir retardos de transmisión.
- Multiplexación. La "multiplexación" en las redes de satélite es de los tipos TDM, FDM y CDM (apartado 3.7.2).
- Acceso al medio. El "acceso al medio" es del tipo determinista por asignación fija, y sus métodos están basados en las técnicas de multiplexación TDM, FDM y CDM, recibiendo por ello los nombres específicos de TDMA, FDMA y CDMA. En TDMA las diferentes opciones de asignación dinámica de intervalos de tiempo son:
  - Aloha ranurado. Un conjunto de intervalos de tiempo están disponibles para varias estaciones, que pueden transmitir sus paquetes de datos cuando deseen. Esto puede dar lugar a que dos o más estaciones transmitan sus paquetes en el mismo intervalo de tiempo, produciéndose una colisión lo que obliga a una retransmisión. Este método sólo se recomienda para estaciones que generen poco tráfico o que sea aleatorio.
  - Por reserva. Exige una petición previa antes de enviar la información, normalmente por el método anterior. Este método es adecuado para tráfico a ráfagas, es decir, con mensajes largos pero discontinuos en el tiempo.

  Una variante de FDMA es el acceso múltiple con asignación por demanda o DAMA (*Demand Assigned Multiple Access*), que permite

aprovechar al máximo los márgenes de frecuencia cuando el tráfico que genera cada estación terrestre es esporádico, debido a que los márgenes se asignan a las estaciones terrestres solamente durante el tiempo necesario para establecer la comunicación. Una vez finalizado el tiempo, el margen de frecuencia se libera y queda disponible para cualquiera de las estaciones que lo solicite temporalmente, estableciéndose el control de los márgenes de frecuencia mediante una de las estaciones terrestres.

En las redes de satélite el "acceso al medio" se entiende como la capacidad de que varias estaciones terrestres compartan un mismo "transpondedor" del satélite. Los equipos transpondedores de a bordo permiten canalizar el ancho de banda del satélite, considerado éste como un medio físico. Así, en un caso general, el ancho de banda de un satélite convencional es de 500 MHz, y se divide mediante 12 transpondedores en "canales" de 36 MHz, con las bandas de guarda aparte.

Por todo lo anterior un transpondedor se define como un conjunto de bloques funcionales existentes entre la salida de la antena receptora y la entrada de la antena emisora, asociados al tratamiento de las señales comprendidas en una determinada subdivisión del ancho de banda del satélite. De hecho la tarificación del coste de un servicio de satélites se basa en el tiempo de uso del transpondedor por una determinada estación. Cuanto más compartido esté un transpondedor menor es la cuota a pagar por cada estación, debido a que el coste se distribuye entre todas las estaciones.

- Bandas de frecuencia Las bandas de frecuencia utilizadas en las redes de satélites están comprendidas entre la radiofrecuencia (RF) y las microondas (apartado 3.3.2), y su denominación, margen de frecuencias, frecuencias de enlaces ascendente/descendente y servicios típicos son los siguientes:
  - L (1 GHz a 2 GHz): 1,6/1,5 GHz, MSS (*Mobile Satellite Services*), telemetría en GNSS y telefonía móvil DCS (*Digital Cellular System*)
  - S (2 GHz a 4 GHz): investigación del espacio exterior, y servicios móviles (MSS)
  - C (4 GHz a 8 GHz): 6/4 GHz, y servicios fijos (FSS, *Fixed Satellite Services*)

- X (8 GHz a 12 GHz): 8/7 GHz, teledetección, meteorología y FSS militares
- Ku (12 GHz a 18 GHz): 14/12 GHz, radiodifusión por satélite (DBS, *Direct Broadcast Satellite*), y FSS
- K (18 GHz a 27 GHz): DBS y FSS
- Ka (27 GHz a 40 GHz): 30/20 GHz y FSS

El aprovechamiento del espectro radioeléctrico en los servicios basados en satélites, hace que sea necesario reutilizar el ancho de banda disponible en ellos mediante los siguientes procedimientos:

- Aislamiento espacial
- Discriminación de polarización

El "aislamiento espacial" se basa en la utilización de antenas con capacidad de dirigir haces de radiación hacia zonas geográficas diferentes, de tal forma que si los haces están suficientemente separados entre sí pueden operar a las mismas frecuencias. La "discriminación de polarización" utiliza la característica de polarización de las señales electromagnéticas. Esta característica se añade a las ya conocidas de amplitud, frecuencia, fase, período y longitud de onda (λ) de la señal senoidal (apartado 3.7.1).

Se demuestra físicamente que en la propagación de una señal electromagnética, los campos eléctrico y magnético vibran en un plano perpendicular al sentido de la propagación, con diferentes orientaciones dentro de dicho plano (apartado 3.5.2). La orientación de la vibración de dichos campos en la propagación de una onda electromagnética se denomina "polarización" de la señal.

Las antenas con capacidad de discriminar la polarización, permiten que señales de la misma frecuencia pero con distinta polarización puedan ser transmitidas simultáneamente en el mismo haz de radiación sin interferencia.

- Modulación. La modulación en las redes de satélite es de los siguientes tipos:
  - Moduladora analógica
    - Portadora analógica (FM)

- Moduladora digital
  - Portadora analógica
    - bit a bit (BPSK, *Binary Phase Shift Keying*)
    - Multibit (QPSK, *Quadrature Phase Shift Keying*)

La ventaja que representó la modulación en FM frente a la modulación en AM, en el ámbito de la radiodifusión sonora, fue la mejora de calidad en recepción. Esto es debido a que las emisiones interferentes que actúan en el espacio libre afectan más a la amplitud de la portadora que a su frecuencia. En las redes de satélites la FM se emplea para servicios fijos analógicos como la difusión directa de señales de televisión (DBS).

Las ventajas de las modulaciones BPSK y QPSK en comunicaciones por satélite son las siguientes: la portadora analógica tiene una amplitud constante, puede ser transmitida a la máxima potencia, las variaciones espúreas de amplitud (por ruido, atenuación, etc.) son eliminadas en el demodulador, y realiza un uso muy eficiente del ancho de banda. En las redes de satélite las modulaciones digitales se utilizan en servicios fijos y móviles FSS y MSS.

Como resumen de las técnicas de multiplexación, acceso al medio y modulación en los enlaces de las redes de satélites, se presentan a continuación los siguientes ejemplos:

- Enlace FDM/FM/FDMA

  FDM significa que en la estación terrestre emisora los canales de la información analógica que se va a transmitir, por ejemplo de vídeo, voz y datos, se trasladan de su forma original o en “banda base” y se multiplexan en frecuencia o en “banda ancha”. A continuación el resultado modula en frecuencia (FM) a una portadora, y ésta accede al transpondedor del satélite por el método de FDMA.

  En la estación terrestre receptora se efectúa el proceso inverso para recuperar la información y devolverla a sus canales originales en banda base. Para ello primeramente demodula en frecuencia a la portadora, y posteriormente demultiplexa el resultado en frecuencia.

- Enlace TDM/QPSK/TDMA

  La estación terrestre emisora multiplexa en el tiempo varios canales digitales (TDM), a continuación el resultado modula digitalmente con desplazamiento de fase en cuadratura (QPSK) a la portadora, y finalmente ésta accede al transpondedor del satélite por el método de TDMA. En la estación terrestre receptora se extrae la información original efectuando en proceso inverso, es decir, demodulando primeramente a la portadora en fase y a continuación demultiplexando el resultado en el tiempo.

Las redes de satélites ofrecen servicios fijos (FSS) igual que las redes terrestres, a través de agencias internacionales como Intelsat (1964), Eutelsat (1977), Astra, y otras, y nacionales como Hispasat (1988), proporcionando: telefonía internacional, fax, télex, televisión y vídeo internacionales, DBS y servicios de comunicaciones de empresa.

3.9.4. Buses de campo

Los buses de campo (FB, *FieldBus*) son buses de transferencia de información en serie utilizados en la industria, y están orientados a la transmisión de datos, tanto en los procesos discretos (automóvil, alimentación, bienes de equipo, electrodomésticos, etc.) como en los continuos (petroquímicas, electroquímicas, metalurgia, electricidad, agua, gas, etc.), e incluso en los incorporados a productos tales como los de la industria automovilística (motores de explosión, transmisión, climatización, etc.) o de la mecatrónica (robots, manipuladores, efectores, etc.). La aparición de los buses de campo se debió primordialmente a los siguientes factores: la necesidad de reducir el cableado en las instalaciones, la mayor integración de los datos de la planta en los sistemas de información de la empresa, la creciente capacidad de proceso (inteligencia) de los sensores y actuadores, la tendencia a descentralizar los recursos de procesamiento para crear sistemas de proceso distribuido, y la demanda de los mercados de equipamiento de control (PLCs, reguladores, sensores, actuadores, máquinas CNC, etc.) con conexiones normalizadas (abiertas) que garanticen la intercambiabilidad de los productos, la interoperabilidad de los equipos, la compatibilidad de los programas diseñados y su facilidad de desarrollo.

Los buses de campo han extendido su aplicación a los entornos residencial (domótica) y de oficinas (inmótica) o edificios y áreas inteligentes, tanto para aumentar la calidad de vida de sus ocupantes, como para conseguir una mejor

gestión energética, incluyendo la disponibilidad de mayores facilidades para la realización de sus tareas habituales (ocio, trabajo, comunicaciones, etc.).

La interconexión de sistemas electrónicos en el ámbito industrial incorpora los resultados obtenidos en las redes de comunicaciones convencionales LAN, MAN y WAN, para el diseño de sistemas distribuidos que permitan la integración de los datos, procesos y equipamientos con objeto de alcanzar un uso eficiente de los recursos y una mejora continua de su productividad.

El esfuerzo de normalización en el diseño de redes de comunicaciones fue iniciado en 1978 por la Organización Internacional de Normalización (ISO, *International Standard Organization*), presentando un modelo de interconexión de sistemas abiertos (OSI, *Open System Interconnection*) que sirvió de referencia para el desarrollo de protocolos normalizados, cuya versión final es de 1984. Este modelo define la funcionalidad de una red de comunicaciones mediante siete niveles: físico, enlace, red, transporte, sesión, presentación y aplicación. Cada nivel se relaciona con el nivel superior a través de interfaces y le proporciona determinados servicios mediante protocolos específicos. La influencia de este modelo se impuso rápidamente en los desarrollos comerciales de las redes LAN, surgiendo en los años 80 normas aceptadas mundialmente, como IEEE 802.3 e IEEE 802.5, la primera a partir de la red comercial Ethernet de Xerox, y la segunda basada en la red Token Ring de IBM.

Si bien estas redes locales, sobre todo Ethernet, fueron destinadas en las empresas del ámbito industrial a desempeñar el papel de redes de oficina, para enlazar PCs, estaciones de trabajo, grandes ordenadores (mainframes) y periféricos compartidos (impresoras, unidades de almacenamiento, etc.) en tareas de gestión, en las plantas de fabricación de esas mismas empresas seguía existiendo la necesidad de incorporar las islas de automatización, que habían surgido a lo largo de los avances tecnológicos producidos a partir de los años 50, a la red de comunicación general de la empresa.

Estas islas de automatización eran a comienzos de los años 80 un mercado cautivo de los fabricantes de equipamiento industrial de todo tipo (maquinaria, sensores, actuadores, controladores, reguladores, PLCs, robots, manipuladores, etc.), de tal manera que cada equipo con idéntica funcionalidad era incompatible con el de otro fabricante y sólo se comunicaba con los de su firma. Esto en la práctica, además de las cargas económicas indirectas que imponía a la política tecnológica de la empresa (desarrollo de un software específico para cada equipo, inflexibilidad en la sustitución de los módulos de un sistema, etc.), implicaba el despliegue de una gran cantidad de interconexiones punto a punto,

y tal diversidad de pequeñas redes incompatibles que los costes de cableado y su mantenimiento eran muy importantes.

La industria automovilística, más avanzada en la automatización de plantas de productos de gran consumo y por lo tanto con grandes inversiones, pudo imponer desde el sector de las empresas clientes de tecnología industrial una norma de comunicaciones que permitiera enlazar su red de gestión con la planta industrial. La General Motors se planteó este objetivo cuando en 1980 se encontró con que en sus fábricas, en las que tenía instalados alrededor de los 20.000 PLC y 2000 robots procedentes de distintos fabricantes, solamente podían comunicarse entre sí el 15%. En 1984 la General Motors consiguió formar un grupo de usuarios con 36 compañías de la nueva norma denominada MAP (*Manufacturing Automation Protocol*), cuyas revisiones fueron: 1.0 (1982), 1.8 (1983), 2.0 (1984), 2.1 (1985) y 3.0 (1987). Esta norma presenta protocolos en los siete niveles del modelo OSI, y como aspectos más destacables se mencionan la adopción de la norma IEEE 802.4 (*Token Bus*) en el nivel de enlace, y el desarrollo de un protocolo de mensajes de fabricación en tiempo real en el nivel de aplicación denominado MMS (*Manufacturing Message Specification*), que por su éxito en la industria ha sido adoptado como EIA RS-511 e ISO 9506.

Los buses de campo representan el más bajo nivel de redes de comunicación en una planta de fabricación, y surgieron como consecuencia de las necesidades de: reducción de materiales, costes de instalación, peso y cableado, facilidad de reconfiguración de los sistemas, tiempos de transmisión reducidos, fiabilidad en la transferencia de datos, seguridad intrínseca en medios peligrosos, mejora de la señal transmitida y aumento del flujo de información. La satisfacción de estas necesidades hace que los buses de campo sean muy útiles en aplicaciones CIM (*Computer Integrated Manufacturing*).

Según la definición de la IEC (*International Electrotechnical Commission*) y la ISA (*Instrument Society of America*) un bus de campo es una conexión serie digital que permite la transferencia de datos entre elementos primarios de automatización (instrumentos de campo), que realizan funciones de medida y control, y elementos de automatización y control de más alto nivel. De acuerdo con esta definición se presenta la estructura piramidal (pirámide de la productividad) de los niveles de comunicación en una planta industrial, en la que el bus de campo ocupa la base como red local específicamente industrial y las LAN representan el nivel más alto. Los buses de campo dan cobertura a las áreas de una planta industrial (distancias inferiores al kilómetro), que tienen en

un nivel inferior las conexiones locales de equipos terminales y sistemas de tarjetas y a un nivel superior las redes de comunicación LAN, MAN y WAN.

Un bus de campo al igual que una LAN satisface los dos primeros niveles del modelo OSI (físico y enlace) y el último (aplicación), teniendo además un nivel que se ocupa de la gestión de los anteriores. No obstante presenta mensajes más cortos (órdenes, eventos, medidas, etc.), con tiempos de respuesta entre 100 ms y 5 ms (operación en tiempo real) y alta seguridad en la comunicación (fiabilidad), sobre distancias comprendidas entre los 200 m y los 2 km, a velocidades que suelen ser inferiores a 1 Mbps. Normalmente el medio físico de un bus de campo de bajo coste consiste en un par de hilos (apartado 3.4.1), con interfaz EIA RS-485, aunque se encuentran aplicaciones con coaxial (apartado 3.4.1), fibra óptica (apartado 3.5), radio (apartado 3.3.2.1), e infrarrojos (apartado 3.3.2.3). Como criterios de diseño o selección de un bus de campo se pueden considerar entre otros los siguientes: tiempo de transmisión, fiabilidad en el intercambio de datos, capacidad de operaciones seguras en áreas peligrosas, tamaño de los paquetes de datos, suministro de alimentación eléctrica, número máximo de nodos, longitud del bus, medio de transmisión, coste por nodo, capacidad de enlazar con otras redes e interoperabilidad.

Como ya se ha mencionado en la génesis de MAP, la aparición en los años 80 de redes industriales cerradas (propietarias) que habían sido desarrolladas por los fabricantes de equipamiento de campo y control, hizo sentir la necesidad de crear normas abiertas de redes de bajo coste para la comunicación entre la instrumentación de campo y los elementos de control. Esta tarea fue asumida por la IEC y la ISA a través de su comité SP50 para especificar una norma de bus de campo, y en 1992 un consorcio de empresas fabricantes de redes industriales asumieron en su proyecto ISP (*Interoperable System Project),* que en aquellas áreas donde la especificación ISA/IEC SP50 esté incompleta se utilizase tecnología probada de las normas francesa WorldFIP (desarrollada a partir de 1985) y alemana PROFIBUS (desarrollada a partir de 1987).

Las principales ventajas de los buses de campo en cuanto a su instalación y aplicación se pueden resumir en las siguientes:

- Bajo coste de su instalación y del posterior mantenimiento.
- Disminución del cableado de los recursos presentes en la planta de producción industrial.
- La escalabilidad (incremento de nodos a la red) es sencilla y económica.

- Su operación es en tiempo real, siendo apropiada su utilización en los procesos industriales caracterizados por el tiempo de vida corto de sus datos.
- Soportan un elevado tráfico de información elemental con mensajes limitados en forma de órdenes de máquinas, datos de sensores y alarmas.
- Integridad de la transmisión con baja tasa de errores.
- Posibilidad de redundancia en el medio.
- Adecuados para la aplicación en entornos severos y peligrosos.

Las características principales y más comunes entre los buses de campo en cuanto a los parámetros de comunicación son las siguientes:

- Medio físico: Cable de pares trenzados y apantallados, cable coaxial o fibra óptica.
- Topología: Bus serie.
- Tipo de modulación: Banda base (apartado 3.8.1).
- Método de acceso al medio: Fundamentalmente es centralizado por sondeo, aunque también se utilizan métodos distribuidos (paso de testigo) y probabilistas (contienda).
- Velocidad de transferencia: Suele llegar hasta 1 Mbps, aunque en la actualidad con fibra óptica se alcanzan velocidades superiores pudiendo llegar hasta los 5 Mbps entre las normalizadas. Existen dispositivos para buses de campo que alcanzan velocidades de 12 Mbps pero están aun sin normalizar.
- Distancias medias: 1000 m sin repetidores.
- Número de equipos de proceso conectables: Variable (de 10 a 100) según el medio físico utilizado.

3.10. CONCLUSIONES

Se ha desarrollado una serie de conceptos fundamentales en la instalación de redes de comunicación industriales, cuyo conocimiento permite dominar el diseño y la selección de los medios físicos que componen su infraestructura. Este capítulo incluye además la adecuada comprensión de las características del equipamiento que ha de manejarse para la correcta transmisión de información,

no solamente el que está en conexión directa con el medio concreto, sino también el que ha de ser incorporado a los equipos de usuario o terminales de acceso a la información. Dada la variedad de medios existentes en la práctica, el alumno se prepara convenientemente para distinguir las ventajas e inconvenientes en el uso de las diferentes opciones que pueden ser empleadas en un caso particular. Los casos prácticos estudiados y la variedad de ejemplos reales que se mencionan en cada apartado contribuyen a consolidar los conceptos básicos adquiridos.

3.11. BIBLIOGRAFÍA

M. Castro y A. Colmenar. Guía multimedia: Sistemas básicos de comunicaciones. Ed. RA-MA, 1999.

N. Cox, Ch.T. Manley y F.E. Chea. Guía de redes multimedia. Ed. Osborne/McGraw-Hill, 1996.

B.R. Elbert. The satellite communication applications handbook. Ed. Artech House, 1997.

J.M. Hernando. Comunicaciones móviles. GSM. Fundación Airtel, 1999.

P. Mariño. Las nuevas tecnologías de la información y las comunicaciones. Ed. FSG, 2001.

P. Mariño. Las comunicaciones en la empresa; normas, redes y servicios. Ed. RA-MA, 2003.

N. Oliva, M. Castro y otros. Sistemas de cableado estructurado. Ed. RA-MA, 2006.

W. Stallings. Comunicaciones y redes de computadores. Ed. Alhambra-Longman, 2000.

W. Stallings. Redes e Internet de alta velocidad. Ed. Alhambra-Longman, 2004.

3.12. EVALUACIÓN

3.12.1. Ejercicios resueltos

Ejercicio 3.1

Calcular la velocidad máxima de transmisión binaria (bps) en un cable de pares telefónico cuya relación señal-ruido es de 30 dB, y el ancho de banda del canal de 4 kHz.

Solución

Para aplicar la ecuación de Shannon hay que pasar la S/N del enunciado de dB a relación de potencias, por tanto:

(S/N) dB = 10.log (S Watios/N Watios)

de aquí resulta 30 dB = 10.log (S/N)

y despejando

S/N = antilog 3 = 10^3 = 1000

Aplicando la ecuación de Shannon (ecuación 3.1)

C = 4000 Hz.$\log_2$ (1 + 1000)

aproximando a un valor mayor 1001 ≅ 1024 = 2^{10}

resulta

C máxima = 4000.$\log_2$ 2^{10} = 4000.10 = 40 kbps

Ejercicio 3.2

Una fibra óptica multimodo de índice gradual, que es iluminada por un fotoemisor en la ventana de 850 nm con una potencia de 100 mW, presenta una atenuación de 5 dB/km, y un retardo de 0,5 ns/km. Calcular: a) la distancia máxima a la que debe colocarse un fotorreceptor, en el extremo de salida de la fibra, cuyo umbral de detección está en 1 mW. b) el retardo de la señal infrarroja en el fotorreceptor para esa distancia.

Solución

a) Dado que la potencia de entrada es de 100 mW (Pe) y la de salida 1 mW (Ps), la atenuación producida en la fibra en decibelios (dB) es (ecuación 3.2) de:

Atenuación = 10.log (Pe/Ps) = 10.log (100/1) = 10.log 10^2 = 20 dB

Como en cada km se pierden 5 dB, la distancia máxima (d) será:

d = 20 dB/5 dB/km = 4 km

b) Como en cada km se retrasa 0,5 ns, el retardo total (tr) será:

tr = 4 km.0,5 ns/km = 2ns

Ejercicio 3.3

Calcular los retardos de transmisión (segundos) en una red de satélites geoestacionarios (GEO) en los siguientes casos: a) telemetría sobre un satélite desde una estación terrestre b) conversación telefónica entre dos estaciones terrestres y un "salto" de satélite c) conversación telefónica entre dos estaciones terrestres y dos "saltos" de satélite con una estación terrestre intermedia.

Solución

a) Distancia aproximada del GEO a la estación terrestre: A = 36.000 km

Velocidad de la luz: C= 300.000 km/s

Tiempo de ida y vuelta de la señal electromagnética entre la estación terrestre y el GEO (Tr: Tiempo de retardo):

Tr = 2.A/C

por tanto Tr = 72.000 km/300.000 km/s = 0,25 s

b) Del resultado anterior el tiempo de salto será (Ts):

Ts = 2.Tr = 2.0,25 s = 0,5 s

c) De nuevo del resultado anterior, teniendo en cuenta que en este caso se producen dos saltos de satélite, el retardo T resultará ahora:

T = 2.Ts = 2.0,5 s = 1 s

Ejercicio 3.4

Calcular la duración (T) de una trama PCM para telefonía multicanal con 30 señales de voz y 2 de señalización, teniendo en cuenta que cada muestra está representada por una combinación binaria de 8 bits.

Solución

Un canal de voz ocupa 4 kHz, y la frecuencia de muestreo (Fm) el doble:

Fm = 2.4 kHz = 8 kHz, con lo que la velocidad de la trama es de

V = Fm.(nº canales).(nº bits/muestra) = 8000 (muestras/s).32.8 (bits/muestra)

con lo que resulta V = 2,048 Mbps.

Para esta velocidad, la duración máxima de un bit (Z) es de:

Z = 1/V = 1/2,048 Mbps = 0,488 μs

Dado que la trama tiene 32 combinaciones de 8 bits, y que cada bit dura 0,488 μs, la duración (T) de la trama PCM es en total de:

T = 32.8 (bit).0,488 (μs/bit) = 125 μs

3.12.2. Evaluación objetiva

1. La banda 6 del espectro electromagnético es:
   A) Utilizada para generar infrasonidos
   B) La de las radiaciones ultravioleta
   C) Está en el margen de las microondas
   D) Presenta frecuencias entre 300 kHz y 3 MHz

2. Una λ de un femtómetro (1fm):
   A) Está en el espectro visible
   B) Pertenece al margen de las microondas
   C) Equivale a la billonésima parte del milímetro
   D) Es de una señal de radiofrecuencia (RF)

3. Una red fija de acceso inalámbrico en banda ancha es:
   A) WLAN
   B) GSM
   C) GNSS
   D) LMDS

4. Un LIDAR es:

   A)Un SÓNAR

   B)Un electrocardiógrafo

   C)Un satélite GNSS

   D)Un RADAR en el ultravioleta

5. En un armario de cableado estructurado los latiguillos sirven para:

   A) Conectar cables entrantes y salientes del armario

   B) Cambiar conexiones entre equipos electrónicos internos y regletas

   C) Hacer conexiones removibles entre los elementos del armario

   D) Crear estanterías de diferentes alturas adaptables a los equipos electrónicos

6. La multiplexación que asigna a un canal del medio un margen de su espectro es:

   A) WCDM

   B) FDM

   C) CDM

   D) TDM

7. En la modulación FSK sucede lo siguiente:

   A) La moduladora analógica cambia su amplitud

   B) La portadora digital cambia su anchura

   C) A un estado de modulación de la portadora le corresponden varios bits

   D) La portadora analógica cambia solamente su frecuencia en dos valores

8. La modalidad xDSL simétrica con velocidades entre 1,544 Mbps a 2,048 Mbps es:

   A) ADSL

   B) SDSL

   C) VDSL

   D) HDSL

9. Un bus de campo opera en:
   A) El nivel mayor en las redes de comunicación industriales
   B) El nivel menor en las redes de comunicación industriales
   C) No existe en las redes de comunicación industriales
   D) La agrupación de todos los niveles en las redes de comunicación industriales

10. La distancia típica de trabajo de los buses de campo es:
   A) menos de 10 metros
   B) menos de 100 metros
   C) desde decenas de metros hasta algún kilómetro
   D) decenas de kilómetros

TEMA 4
ACCESO AL MEDIO

4.1. Introducción y orientación para el estudio
4.2. Objetivos
4.3. Capa de enlace: enlace lógico y acceso al medio
4.4. Estructura del mensaje
4.5. Detección de errores
4.6. Control de errores
4.7. Control de flujo
4.8. Control de acceso al medio
4.9. Evaluación de redes
4.10. Conclusiones
4.11. Bibliografía
4.12. Evaluación

4.1. INTRODUCCIÓN Y ORIENTACIÓN PARA EL ESTUDIO

La capa de enlace es la que se encarga de la comunicación fiable y eficiente entre dos equipos[1] adyacentes. En entornos industriales, fiabilidad y eficiencia son requisitos indispensables para proporcionar protocolos de comunicaciones de tiempo real, que son aquellos que garantizan determinismo temporal en el intercambio de mensajes.

La Figura 4.1 ilustra el escenario de aplicación, donde existe un medio de transmisión o red[2] compartido por dos o más equipos, que intercambian mensajes. Cuando los equipos comparten un mismo medio de transmisión, los mecanismos de control de acceso al medio, unidos a las técnicas de codificación de mensajes, de control de errores y de flujo, son cruciales para el correcto funcionamiento del sistema.

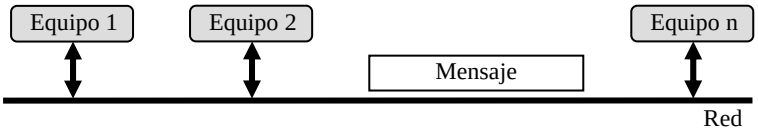


FIGURA 4.1. Escenario de aplicación.

Este capítulo se centrará en describir y analizar las políticas de control de acceso al medio, valorando su aplicabilidad en entornos industriales. Además, se revisará la forma de sincronizar y estructurar los mensajes (¿Cómo se decide si un bit pertenece a un mensaje, y cual es su significado?), los mecanismos de detección de errores (¿Cómo se detecta la alteración del estado lógico de un bit?), las técnicas de control de errores (¿Cómo se detecta la pérdida de mensajes?) y las técnicas de control de flujo (¿Cómo se evita que un equipo sobrecargue de mensajes a otro equipo?).

[1] A lo largo de este capítulo, las entidades que transmiten y reciben información (entiéndase autómatas programables, dispositivos de campo, módulos de entradas y salidas, sensores y actuadores inteligentes, etc.) son llamadas indistintamente equipos, nodos o dispositivos.
[2] Red o bus son términos usados de forma indistinta en este capítulo para designar el medio de transmisión.

4.2. OBJETIVOS

Los objetivos de este capítulo son, primero, dar respuesta a las preguntas planteadas en el apartado anterior, y, segundo, proporcionar los conocimientos y herramientas claves para poder evaluar y así determinar si una política de acceso al medio es o no adecuada para un sistema de comunicaciones industriales.

4.3. CAPA DE ENLACE: ENLACE LÓGICO Y ACCESO AL MEDIO

Las técnicas de transmisión de datos por si solas no garantizan una comunicación fiable entre equipos adyacentes, debido a la existencia de posibles errores, pérdidas de mensajes, etc. Por tanto, en cada dispositivo es necesario incluir una serie de mecanismos que garanticen la comunicación efectiva:

- Estructuración y sincronización del mensaje: los datos que se envían se denominan tramas[3]. Se deben determinar el comienzo y final de cada trama, así como su estructura y contenido.

- Detección de errores: se deben detectar los errores en la transmisión de bits.

- Control de errores: hay que detectar que las tramas enviadas lleguen a su destino, y proporcionar mecanismos para recuperar o reclamar aquellas que no lleguen correctamente.

- Control de flujo: el equipo emisor no debe sobrecargar al equipo receptor con excesivos datos. Se deben habilitar mecanismos para su control.

- Control de acceso al medio: dos o más equipos compartiendo un mismo medio de transmisión pueden tratar de enviar datos en el mismo instante,

[3] A lo largo de este capítulo, el término trama y el termino mensaje se utilizarán de forma indistinta para indentificar los datos que se envían entre equipos que comparten un mismo medio de transmisión.

produciéndose una colisión. Para evitarlo, se deben de establecer mecanismos de coordinación y arbitraje.

Tomando como referencia la torre OSI (*Open Systems Interconnection*), todos estos mecanismos que garantizan una comunicación fiable y efectiva se implementan en la capa de enlace. Para redes de área local (LAN, *Local Area Networks*) en entornos de oficina, la capa de enlace implementa estos mecanismos en dos subcapas diferenciadas: una subcapa superior llamada Control del Enlace Lógico (LLC, *Logical Link Control*) y una subcapa inferior llamada Control del Acceso al Medio (MAC, *Medium Access Control*). Para redes en entornos industriales, esta distinción generalmente no existe de forma explícita.

La capa de enlace de datos está en contacto con la capa superior, dando servició a ésta a través de primitivas específicas. Los servicios que ofrece, que se centran en la transferencia de información entre dispositivos, se pueden clasificar según la existencia de confirmación y/o conexión.

Confirmación se produce cuando el equipo receptor proporciona un acuse de recibo al equipo emisor. Por lo general, los sistemas de comunicaciones industriales, de una forma u otra, proporcionan este servicio, garantizando así fiabilidad. Por ejemplo, en Profibus DP, el intercambio más común de datos entre dos dispositivos se lleva a cabo mediante un mecanismo de pregunta/respuesta dentro de un intervalo de tiempo prefijado, con lo cual, la respuesta sirve de confirmación. Sin embargo, también existe en el mismo Profibus DP mensajería que no requiere confirmación, como la que un equipo envía a todos los otros (*broadcast*), que ni confirman ni responden.

La conexión se establece cuando se garantiza un canal lógico entre emisor y receptor, existiendo un control de flujo y errores. Este concepto es más propio de protocolos que especifican la comunicación entre equipos no adyacentes, como por ejemplo los estándares TCP/IP (*Transmission Control Protocol*, *Internet Protocol*) usados en Internet. En protocolos para entornos industriales, en operación normal, siempre existe conexión, con lo cual, no se requieren servicios específicos en modo conexión.

La aplicación de servicios orientados a la conexión y/o confirmación depende de si los protocolos son orientados a nodos o a mensajes. La Figura 4.2 ilustra estos dos tipos de protocolos. En los protocolos orientados a nodos, el intercambio de datos está basado en el direccionamiento de nodos. En estos casos, los mensajes transmitidos contienen las direcciones de los nodos origen y destino. Este tipo de protocolos, de gran implantación en entornos industriales

(como por ejemplo WorldFip o Profibus), están orientados a la conexión, con confirmación. En los protocolos orientados a mensajes, el intercambio de datos está basado en la identificación del contenido del mensaje. En estos casos, no se especifica el destinatario del mensaje, y son los nodos los que deciden si lo aceptan. Este tipo de protocolos son de gran implementación por ejemplo en los automóviles (como por ejemplo CAN).

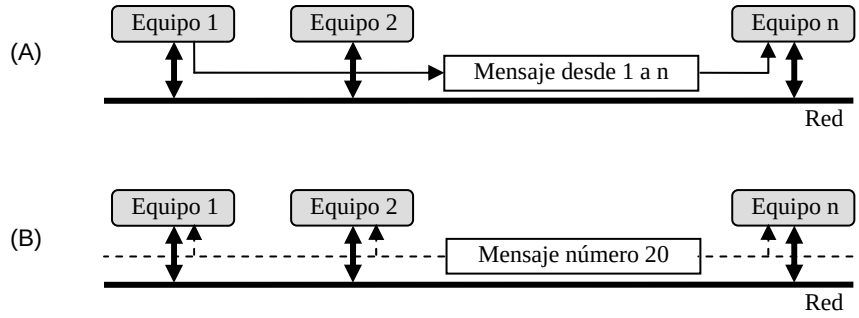


FIGURA 4.2. Protocolos orientados al nodo (A) o al mensaje (B).

4.4. ESTRUCTURA DEL MENSAJE

Los datos que se envían se denominan tramas, siendo las unidades básicas de transmisión. Una trama es una secuencia estructurada de bits, donde la estructura determina distintos campos.

Para sincronizar y determinar donde termina una trama y donde empieza la siguiente, existen varias técnicas (no excluyentes entre si):

- Por tiempo (inactividad): se utilizan intervalos de tiempo (medio de transmisión inactivo) para delimitar tramas. Por ejemplo, en el protocolo CAN (*Controller Area Network*) se establece que entre dos tramas consecutivas, el medio de transmisión tiene que estar inactivo durante un tiempo determinado conocido como Espacio entre Tramas (IFS, *InterFrame Space*)

- Por tamaño: se usa un campo de control en la trama donde se especifica el tamaño de la trama. Por ejemplo, CAN reserva 4 bits en la trama, que forman el campo de Código de Longitud de Datos (*Data Length Code*), donde se especifica el número de bytes de datos que contiene la trama.

- Por delimitadores: se añade un delimitador (bit o secuencia característica de bits) al inicio y al final de la trama. Por ejemplo, el delimitador de inicio en CAN es el primer bit que se envía, bit de Inicio de Trama (*Start of Frame*), que es un cero. Y el delimitador de final de trama (*End of Frame*) es una secuencia de siete unos consecutivos. La posibilidad de que los delimitadores aparezcan entre los datos a enviar se soluciona con las técnicas de relleno. Por ejemplo, en CAN, una secuencia de siete unos nunca puede aparecer en los datos ya que el protocolo obliga a que después de 5 bits consecutivos con el mismo estado lógico (cero o uno), se tiene que añadir un bit de relleno con el estado lógico contrario (*bit-stuffing*).

Cada protocolo establece los campos de las tramas. La estructura genérica de una trama se ilustra en la Figura 4.3.

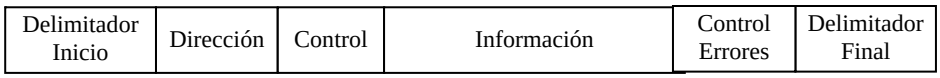


FIGURA 4.3. Estructura general de una trama.

Campos típicos son los delimitadores de los extremos (de inicio y final de trama), campos de control (como el que indica la longitud de los datos a enviar) y campo de información, que es el que contiene los datos que se desean enviar. Además, las tramas suelen tener campos para especificar las direcciones o los identificadores, así como campos que especifican el contenido de la información. Finalmente, las tramas acostumbran a tener un campo que se utiliza para detectar errores. En el apartado 4.9 se ilustran los formatos de las tramas CAN, Profibus DP y Ethernet.

El campo de información acostumbra a tener una longitud variable, de unos pocos bytes (por ejemplo en CAN es de un máximo de 8 bytes) a unos cuantos centenares de bytes (por ejemplo en Profibus DP es de un máximo de 244 bytes). Si se compara la capacidad de transportar datos con Ethernet (1500 bytes) o con tramas IP (sobre 65.535 bytes), se concluye que los protocolos para entornos industriales tienen una capacidad menor, justificada por el tipo de información que deben transportar. Cabe destacar que los protocolos de comunicaciones especifican dos tipos de tramas: tramas de datos y tramas de control. Las tramas de datos son las que transportan información en operación normal. Las tramas de control se utilizan en las etapas iniciales de configuración y parametrización, pero también en operación normal, en funciones de control de flujo y errores, o en operaciones de diagnosis.

4.5. DETECCIÓN DE ERRORES

La capa de enlace tiene que asegurar que todas las secuencias de bit recibidas sean correctas, en el sentido que todo 0 que se transmitió sea recibido como 0, y que todo 1 que se transmitió sea recibido como 1. Las alteraciones de los estados lógicos de los bits se deben principalmente a la presencia de ruido e interferencias electromagnéticas.

Existen distintas técnicas para asegurar la recepción correcta de datos. Estas técnicas, llamadas de detección y corrección de errores, están resueltas por hardware y se basan en añadir bits redundantes en los datos con el objetivo que el equipo receptor pueda comprobar si en los datos recibidos hay errores, e incluso corregirlos. La Figura 4.4 ilustra este método.

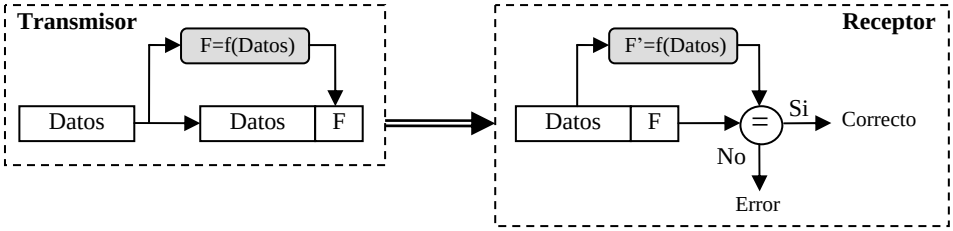


FIGURA 4.4. Detección de errores.

Estas técnicas, dependiendo de si se aplican a palabras o tramas, se clasifican en códigos de palabras o códigos de tramas. Ambas técnicas pueden ser usadas por un mismo protocolo si las tramas se transmiten por palabras.

4.5.1. Códigos de palabras

Estos códigos se basan en codificar las palabras de datos a enviar con más bits de los necesarios, añadiendo así información redundante para la detección de errores. Cada palabra de datos que se envía de k bits se codifica con n bits, siendo n>k.

Un ejemplo seria añadir el bit de paridad, de manera que la cantidad de bits a 1 (ó a 0) de la palabra enviada sea par o impar. La Tabla 4.1 muestra un ejemplo de bit de paridad par a 1 para palabras de 3 bits.

TABLA 4.1. Ejemplo de bit de paridad par a 1

Palabra a enviar	Palabra enviada
000	000**0**
001	001**1**
010	010**1**
011	011**0**
100	100**1**
101	101**0**
110	110**0**
111	111**1**

En el ejemplo anterior, si el receptor recibe 0001, conociendo el sistema de codificación, será capaz de detectar que ha ocurrido un error. Aun así, no será capaz de corregirlo, ya que, si se supone que sólo un bit ha sido alterado, las posibles palabras que podrían haberse transmitido son 0000, 0011, 0101 y 1001. En consecuencia, usando esta técnica, no sólo no es posible corregir errores, sino que sólo es posible asegurar que ha ocurrido un error si sólo se ha alterado el estado lógico de un número impar de bits.

Un concepto muy utilizado para medir la capacidad de detección (y corrección) de errores es la distancia de Hamming. Esta medida indica cuantos bits erróneos tienen que ocurrir para que ocurra un bit erróneo no detectable. Por ejemplo, una distancia de 4 significa que se pueden detectar de forma fiable hasta 3 bits erróneos. Para aplicaciones industriales, se espera que la distancia de Hamming sea de 4 como mínimo.

El concepto de distancia de Hamming aplicado a palabras que forman sistemas de codificación (como el ejemplo indicado en la Tabla 4.1) se define de la siguiente forma: esta distancia se define como el número de bits con los que difieren dos combinaciones cualesquiera del sistema de codificación. Por ejemplo, con la técnica del bit de paridad, la distancia mínima de Hamming es 2, ya que éste es el numero de bits de estado lógico distinto que existe entre cualquier combinación. Utilizando el concepto de distancia de Hamming se puede definir la condición con la cual es posible detectar errores o detectarlos y corregirlos:

- Para detectar t errores, la distancia de Hamming (d_m) entre cualquier pareja de palabras de un sistema de codificación es $d_m = t + 1$.

- Para detectar y corregir t errores, la distancia de Hamming entre cualquier pareja de palabras de un sistema de codificación es $d_m = 2t + 1$

Por ejemplo, con la técnica del bit de paridad, donde d_m es 2, utilizando la primera ecuación se ve que se podrán detectar errores de 1 bit en cada palabra. Errores de más de un bit no serán necesariamente detectables. Errores de 1 o más bits no son corregibles. De la segunda ecuación se deduce que para poder permitir la corrección de errores, hay que añadir muchos bits redundantes en cada palabra. Esto resulta muy costoso, y no se utiliza en sistemas de comunicaciones industriales.

Existen otras técnicas de detección de errores a nivel de palabras basadas en aritméticas alternativas a la del bit de paridad que pueden ser un poco más efectivas. Sin embargo, estas técnicas son buenas si la tasa de errores se distribuye uniformemente, esto es, cuando los bits erróneos aparecen de vez en cuando. Cuando los bits erróneos aparecen en ráfagas (que es el caso más realista) los códigos de palabra se muestran insuficientes, y se requieren mecanismos más potentes, como los códigos de trama.

4.5.2. Códigos de tramas

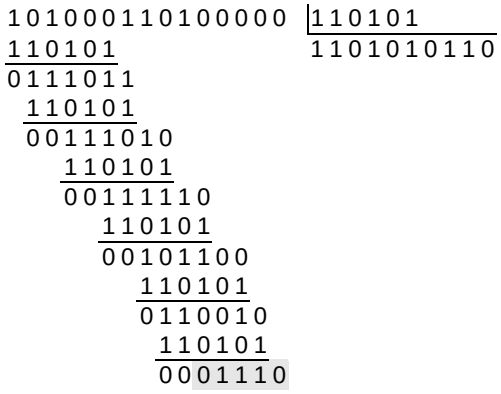
Los códigos de tramas se basan en aplicar operaciones lógicas o algebraicas a todos los bits de la trama, o a un subconjunto, obteniendo como resultado un código que se añade como campo de control de errores. Este código, generado en el transmisor, es inspeccionado en el receptor, para comprobar si se han producido errores. De los distintos códigos existentes, los más utilizados en protocolos industriales son los códigos de secuencia de comprobación de tramas (FCS, *Frame Check Secuence o checksum*) que utilizan métodos aritméticos simples, y los códigos de comprobación de redundancia cíclica (CRC, *Cyclic Redundany Check*), que utilizan aritmética polinómica.

Los *checksum* se basan generalmente en sumar el contenido de los bytes que forman la trama (excluyendo delimitadores) y aplicar al resultado una operación binaria. Por ejemplo, las tramas Profibus de configuración e intercambio de datos incorporan un campo de 1 byte que codifica el FCS, calculado como la suma de los bytes ASCII de los campos principales (desde el campo de dirección de destino, DA o *Destination Address*, al campo de datos, DU o *Data Unit*) modulo 256. Cabe destacar que Profibus, aparte de usar el FCS para detectar errores, también utiliza la técnica del bit de paridad (par a 1) en la transmisión de los bytes a través del estándar de referencia RS485. La

combinación de ambas técnicas resulta en una distancia de Hamming de 4, es decir, se pueden detectar hasta 3 bits erróneos por trama.

Los códigos de comprobación de redundancia cíclica, CRC, son verificadores de errores extensamente usados en protocolos para entornos industriales, así como para entornos no industriales. La idea básica es que dada una trama M de k bits, el transmisor genera una secuencia F de n bits (*checksum*) de tal manera que la trama resultante a enviar que es de T bits (k+n) sea exactamente divisible por un patrón predeterminado P (de n+1 bits). En recepción, se divide la trama T por el mismo patrón P, y si el resto es 0, se asume que no ha ocurrido ningún error. La secuencia F es el resto de dividir los bits del mensaje desplazados n posiciones a la izquierda por el patrón P. En la Tabla 4.2 se ilustra el mecanismo de cómputo de los códigos de comprobación de redundancia cíclica.

TABLA 4.2. Los códigos de comprobación de redundancia cíclica

Trama a enviar (M)	1010001101 (10 bits)
Patrón (P):	110101 (6 bits)
Checksum (F):	a calcular (5 bits)
Operación	
	
Trama enviada	101000110101110 (15 bits)

Una forma alternativa de ver este proceso es expresar los datos mediante polinomios de variable X con coeficientes binarios. Así, el polinomio P(x) correspondiente al patrón P anterior seria $P(x)= x^5 + x^4 + x^2 + 1$. El polinomio P(x) también es conocido como G(x), polinomio generador, que corresponde a un número primo.

Por ejemplo, CAN utiliza el polinomio generador $P(x) = x^{15} + x^{14} + x^{10} + x^8 + x^7 + x^4 + x^3 + 1$ de 16 bits, que genera un CRC de 15 bits, a los cuales se le añade un 0 para obtener un campo de comprobación de errores de 2 bytes. Este polinomio generador es eficiente para tramas de menos de 127 bits, como las de CAN. Este CRC tiene una distancia de Hamming de 6, es decir, se pueden detectar hasta 5 bits erróneos por trama. Además, es capaz de detectar errores en ráfagas de hasta 15 bits.

Existen CRC's estándares y también CRC's específicos. De forma general, los CRC's detectan todos los errores de 1 y 2 bits, todos los errores de 1 bit en un número impar de bits, y todos los errores de ráfagas de errores más pequeñas que n, así como la mayor parte de las ráfagas más grandes que n. Como más largo es el polinomio generador, mayor es la capacidad de detección de errores.

Los cálculos de checksum y CRC en el transmisor, así como su comprobación en el receptor se hacen por hardware, mediante un circuito de registros de desplazamiento y lógica de puertas XOR.

4.6. CONTROL DE ERRORES

El control de errores se refiere a la detección y corrección de errores que pueden darse en la transmisión de tramas, existen dos tipos: trama perdida, se da cuando la trama no llega al receptor; y trama incorrecta, que ocurre cuando en recepción se detectan bits erróneos. En ambos casos, es deseable la retransmisión del mismo mensaje. Los mecanismos de control de errores se pueden clasificar en pasivos o activos, según exista señalización del error, como se ilustra en la Figura 4.5.

En los mecanismos pasivos (esquemas (A) y (B), Fig. 4.5), después de la transmisión de un mensaje, el transmisor espera la confirmación del receptor durante un tiempo especificado. Si el receptor detecta un error en los datos transmitidos o estos no llegan, el receptor no transmite la confirmación. Al expirar el tiempo máximo de confirmación, el transmisor empieza la retransmisión.

En el caso de que la confirmación sea errónea (por ejemplo, por pérdida del mensaje de confirmación), hay que habilitar mecanismos para detectar en el receptor la recepción de un mensaje ya recibido. Posibles mecanismos pasan por numerar los mensajes o por etiquetar los mensajes y las correspondientes confirmaciones con 0 y 1.

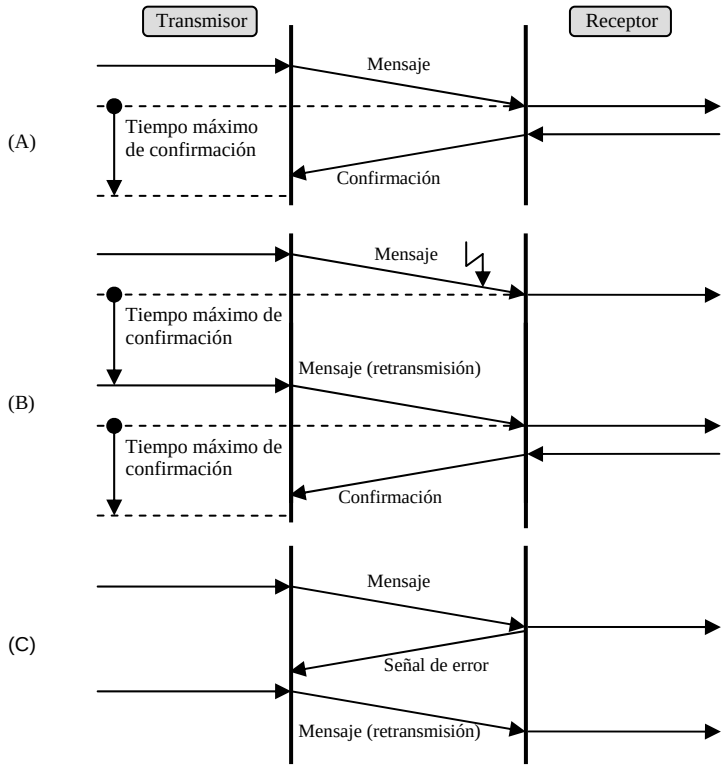


FIGURA 4.5. Esquemas de control de errores: control pasivo (A y B) y activo (C).

Con este tipo de control de errores siempre existe una transmisión confirmada de datos. Una desventaja de estos métodos es que el transmisor tiene que esperar que expire el tiempo máximo de confirmación. Además, este principio de detección de errores no se puede aplicar a mensajes que van destinados a más de un equipo, como por ejemplo los broadcast. El control de errores pasivo, que también se conoce como PAR (*Positive Acknowledgement Retransmission on timeout*) se utiliza por ejemplo en Profibus. Estos métodos pertenecen a la familia de técnicas de control de errores ARQ (*Automatic Repeat Request*) o solicitud de repetición automática, que se clasifican según si la retransmisión se dispara por ser detectado el error en el transmisor o en el receptor.

En protocolos orientados a mensajes, al ser desconocidos los receptores, el control pasivo de errores no es aplicable y se aplican técnicas de control activo de errores, mediante señalización de errores, esquema (C) (Fig. 4.5). El receptor, al detectar un mensaje erróneo, transmite un mensaje de error que será visto por

el transmisor, el cual empezará la retransmisión inmediatamente. Con estas técnicas, los tiempos de recuperación de errores son pequeños. Además, como todos los nodos ven el mensaje de error al mismo tiempo, estas técnicas favorecen la consistencia de datos en toda la red. Un ejemplo de aplicación de esta técnica se encuentra en CAN.

4.7. CONTROL DE FLUJO

El control de flujo es una estrategia para asegurar que un nodo transmisor no sobrecargue de datos a uno o más nodos receptores. Cuando un nodo recibe mensajes, estos se guardan en un espacio de memoria o buffer de recepción. Entonces, los mensajes se sacan del buffer y se procesan. Sin control de flujo se podría dar el hecho de que un receptor tuviera el buffer lleno y recibiera un nuevo mensaje, con lo cual, éste se perdería por desbordamiento (*overflow*) o un mensaje ya existente en el buffer se perdería por sobre-escritura del nuevo.

Los métodos de control de flujo, al igual que los de control de errores, se pueden dividir en pasivos y activos. En los pasivos, donde el método mas usado es el para y espera (*Stop-and-Wait*), el transmisor envía un mensaje y no enviará el siguiente hasta que no reciba una confirmación positiva. El problema de este mecanismo es que sólo un mensaje puede ser trasmitido en cada momento, forzando en el transmisor un tiempo de espera entre mensaje y mensaje. Esta desventaja se puede solucionar con mecanismos más sofisticados, en los cuales se permite la transmisión y confirmación de mensajes por grupos, como el método de ventana deslizante (*Sliding Window*). Estos métodos son apropiados para protocolos orientados al nodo, y normalmente los mecanismos pasivos de control de errores hacen también las funciones de control de flujo.

Los métodos de control de flujo activos se basan en señalizar la situación de sobrecarga. En este caso, el receptor, al detectar la situación de sobrecarga, envía un mensaje de señalización de sobrecarga. Estos métodos son apropiados para protocolos orientados al mensaje. CAN por ejemplo utiliza un método de señalización enviando un mensaje específico de sobrecarga.

4.8. CONTROL DE ACCESO AL MEDIO

El control de acceso al medio se utiliza para arbitrar cual de los transmisores de la red gana el acceso al medio de transmisión. Se utiliza para prevenir que dos o mas nodos intenten transmitir datos al mismo tiempo, previniendo que las

señales en el medio de transmisión interfieran entre si, interferencias llamadas colisiones. Los métodos de control de acceso al medio determinan el rendimiento de cada protocolo, en términos de latencia y comportamiento de tiempo real. Estos métodos se pueden clasificar según el acceso sea determinista o aleatorio (Fig. 4.6).

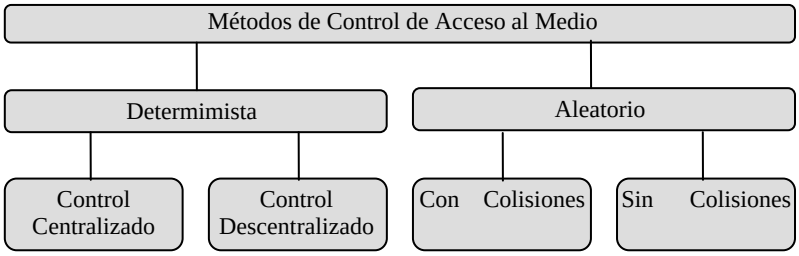


FIGURA 4.6. Clasificación de los métodos de acceso al medio

Los métodos deterministas se pueden clasificar dependiendo de si el derecho de acceso al medio se lleva a cabo de forma centralizada (a través de un maestro o unidad de sincronización) o descentralizada (de forma consensuada entre los nodos, como por ejemplo mediante el paso de testigo de nodo a nodo).

Con acceso determinista al bus, el derecho de acceder al bus está claramente definido antes de cualquier acceso al bus, eliminando la posibilidad de ocurrencia de colisiones. Además, estos métodos permiten un análisis sencillo de los tiempos de reacción del sistema. Los métodos centralizados son de fácil implementación. La desventaja principal es que si la entidad central falla, el sistema entero falla. La métodos descentralizados son más complejos de implementar, pero aportan la ventaja que son más flexibles y generalmente se mantienen operables si un nodo falla o se desactiva.

Con los métodos de acceso aleatorio, los nodos pueden utilizar el bus tan pronto como esté desocupado. Como varios nodos pueden acceder al bus al mismo instante de tiempo, siempre y cuando éste estuviera desocupado, estos métodos se llaman de Acceso Múltiple por Detección de Portadora (o CSMA, *Carrier Sense Multiple Access*). Estos métodos aleatorios se clasifican en métodos con colisiones o sin colisiones, dependiendo de si pueden ocurrir colisiones de mensajes. Los métodos con los cuales pueden ocurrir colisiones, pero éstas pueden ser detectadas se llaman de Acceso Múltiple por Detección de Portadora con Detección de Colisión (o CSMA/CD, *Carrier Sense Multiple Access with Collision Detection*). Métodos donde las colisiones son minimizadas o evitadas se llaman de Acceso Múltiple por Detección de Portadora con Evitación de Colisión (o CSMA/CA, *Carrier Sense Multiple*

Access with Collision Avoidance). En los métodos en que se evitan completamente las colisiones, los nodos tienen la capacidad de detectar el acceso simultáneo al bus durante la fase de arbitrio antes de transmitir el mensaje. Basándose en prioridades, el nodo con el mensaje más prioritario es el que se queda accediendo al bus. Estos protocolos se llaman también de Resolución de Colisiones (CSMA/CD/CR, *Carrier Sense Multiple Access with Collision Detection and Collision Resolution*) o de Arbitraje por Prioridad de Mensaje (*CSMA/CD/AMP, Carrier Sense Multiple Access with Collision Detection and Arbitration Message Priority*).

En los sistemas de comunicaciones con métodos aleatorios con colisiones, el acceso simultáneo al bus se detecta después de una longitud de trama específica o después de la evaluación del código de detección de errores. Un nodo, tras detectar la colisión, la señaliza al bus, y la situación conflictiva se resuelve con una estrategia estadística de espera antes de retransmitir el mensaje.

Los métodos de acceso aleatorio con colisiones tienen un tiempo de acceso al bus no determinista. Para sistemas con métodos sin colisiones, los tiempos de acceso al bus tampoco son deterministas pero si que se pueden calcular las latencias de todos los mensajes en el peor caso. Así, se puede garantizar que los mensajes llegaran en un plazo especificado, ofreciendo garantías de tiempo real.

Los siguientes métodos de acceso al bus son utilizados en sistemas de comunicaciones industriales:

- Control centralizado de acceso al bus mediante una entidad central (maestro) que consulta (*polling*) los nodos del bus (esclavos) siguiendo un orden establecido (maestro/esclavo o principio de consulta).

- Control centralizado de acceso al bus mediante una entidad central (maestro) con un principio de paso de testigo delegado.

- Control descentralizado de acceso al bus mediante paso de testigo.

- Acceso al bus por coordinación de ranuras temporales (TDMA, *Time Division Multiple Access*) o frecuenciales (FDMA, *Frequency Division Multiple Access*).

- Acceso aleatorio al bus mediante CSMA/CD

- Acceso aleatorio al bus mediante CSMA/CR/CA

4.8.1. Maestro/Esclavo

Con el principio maestro/esclavo, un nodo de la red, el maestro, asume el control de acceso al bus en términos de un intercambio de datos orientado al nodo. El nodo de forma cíclica consulta a los otros nodos o esclavos. El maestro puede transmitir datos al nodo destino con una trama de consulta (o petición) y recibir datos del nodo consultado en una trama de respuesta, tal y como se ilustra en la Figura 4.7.

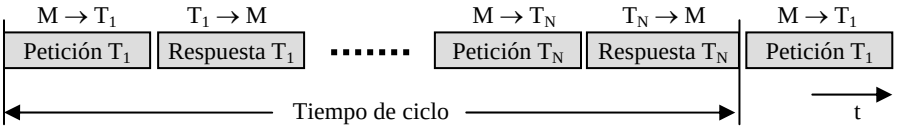


FIGURA 4.7. Acceso al bus por el principio maestro/esclavo.

El tiempo máximo de latencia, que es el tiempo máximo que puede pasar hasta que un esclavo puede transmitir un nuevo mensaje (que en el peor caso es el tiempo de ciclo), viene determinado por la longitud de las tramas, el tiempo de bit (o velocidad de transmisión) y por el número de esclavos. Para obtener latencias cortas, se requieren velocidades de transmisión altas y pocos esclavos.

Las ventajas de este principio son su simplicidad y el determinismo sobre el tiempo máximo de latencia después del cual el bus está disponible para un nodo. Una primera desventaja de este método es que un nodo esclavo es consultado sin tener en cuenta si éste requiere o no acceso al bus. Esto provoca una carga innecesaria en el bus, que se agrava particularmente para los sistemas donde la generación de datos es aleatoria. Pero para sistemas cíclicos, por ejemplo la actualización de las imágenes de proceso de un autómata, es un método muy apropiado.

Una segunda desventaja de este método es que cuando el maestro falla, todo el sistema falla. Este problema se soluciona con técnicas de replicación de maestro. Y una tercera desventaja es que el principio maestro/esclavo sólo permite relaciones de comunicación de 1 a N. Esto significa que el intercambio de datos entre esclavos tiene que pasar forzosamente por el maestro. Para solucionar este problema, algunos protocolos dividen el tiempo de ciclo en dos subciclos. En el primer subciclo se lleva a cabo las transmisiones siguiendo el principio de maestro/esclavo, y en el segundo subciclo, se permiten transmisiones de esclavo a esclavo. Por ejemplo, Profibus DP se basa en el

principio maestro/esclavo, pero también permite otros modelos de comunicación dentro del tiempo de ciclo. Otros protocolos que utilizan este principio son Modbus o AS-I.

4.8.2. Paso de testigo delegado

En este método, la entidad central (maestro o árbitro del bus) induce a los nodos (productores) a transmitir los mensajes (variables) requeridos utilizando mensajes de petición de acuerdo con una planificación temporal predefinida.

Los mensajes transmitidos pueden ser aceptados por los otros nodos (consumidores). El maestro delega el acceso de los otros nodos al bus transmitiendo un mensaje de petición (testigo) que es idéntico al identificador del mensaje requerido o variable. Una vez la variable ha sido transmitida por el nodo, el derecho de acceso al bus vuelve a ser del maestro.

El maestro consulta las variables siguiendo una planificación temporal prefijada. En consecuencia, el sistema tiene un comportamiento determinista, productor/consumidor orientado al mensaje, tal y como se ilustra en la Figura 4.8 (donde A corresponde al árbitro). Las latencias de los mensajes son específicas de las aplicaciones, y dependen de la planificación temporal.

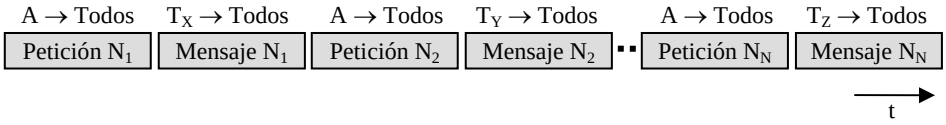


FIGURA 4.8. Acceso al bus por testigo delegado.

Este método corresponde a un sistema distribuido de mensajes controlado de forma centralizada.

El intercambio de datos entre nodos de forma arbitraria es posible ya que cada nodo puede enviar un mensaje a todos los otros nodos, con la ventaja de tener garantizado el determinismo temporal. Una desventaja de este método es la robustez: si el árbitro falla, todo el sistema falla. Se pueden utilizar técnicas de redundancia para solucionar este problema. Protocolos como WorldFip o ControlNet utilizan este principio.

4.8.3. Paso de testigo

El principio de paso de testigo (*token passing*), ilustrado en la Figura 4.9, se basa en que el derecho de acceder al bus se pasa de nodo a nodo a través de la transmisión de un mensaje específico llamado testigo. Cuando un nodo recibe el testigo, realiza intercambio cíclico de mensajes de datos durante un periodo de tiempo limitado (tiempo de posesión del testigo, *token holding time*). Cuando expira el tiempo de posesión del testigo, o cuando el nodo no tiene más datos a enviar, el testigo ha de ser enviado al siguiente nodo lógico. El principio de paso de testigo es relativamente complejo cuando incluye mecanismos dinámicos de inclusión y exclusión de nodos. Además, se requieren mecanismos complejos para establecer el anillo lógico (ordenación de la secuencia de nodos que reciben el testigo), monitorizarlo o reinicializarlo en caso de pérdida del testigo.

Cada nodo participando en el paso de testigo puede intercambiar mensajes con cualquier otro nodo. El tiempo de posesión del testigo y el número de nodos determinan el tiempo de latencia para estos sistemas.

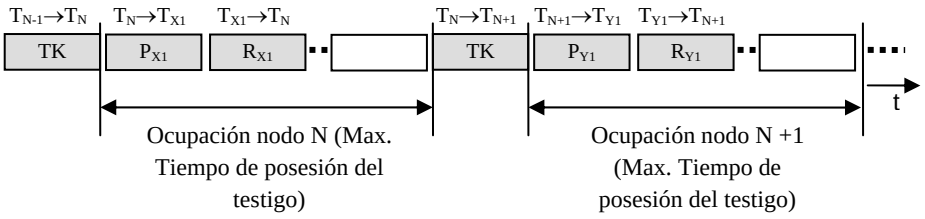


FIGURA 4.9. Acceso al bus por paso de testigo (TK: testigo, P: petición y R: respuesta).

El principio del paso de testigo se utiliza por ejemplo en Profibus DP cuando el sistema tiene más de un nodo maestro. En este caso, la ordenación entre maestros se lleva a cabo por un paso de testigo en anillo lógico.

4.8.4. Acceso múltiple por división de tiempo (TDMA)

El método de acceso múltiple por división del tiempo se basa en reservar ranuras temporales (*slots*) exclusivas a los nodos, siguiendo una planificación temporal predefinida (Fig. 4.10). A diferencia del método de testigo delegado, una entidad central no consulta a los nodos. Se les asigna el bus en base a instantes predefinidos. En cada ranura los nodos pueden enviar hasta un

máximo de mensajes, y si no tienen más datos que enviar, el espacio sobrante se desperdicia (aunque existen variaciones de TDMA donde la longitud de las ranuras se adapta dependiendo de su utilización).

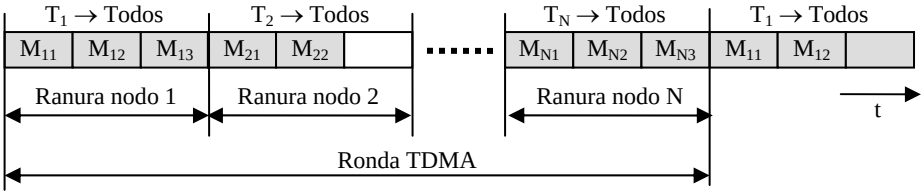


FIGURA 4.10. Acceso al bus por TDMA.

Dependiendo de cómo se planifique, los mensajes en cada ranura se repiten de ronda en ronda, o después de varias rondas, formando un hiperciclo. Los mensajes pueden ser aceptados por cualquier nodo del bus, resultando en un sistema productor/consumidor sincronizado orientado al mensaje.

Un prerrequisito de este sistema es que se requiere una sincronización perfecta entre los relojes locales de cada nodo. Esto se puede conseguir mediante mensajes de sincronización enviados por una entidad central o vía re-sincronización de los nodos en base al conocimiento del los instantes de tiempo de determinados mensajes. El tiempo de latencia queda determinado por el número de nodos, número de mensajes y velocidad de transmisión. Ejemplos de utilización de TDMA son WorldFip o *Time Triggered Protocol* (TTP). Variantes del TDMA reservan ranuras para transmisiones asíncronas, dando más flexibilidad al protocolo.

La misma filosofía de división de la capacidad del canal entre todos los nodos se encuentra en los protocolos que siguen el principio de acceso múltiple por división de frecuencia (FDMA). En este caso, la división no es por ranuras temporales, sino por bandas de frecuencias. Pero estos métodos son más aplicables a la capa física que a la capa de enlace. Por ejemplo, sistemas de comunicación sin cable como Bluetooth o Zigbee utilizan variantes de esta técnica. Con estos métodos se divide el ancho de banda en canales, y la aplicación de técnicas de espectro extendido permite que cada transacción se realice en una banda de frecuencia distinta (escogida de forma pseudo-aleatoria), minimizando la probabilidad de colisiones.

4.8.5. Acceso múltiple por detección de portadora (CSMA)

Con estos métodos, cada nodo tiene el mismo derecho de acceder al bus tan pronto como esté inactivo. Debido a la simplicidad de estos mecanismos, no se desperdicia ancho de banda en funciones de control de acceso al bus (por ejemplo, las consultas en los sistemas maestro/esclavo o las transmisiones del testigo en los sistemas basados en paso de testigo). Además, estos métodos son muy apropiados para sistemas basados en mensajería disparada por eventos, ya que sólo se realiza transmisión de datos cuando realmente es necesario. La principal desventaja es la ocurrencia de colisiones. Su tratamiento desemboca en una familia de protocolos, donde los más extendidos son el CSMA/CD y CSMA/CR/CA.

4.8.5.1. CSMA/CD

En protocolos orientados a nodos, con nodos con igual derecho de acceso al bus, en principio, las colisiones no se pueden evitar. Como consecuencia, hay que poder detectar las colisiones (*Collision Detection*, CD), señalar, y resolver utilizando mecanismos de resolución de colisiones. Las colisiones se resuelven introduciendo en los nodos tiempos de espera aleatorios antes de la retransmisión. Este método es problemático para aplicaciones de tiempo real ya que no se puede garantizar el acceso al bus. El protocolo más representativo de este tipo de acceso es Ethernet, que se usa en los niveles altos de la automatización, donde los requisitos de tiempo real son menos estrictos. La Figura 4.11 ilustra el principio CSMA/CD con dos nodos transmitiendo simultáneamente en el bus. El nodo 1 es el primero en detectar la colisión y la señala transmitiendo una secuencia de alerta (llamada *Jam Signal*). Los dos nodos se retiran del bus y cada uno espera un tiempo aleatorio específico antes de retransmitir. En el ejemplo, como el tiempo de espera para el nodo 2 es más corto, puede acceder al bus sin interferencia del nodo 1.

En general, el tiempo de acceso al bus no se puede especificar de forma determinista. Los tiempos de latencia de los mensajes, cuando el bus no está muy cargado, son cortos. Cuando el bus está muy cargado, el sistema se puede desestabilizar debido a la alta probabilidad de colisiones.

Debido al bajo coste de los controladores Ethernet, a su alta velocidad de transmisión, y a la disponibilidad de cables y conectores específicos para entornos industriales, cada vez hay más tendencia a utilizar Ethernet en aplicaciones de campo. Estos protocolos se denominan genéricamente Ethernet

Industrial. Como Ethernet sólo define la capa física y de enlace, se acostumbra a utilizar con el protocolo TCP/IP, que al estar disponible en la tecnología PC, proporciona muchas ventajas.

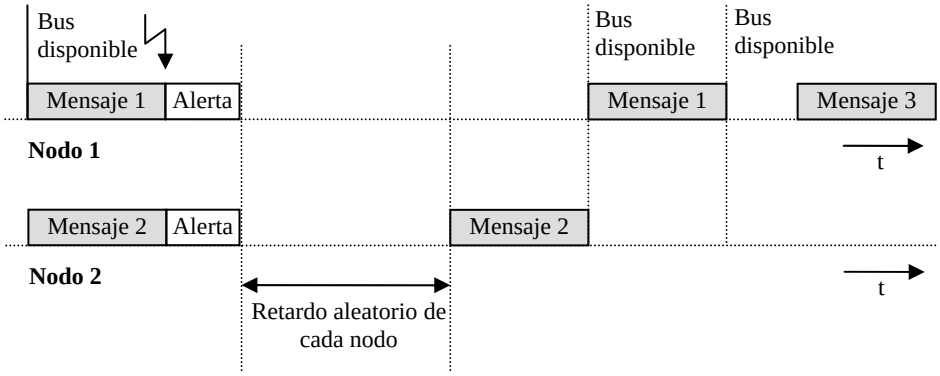


FIGURA 4.11. Acceso al bus por CSMA/CD.

Para aplicaciones industriales de tiempo real, hay distintos mecanismos que se pueden usar para mejorar el determinismo de Ethernet. Una posibilidad es incorporar un protocolo maestro/esclavo por encima de Ethernet, como Modbus sobre Ethernet. En este caso, los esclavos sólo responden a las peticiones del maestro, con lo cual, se controla el tráfico y se eliminan las colisiones. La desventaja de este mecanismo es que se elimina la característica multi-maestro de Ethernet. Otras posibilidades para aumentar el determinismo se basan en utilizar altas velocidades de transmisión o controlar la carga del bus, con lo que se reduce la probabilidad de colisiones, o utilizar la tecnología de conmutadores (*Switched Ethernet*), que puede eliminar la posibilidad de colisiones al separar los dominios de colisión. Más mecanismos pasarían por la introducción de prioridades, o de ranuras temporales.

4.8.5.2. CSMA/CR/CA

Con los métodos CA (*Collision Avoidance*) se evita la colisión de mensajes. Aun siendo posible que dos o más nodos empiecen a transmitir mensajes simultáneamente, se garantiza que durante una fase de arbitraje, sólo un nodo, el que transmite el mensaje de más alta prioridad, se quedará en el bus al finalizar esta fase. Durante la fase de arbitraje, los nodos van transmitiendo el identificador del mensaje bit a bit, al mismo tiempo que comprueban si el bit en el bus se corresponde con el bit que están transmitiendo. En caso negativo, el nodo sabe que otro nodo con un mensaje de más alta prioridad también está

transmitiendo, con lo cual, el de menos prioridad para la transmisión y se convierte en receptor del mensaje de más alta prioridad que se está transmitiendo. Un ejemplo de este tipo de acceso el medio se encuentra en el protocolo CAN.

4.9. EVALUACIÓN DE REDES

Los protocolos de comunicaciones para entornos industriales tienen que proporcionar latencias deterministas en el tiempo de transmisión de mensajes. Los factores que determinan estas latencias y permiten evaluar los distintos protocolos son la estructura de las tramas, la velocidad de transmisión (y de propagación), el número de nodos y el método de control de acceso al medio.

La codificación de las tramas determina el rendimiento del sistema de codificación. Por rendimiento se entiende la relación entre el número de bits útiles (de información) enviados entre el total de bits enviados (ecuación 4.1).

$$\text{Rendimiento} = \frac{\text{Bits utilies}}{\text{Bits enviados}} \tag{4.1}$$

La Figura 4.12 ilustra la codificación de las tramas de intercambio de datos en los protocolos CAN, Profibus DP y Ethernet. CAN puede enviar un máximo de 8 bytes de datos y requiere enviar 47 bits de información adicional. Profibus puede enviar un máximo de 244 bytes de datos, requiriendo el envió de 88 bits adicionales. Finalmente, Ethernet puede enviar un máximo de 1500 bytes de datos (y un mínimo de 46 bytes), requiriendo el envío de 168 bits adicionales.

SOF	ID	RTR	IDE	RB0	DLC	Data	CRC	DEL	ACK	EOF	IFS
1 bit	11 bits	1 bit	1 bit	1bit	4 bits	0..8 bytes	15 bits	1 bit	2 bits	7 bits	3 bits

SD	LE	LEr	SD	DA	SA	FC	DSAP	SSAP	Data	CRC	DEL
1 byte	1 byte	1 byte	1 byte	1 byte	1 byte	1 byte	1 byte	1 byte	0..244 bytes	1 byte	1 byte

PRE	DA	SA	L/T	Data	FCS
8 bytes	6 bytes	6 bytes	2 bytes	46..1500 bytes	4 bytes

FIGURA 4.12. Formato de tramas CAN, Profibus DP y Ethernet (de arriba abajo).

La Tabla 4.3 muestra una comparativa del rendimiento de las tramas de estos tres protocolos dependiendo del número de bytes a enviar[4]. La Tabla 4.3 muestra que para aplicaciones donde hay que transmitir pocos datos, el protocolo CAN proporciona el mejor rendimiento. Si el número de datos incrementa, Profibus DP o Ethernet son los que obtendrán un mejor rendimiento.

TABLA 4.3. Comparativa de rendimiento

Datos (bytes)	1	8	46	244	1500
CAN	15%	58%	58%	58%	58%
Pofibus DP	8%	42%	80%	95%	95%
Ethernet	-	-	68%	92%	99%

Cabe destacar que esta métrica se tiene que considerar juntamente con el número de nodos y el método de control de acceso al medio, y las velocidades de transmisión y propagación que determinan los tiempos de transmisión y propagación.

El tiempo de propagación se define como la distancia recorrida dividida entre la velocidad de propagación (que en medios guiados se puede aproximar por $v_p = 2.10^8$ m/s y en medios no guiados por $v_p = 3.10^8$ m/s). Este tiempo, en entornos industriales, debido a las cortas distancias que recorren las señales, es despreciable frente al tiempo de transmisión. El tiempo de transmisión se define como el número de bits a enviar dividido por la velocidad de transmisión especificada en bits por segundo (que por ejemplo en CAN puede ser de hasta 1 Mbps, en Profibus DP de hasta 12 Mbps, y en Ethernet de 10, 100 o 1000 Mbps). En la Figura 4.13 se ilustran estos tiempos durante una transacción de pregunta/respuesta entre un nodo emisor y un nodo receptor.

Partiendo de estos tiempos, y con los parámetros específicos de cada protocolo, se pueden realizar análisis más completos de los métodos de control de acceso al medio. A continuación[5] se explican dos ejemplos donde se analiza el tiempo de latencia entre el envío de dos mensajes desde un mismo nodo:

[4] Esta comparativa no tiene en cuenta los bits de relleno que los mecanismos de transmisión puedan añadir a las tramas enviadas.

[5] En ambos ejemplos se asume un tiempo de propagación despreciable, así como un tiempo de procesado también despreciable.

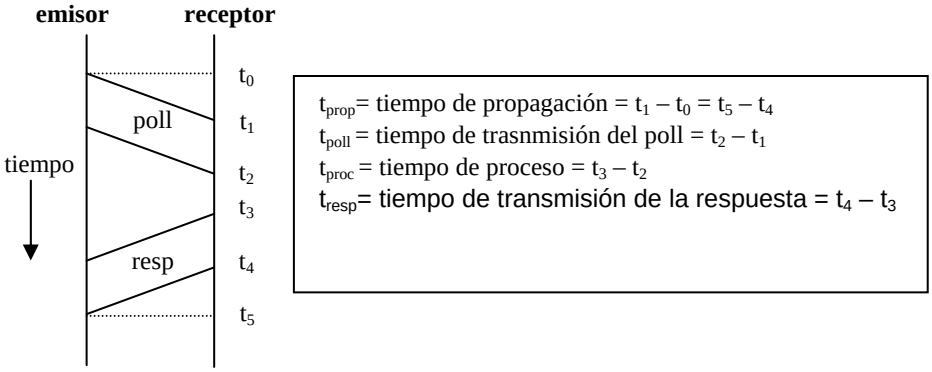


FIGURA 4.13. Análisis de los tiempos en una transacción.

- En un protocolo TDMA con N nodos (donde cada ranura permite la transmisión de un mensaje de X bytes), dada una velocidad de transmisión v_t, el tiempo de latencia entre mensajes de un mismo nodo viene dado por:

$$T_{latencia} = N \cdot (8 \cdot X / v_t) \tag{4.2}$$

- En un protocolo maestro/esclavo con N nodos (donde cada nodo envía una petición de X bytes y cada esclavo responde con un mensaje de Y bytes, dada una velocidad de transmisión v_t, el tiempo de latencia entre mensajes de un mismo esclavo viene dado por:

$$T_{latencia} = N \cdot (8 \cdot (X + Y) / v_t) \tag{4.3}$$

Este tipo de análisis se puede extender a analizar el rendimiento del método de control de acceso al medio (relación de bits útiles dividido por el total de bits enviados, a nivel de transacciones).

4.10. CONCLUSIONES

Los mecanismos de acceso al medio de los protocolos de comunicaciones industriales determinan el área de aplicabilidad. En este capítulo se han descrito y evaluado los mecanismos de acceso al medio mas utilizados en protocolos industriales. Para aplicaciones con restricciones temporales de tiempo real (como por ejemplo automatización de líneas de producción o sistemas

electrónicos del automóvil), los protocolos tienen que basarse en mecanismos de acceso al medio que ofrezcan servicios de comunicación deterministas. Para aplicaciones con restricciones temporales no estrictas (como por ejemplo sistemas de supervisión y monitorización), los protocolos tienen que proporcionar una alta capacidad de respuesta. En todos ellos es necesario habilitar mecanismos de detección y control de errores, así como mecanismos de control de flujo, para garantizar una comunicación fiable y eficiente entre los dispositivos.

4.11. BIBLIOGRAFÍA

W. Stallings. Data and computer communications. Ed. Prentice Hall, 2007.

R. Zurawsky. Special issue on industrial communication systems. Proceedings of the IEEE, vol. 93, nº 6, Junio de 2005.

4.12. EVALUACIÓN

4.12.1. Ejercicios resueltos

Ejercicio 4.1. Lógica para la generación/detección del bit de paridad

La generación del bit de paridad en el transmisor, así como su comprobación en el emisor, se realiza por hardware. La Figura 4.14 muestra la lógica digital para generar y detectar el bit de paridad de una palabra de 2 bits.

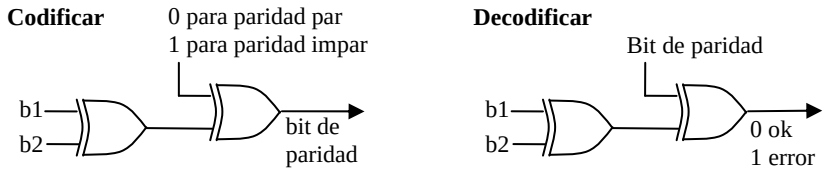
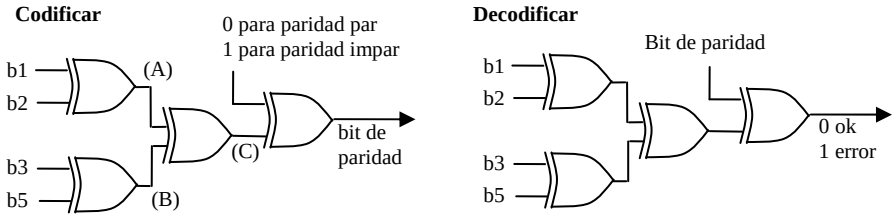


FIGURA 4.14. Codificación y decodificación del bit de paridad (palabras de 2 bits).

Diseñar la lógica digital para codificar y decodificar el bit de paridad de una codificación formada por palabras de 4 bits, y demostrar su correcto funcionamiento.

Solución

La Figura 4.15 muestra el circuito y la tabla lógica que demuestra el funcionamiento.



b1 b2 b3 b4	(A)	(B)	(C)	Par	Bit de paridad
0 0 0 0	0	0	0	0	0
0 0 0 1	0	1	1	0	1
0 0 1 0	0	1	1	0	1
0 0 1 1	0	0	0	0	0
0 1 0 0	1	0	1	0	1
0 1 0 1	1	1	0	0	0
0 1 1 0	1	1	0	0	0
0 1 1 1	1	0	1	0	1
1 0 0 0	1	0	1	0	1
1 0 0 1	1	1	0	0	0
….					

FIGURA 4.15. Codificación y decodificación del bit de paridad (palabras de 4 bits).

Ejercicio 4.2. Detección de errores en Modbus-ASCII

Las tramas en Modbus ASCII pueden ser de hasta 525 octetos, correspondientes al formato de la Figura 4.16 (donde cada carácter es de 8 bits).

Delim. I1	Dir. Dest.	Función	Datos	LRC	Delim. F1	Delim. F2

Delim.I1	Carácter sincronización	1 carácter	:	3AH
Dir. dest.	Dirección esclavo	2 caracteres	H	L
Función	Código función	2 caracteres	H	L
Datos	Información	2*N caracteres		
LRC	Código detect. errores	2 caracteres	H	L
Delim.F1	1er delimitador final	1 carácter	CR	0DH
Delim.F2	2º delimitador final	1 carácter	LF	0AH

FIGURA 4.16. Análisis de los tiempos en una transacción.

El nodo prepara la trama en un registro de memoria, empezando por la dirección (Dir. Dest.) y terminando con el código de detección de errores (LRC o *Longitudinal Redundancy Check*). Este se calcula sumando el contenido del registro por octetos, y dejando el resultado de la suma en complemento a 2 en otro octeto (aprovechando los 8 bits de menor peso). Dado el siguiente mensaje: dirección esclavo 6, código de función 5 (forzar un bit a 1 o 0), y datos: 173 (dirección del bit, en hexa 00AC) y FF00 para indicar que se fuerza a 1, calcular el LCR.

Solución

El cálculo del LCR se detalla a continuación.

Dirección	0	0	0	0	0	1	1	0
Código función	0	0	0	0	0	1	0	1
Datos	0	0	0	0	0	0	0	0
	1	0	1	0	1	1	0	0
	1	1	1	1	1	1	1	1
	0	0	0	0	0	0	0	0
Suma	1	0	1	1	0	0	1	0
Suma	1	0	1	1	0	0	1	0
Complemento a 1	0	1	0	0	1	1	0	1
Complemento a 1	0	1	0	0	1	1	0	1
+1	0	0	0	0	0	0	0	1
Complemento a 2	0	1	0	0	1	1	1	0

El LRC resultante es 4Eh.

Ejercicio 4.3. Control de errores

El autómata que se implementa en el transmisor para el control pasivo de errores (ilustrado en la Figura 4.5 (A) y (B)) con numeración binaria de tramas se muestra en la Figura 4.17. Las tramas que se envían van numeradas 0, 1, 0, 1,…

El funcionamiento del autómata es el siguiente:

- Transmisión-0: se envía la trama d0, activando el temporizador y se pasa al estado Confirmación-0?

- Confirmación-0: se espera la respuesta del receptor. Una confirmación 0 implica una correcta recepción de la trama y se pasa al estado

Transmisión-1. Si no se recibe nada y expira el temporizador (*Timeout*), se pasa al estado de Retransmisión-0.

- Retransmisión-0: se reenvía la trama d0, activando el temporizador, y pasando al estado de Confirmación-0?

- Transmisión-1: se envía la trama d1, activando el temporizador y se pasa al estado Confirmación-1?

- Confirmación-0: se espera la respuesta del receptor. Una confirmación1 implica una correcta recepción de la trama y se pasa al estado Transmisión-0. Si no se recibe nada y expira el temporizador (*TimeOut*), se pasa al estado de Retransmisión-1.

- Retransmisión-1: se reenvía la trama d1, activando el temporizador, y pasando al estado de Confirmación-1?.

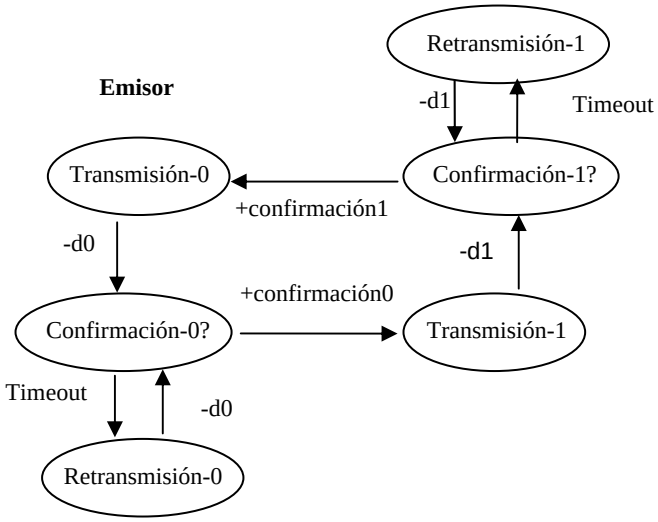


Figura 4.17. Autómata de control pasivo de errores en el transmisor.

Diseñar y explicar el autómata complementario para el nodo receptor

Solución

El autómata de la Figura 4.18 proporcional la lógica en el emisor.

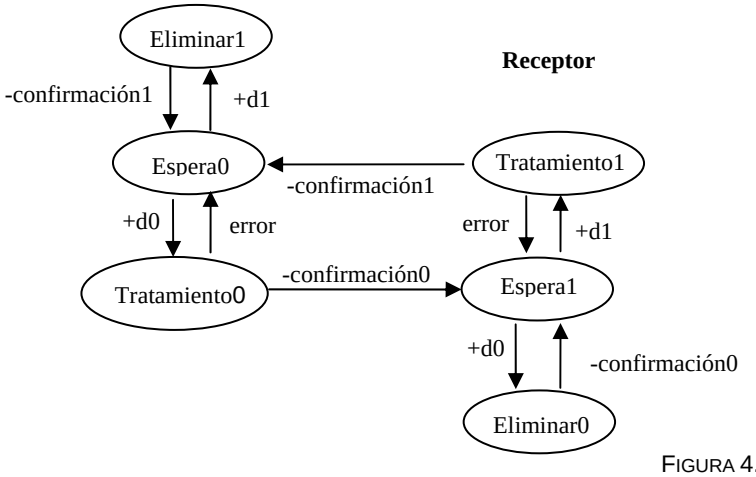


FIGURA 4.18.
Autómata de control pasivo de errores en el receptor.

Explicación de los estados y transiciones:

- Espera0: al recibir una trama d1 (la anteriormente recibida), se pasa al estado Eliminar1. Al recibir una trama d0 (la que se espera) se pasa al estado de Tratamiento0.

- Tratamiento0: se envía un confirmación0 y se pasa al estado Espera1

- Eliminar1: se envía una confirmación1.

- Espera1: en recibir una trama d0 (la anteriormente recibida), se pasa al estado Eliminar0. En recibir una trama d1 (la que se espera) se pasa al estado de Tratamiento1.

- Tratamiento1: se envía un confirmación1 y se pasa al estado Espera1

- Eliminar0: se envía una confirmación0.

Ejercicio 4.4. Tiempo de ciclo en un protocolo de acceso al medio aleatorio con resolución de colisiones

En una aplicación de control distribuido se utiliza un protocolo de acceso al medio aleatorio con resolución de colisiones. En cada ciclo de trabajo, cada uno de los 50 nodos tiene que enviar 1 mensaje de 150 bits.

a. ¿Si la velocidad de transmisión es de 250 kbps, cual será el tiempo mínimo de latencia entre dos mensajes consecutivos de un mismo nodo?

b. ¿Si la tasa de errores del canal es de 10^{-3} y los errores se tratan con un mecanismo de retransmisión inmediata por señalización, cual será el tiempo de latencia?

Solución

a. Teniendo en cuenta que no hay pérdidas de mensajes por colisiones, el tiempo de ciclo vendrá dado por el tiempo de transmisión de los 50 mensajes.

$$T_{latencia} \geq 50 \cdot (150 bits / 250 bps) = 30 ms$$

b. Si la tasa de errores es de 10^{-3}, de media, al enviar los 50 mensajes (50.150bits = 7500 bits) se producirán 7,5 (7500.10^{-3} = 7,5) errores. En el peor caso, estos se distribuyen en 8 mensajes. De los 8 nuevos mensajes a retrasmitir (8.150 bits = 1200 bits), se producirán 1.2 errores (1200.10^{-3} = 1,2), que en el peor de los casos se distribuyen en 2 mensajes. En consecuencia, el tiempo de latencia tiene que ser mayor o igual a la transmisión de los 50 mensajes originales, más los 8+2 de retransmisión. El tiempo de ciclo será (considerando despreciable la señalización del error):

$$T_{latencia} \geq (50 + 8 + 2) \cdot (150 bits / 250 bps) = 36 ms$$

Ejercicio 4.5. Rendimiento de un protocolo de acceso al medio por paso de testigo en anillo lógico.

Se ha diseñado una aplicación para 20 nodos sobre un protocolo de control de acceso al medio por paso de testigo en anillo lógico. Los parámetros del protocolo y aplicación son:

- Tiempo de posesión del testigo, T_t: 50 ms (se supone que un nodo en posesión del testigo siempre utiliza todo el T_t)
- El rendimiento de cada trama de datos es del 90%
- Tamaño del testigo: 50 bits
- La velocidad de transmisión es de 10 kbps
- El canal está libre de errores

¿Cual es el rendimiento (bits útiles/total de bits)?

Solución

Cada nodo transmitirá 500 bits (10 kbps.50ms), de los cuales el 90% son datos útiles, o sea, 450 bits. El total de datos útiles de los 20 nodos es 9000 bits (20x450 bits). El total de bits enviados es, por cada nodo, de 550 bits (los 500 bits que transmite más los 50 bits de transmisión del testigo). En consecuencia, el total de bits de los 20 nodos es de 11.000 bits (20x550 bits). El rendimiento es del 81% (9000/11.000 bits).

4.12.2. Evaluación objetiva

1. Cual de estos mecanismos no se especifica en la capa de enlace:

   A) Control de errores
   B) Control de flujo
   C) Control de acceso al medio
   D) Control de enrutamiento

2. ¿En que posición se encuentra el campo de detección de errores dentro de una trama?

   A) Al inicio
   B) En medio
   C) Al final
   D) En cualquier lugar

3. ¿En un esquema de control pasivo de errores, qué nodo activa el temporizador?

   A) El emisor
   B) El receptor
   C) Ninguno de los dos
   D) Los dos

4. Los protocolos con métodos de control de acceso al medio aleatorios con colisiones:

   A) Nunca se utilizan en al automatización
   B) Se utilizan en todos los niveles de la automatización
   C) Se utilizan en los niveles altos de la automatización
   D) Se utilizan en los niveles inferiores de la automatización

5. El protocolo maestro/esclavo es un método de acceso al medio:

   A) Determinista con control centralizado
   B) Determinista con control descentralizado
   C) Aleatorio con colisiones
   D) Aleatorio sin colisiones

6. El tiempo de transmisión de un mensaje depende:

   A) De la longitud de la red
   B) Del tamaño en bits a enviar
   C) Del número de nodos
   D) Del protocolo de acceso al medio

7. ¿Cuál de las siguientes afirmaciones no ayuda a aumentar el determinismo de Ethernet?

   A) Implementar un maestro/esclavo por encima de Ethernet
   B) Utilizar más nodos en la red
   C) Aumentar la velocidad de transmisión y controlar la carga
   D) Usar la tecnología de conmutadores

8. Si la codificación de la trama en un protocolo industrial tiene un rendimiento del 80%,

   A) De cada 100 bits que se envían, 80 son de datos útiles
   B) De cada 100 bits que se envían, 20 tienen que ser retransmitidos
   C) De cada 100 bits que se envían, 20 son de datos útiles
   D) No se puede garantizar comunicación de tiempo real

9. El método de *Stop-and-Wait*:

   A) Se basa en dejar que los nodos trasmitan siempre que el bus esté libre

   B) No es apropiado para el control de flujo

   C) Permite que un nodo transmisor no sobrecargue a un nodo receptor

   D) Ayuda a reducir las alteraciones del estado lógico de los bits

10. Que método de acceso al medio no es aconsejable para hacer el control delsistema antibloqueo del freno de un automóvil

   A) Maestro/esclavo

   B) CSMA/CD

   C) TDMA

   D) TTP

TEMA 5
REDES DE ÁREA LOCAL. COMPONENTES

5.1. Introducción y orientaciones al estudio
5.2. Objetivos
5.3. Características generales de las LAN
5.4. Topologías y medios de transmisión
5.5. Modelo IEEE 802, IEEE 802.3 y ETHERNET
5.6. Dispositivos de interconexión de redes
5.7. Redes de área local de alta velocidad
5.8. Redes inalámbricas
5.9. Aplicaciones
5.10. Conclusiones
5.11. Bibliografía
5.12. Evaluación

5.1. INTRODUCCIÓN Y ORIENTACIONES AL ESTUDIO

En este capítulo se van a tratar los diferentes aspectos a tener en cuanta a la hora de diseñar y ampliar una red de área local o LAN. También se describirán las redes LAN más habituales y sus aplicaciones.

Aunque estos elementos se describan de forma individual, no son independientes unos de otros por lo que se debe ver como una parte de un todo que esta en continua interacción.

Al principio del capitulo se describirán las características de una red de de área local como: topologías y medios de transmisión. En muchos casos los medios de transmisión y la topología están fuertemente relacionados. Así una topología puede determinar que medio de transmisión se debe utilizar y viceversa.

Posteriormente se definirá uno de los estándares más utilizado, IEEE 802.3 salvo por dos pequeñas diferencias la norma IEEE 802.3 es muy similar a las redes Ethernet.

Las mejoras tecnológicas han traído nuevos esquemas de diseño. Estos esquemas han dado lugar a redes de alta velocidad y redes inalámbricas. En este capítulo veremos las características de cada una de ellas.

5.2. OBJETIVOS

Podemos dividir este capítulo en tres partes. La primera parte pretende dar a conocer al lector las características de una red, algunos estándares de red como el modelo IEEE 802.3 y Ethernet y por último conocer los diferentes dispositivos que permiten la expansión e interconexión entre redes. La segunda parte describirá algunas de las redes de área local más actuales como son: redes de alta velocidad y redes inalámbricas. Por último la tercera parte pretende que el lector vea algunas de las aplicaciones de una red de área local en la industria.

5.3. CARACTERÍSTICAS GENERALES DE LAS LAN

Las Redes de Área Local (RAL) o LAN (*Local Area Network*) generalmente son redes privadas que se encuentran en un edificio o en un campus de pocos kilómetros de longitud.

Una de sus principales aplicaciones es la de conectar estaciones de trabajo, PCs, impresoras, etc. de una oficina o fabrica para compartir recursos e intercambiar información.

Los principales elementos de una red RAL son los siguientes:

- Topología.
- Medio de transmisión.
- Disposición.
- Técnicas de acceso al medio.

En este tema se irá describiendo cada uno de ellos de manera detallada.

5.4. TOPOLOGÍAS Y MEDIOS DE TRANSMISIÓN

Como ya se ha comentado en apartados anteriores dos aspectos a tener en cuenta para el diseño de una red de área local son: la topología de la red y el medio de transmisión. En este apartado se va a tratar cada uno de ellos detalladamente.

5.4.1. Topologías

En el mundo de las redes de comunicación el término topología hace referencia a la disposición física en la que se conectan los diferentes elementos de una red o nodos.

Las topologías más comunes para una red de área local son la topología en bus, en árbol, en anillo y en estrella. A continuación se describirá cada una de ellas. Conviene recordar que estas topologías pueden mezclarse dado lugar a topologías como anillo-estrella.

5.4.1.1. Topología en bus

Todos los elementos de la red o nodos se conectan de forma directa, a través de una interfaz física o toma de conexión, a un medio de transmisión lineal o bus. El bus está delimitado en sus extremos por unas resistencias o terminadores que tienen como misión absorber las señales que viajan por el bus (Fig. 5.1).

La comunicación entre el nodo y la toma de conexión es full-duplex, por lo que es posible la transmisión y recepción de datos a través del bus. Cuando un nodo transmite información, esta se propaga a través del bus en ambos sentidos y es recibida por el resto de nodos.

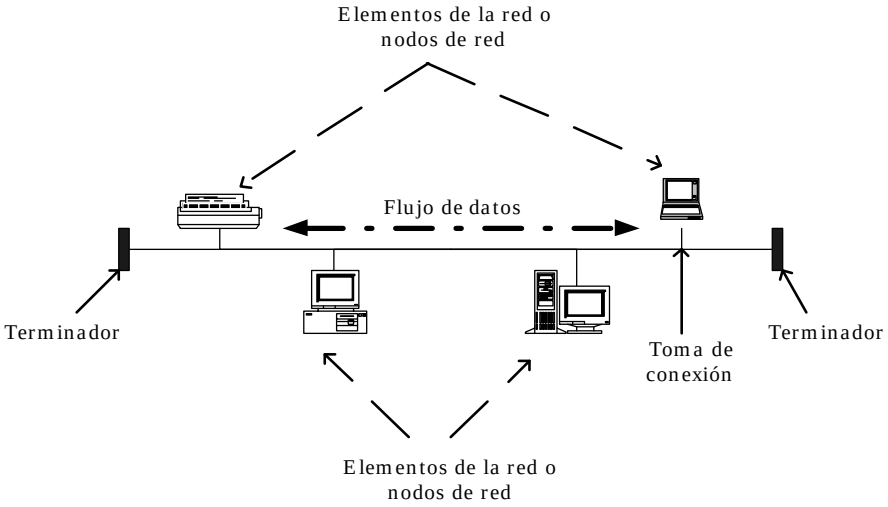


FIGURA 5.1. Topología en bus.

Si esto es así, ¿Cómo se indica a quien va dirigida la información que se propaga por el bus?, para solucionar este problema, a cada nodo de la red se le asigna un identificador o dirección única. Así cuando un nodo desea enviar información a otro nodo lo primero que hace es dividir dicha información en pequeños bloques llamados tramas. Cada trama consta de una porción de la información a enviar y una cabecera de control que contiene la dirección del nodo destino. A continuación se verá un pequeño ejemplo:

1. Imagine un bus con 4 nodos. Cada uno de estos nodos tiene un identificador único: A, B, C y D (Fig. 5.2).

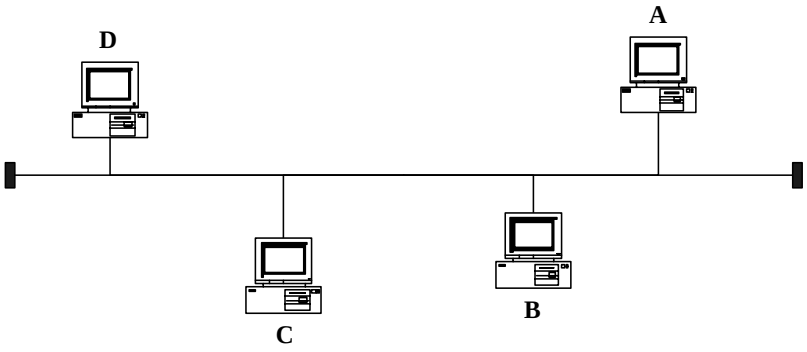


FIGURA 5.2. Bus con 4 nodos.

2. El nodo B desea enviar información al nodo D. Para ello divide dicha información en tramas, colocando en cada cabecera de la trama el identificador o dirección del nodo destino y las envía a través del bus (Fig. 5.3).

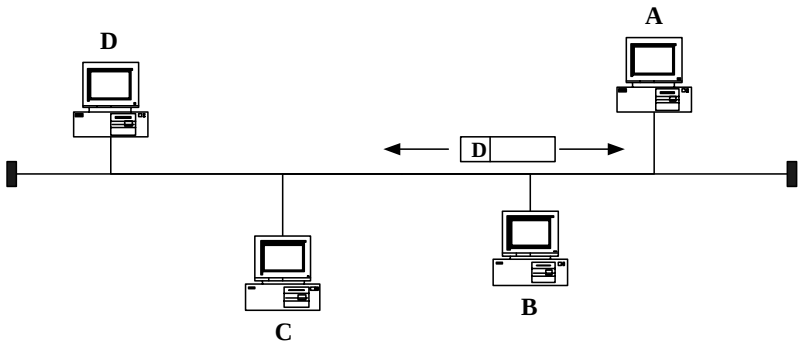


FIGURA 5.3. El nodo B transmite una trama para el nodo D.

3. La trama se propaga a través del bus en ambos sentidos y llega al nodo A y C. Estos nodos comprueban la dirección destino de la trama y la ignoran (Fig. 5.4).

4. El nodo D observa que la trama va dirigida a el y copia los datos de esta mientras que pasa. Mientras tanto la otra copia de la trama llega al terminador y es eliminada del bus (Fig. 5.5).

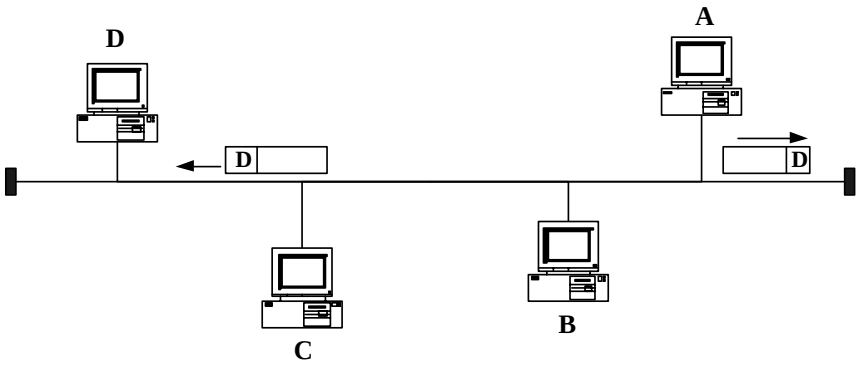


FIGURA 5.4. El nodo C y A ignoran la trama que va dirigida al nodo D.

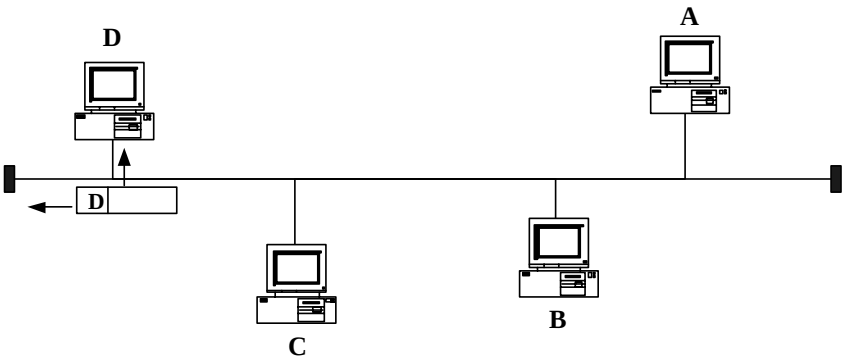


FIGURA 5.5. El nodo C y copia la trama conforme pasa.

Otro aspecto a tener en cuenta es establecer un mecanismo de control de acceso que permita regular la transmisión de datos. Este mecanismo deberá evitar que dos nodos intenten transmitir información simultáneamente o que un nodo decida transmitir información durante un largo periodo de tiempo. De estos mecanismos se hablará más adelante.

5.4.1.2. Topología en árbol

Es una generalización de la topología en bus en la que el cable se desdobla en varios ramales sin bucles cerrados (Fig. 5.6). Al igual que la topología en bus, las transmisiones se propagan por cada ramal de la red y llegan a todos los nodos o estaciones de la red.

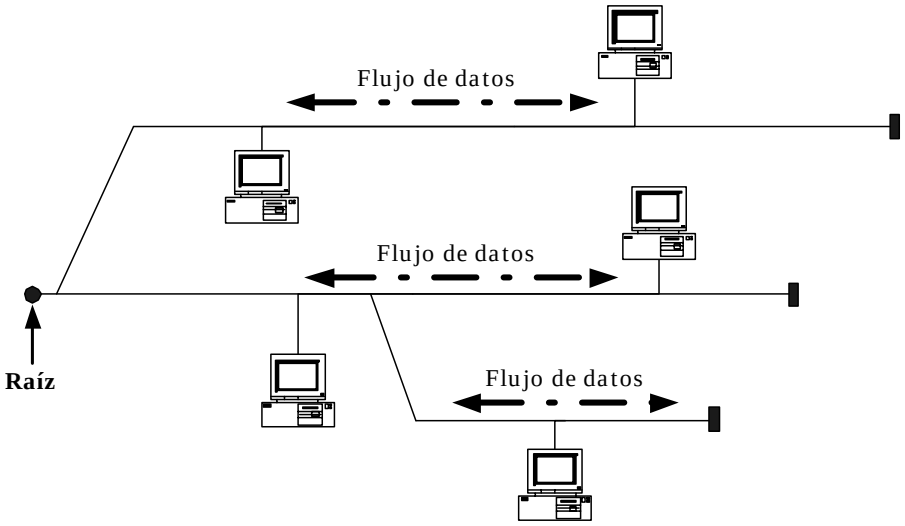


FIGURA 5.6. Topología en árbol.

5.4.1.3. Topología en anillo

La red la componen un conjunto de repetidores unidos por enlaces punto a punto unidireccionales, formando un bucle cerrado o anillo (Fig. 5.7). Un repetidor es un dispositivo que tiene como misión recibir datos del nodo o estación de trabajo y transmitirlos secuencialmente bit a bit. Al igual que en las topologías de bus y anillo cuando una estación de trabajo transmite información a través del repetidor, esta se divide en tramas. Esta tramas viajan por el anillo en un solo sentido, en el sentido de las agujas del reloj o en sentido contrario.

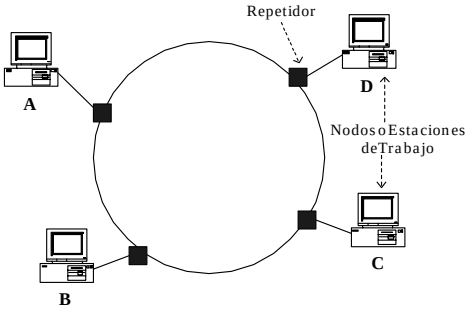


FIGURA 5.7. Topología en anillo.

A continuación se va a analizar un ejemplo de cómo transmitir la información a través de una red en anillo. Para ello se utilizará la red en anillo que se muestra en la figura 5.7. Dicha red está formada por 4 nodos, cada uno de estos nodos tiene un identificador único: A, B, C y D.

1. El nodo A desea enviar información al nodo C. Para ello divide dicha información en tramas, colocando en cada cabecera el identificador o dirección del nodo destino y las envía a través del repetidor (Fig. 5.8).

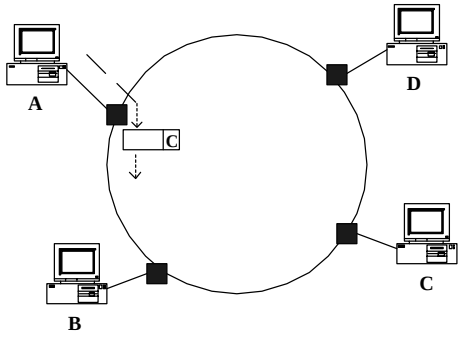


FIGURA 5.8. El nodo A envía datos para el nodo C a través del anillo.

2. La trama se propaga a través del anillo en un único sentido y llega al nodo B. Estos nodos comprueban la dirección destino de la trama y la ignoran (Fig. 5.9).

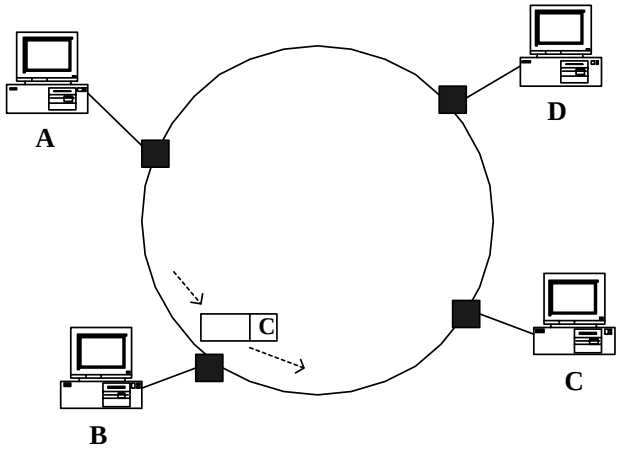


FIGURA 5.9. El nodo B ignora la trama ya que no va dirigida a él.

3. El nodo C observa que la trama va dirigida a el y copia los datos de esta mientras que pasa (Fig. 5.10).

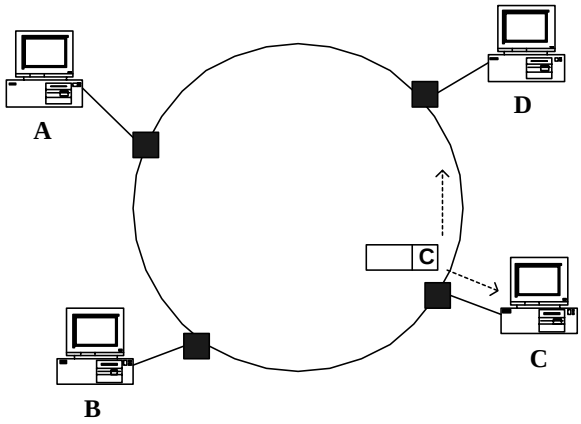


FIGURA 5.10. El nodo C copia la trama mientras pasa.

4. El nodo D ignora la trama y esta vuelve a pasar por el nodo A el cual recoge la trama y la saca del anillo (Fig. 5.11).

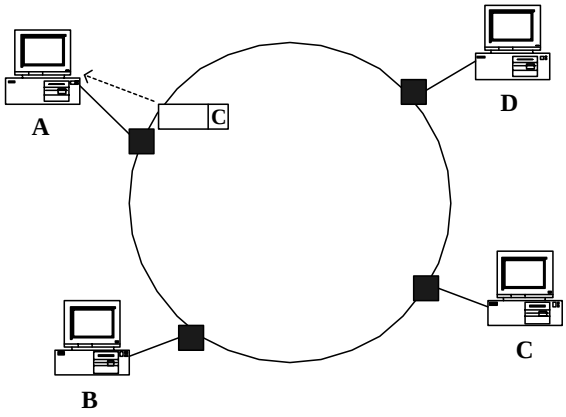


FIGURA 5.11. El nodo A saca del anillo la trama que envió.

La topología en anillo puede ser utilizada para proporcionar enlaces de muy alta velocidad sobre distancias largas. Un anillo puede proporcionar, potencialmente, mejor rendimiento que cualquier otra topología. Una desventaja, es que un fallo de un solo enlace o de un repetidor puede inutilizar la red entera.

5.4.1.4. Topología en estrella

Cada nodo de la red está conectado a un nodo central común, normalmente este nodo común suele ser un concentrador, conmutador o repetidor (Fig. 5.12).

Generalmente la conexión entre un nodo y el nodo central común se hace a través de un enlace bidireccional. Aunque también se pueden utilizar dos enlaces punto a punto, uno para la transmisión de información y otro para la recepción de información.

Existen dos alternativas para el funcionamiento del nodo central:

1. El nodo central trabaja en modo difusión. Así, la trama enviada por un nodo es reenviada por el nodo central a todas las estaciones que tiene conectadas. En este caso al nodo central se le conoce como concentrador.

2. El nodo central trabaja como conmutador de tramas. Así, la trama enviada por un nodo es almacenada por el nodo central y transmitida al enlace del nodo destino.

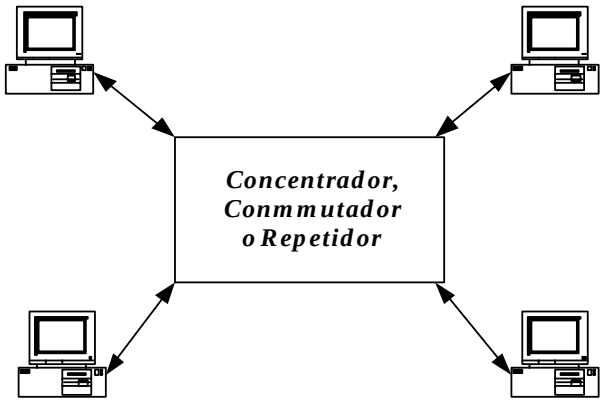


FIGURA 5.12. Topología en estrella.

La topología en estrella se aprovecha de la disposición natural del cableado de los edificios. Generalmente es mejor para distancias cortas y puede ofrecer velocidades elevadas a un número pequeño de dispositivos.

5.4.2. Medios de transmisión

Otro de los aspectos clave a la hora de diseñar una red de área local o LAN es el medio físico que transporta la información que viaja por la red. La elección del medio físico está muy ligada a la topología que se elija y viceversa. Otros factores que desempeñan un papel importante en dicha elección son:

- Ancho de banda. Esta definido por el espectro de frecuencias que el medio puede transferir. Cuanto mayor sea el ancho de bandas, las velocidades de transmisión serán mayores.
- Alcance del entorno. Debe proporcionar servicio a la gama de entornos requeridos.
- Fiabilidad: debe satisfacer los requisitos de disponibilidad.
- Seguridad. Grado de dificultad con que las señales transportadas pueden ser interceptadas.
- Coste.

Los medios de transmisión se pueden clasificar en medios guiados como: cables de cobre y fibra óptica y en medios no guiados, transmisión inalámbrica. A continuación se van a describir algunos de los medios más significativos.

5.4.2.1. Cable de par trenzado

Es uno de los medios de transmisión más viejo. Consiste en dos alambres de cobre aislados que se trenzan en forma helicoidal, igual que una molécula de ADN (Fig. 5.13). Se pueden utilizar tanto para comunicaciones analógicas, sistema telefónico analógico, como para comunicaciones digitales.

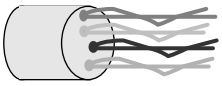


FIGURA 5.13. Par trenzado sin apantallar.

El cable de par trenzado puede ser:

- Apantallado o STP (*Shielded Twister Pair*): están cubiertos por un blindaje que minimiza las interferencias electromagnéticas y diafonía. Fueron introducidos por IBM a principios de la década de los 80, pero

debido a su coste y volumen no ganaron popularidad fuera de las instalaciones de IBM.

- Sin apantallar o UTP (*Unshielded Twister Pair*): debido a su bajo coste, sencillez de instalación y utilización en el tendido telefónico los cables UTP se utilizan en un gran número de redes LAN. Los UTP se clasifican por categorías, las más utilizadas son la categoría 3 y 5. A continuación se describirá cada una de las categorías existentes:
  - Categoría 1: solamente voz (cable telefónico).
  - Categoría 2: se utilizan para transmisión de datos a bajas velocidades, inferiores a 4 Mbps.
  - Categoría 3: Comenzaron a utilizarse en redes Ethernet a 10 Mbps, con longitudes de segmento inferiores a 100 m y máxima longitud de red de 500 m. Posteriormente, se extendió su uso para redes con paso de testigo a 4 Mbps y 16 Mbps y redes de alta velocidad a 100 Mbps; a esta velocidad se necesitan varios pares, normalmente 4, para alcanzar dicha velocidad de transmisión.
  - Categoría 4: soporta velocidades de 20 Mbps. En redes Token Ring hasta 16 Mbps.
  - Categoría 5: son muy utilizados debido a que soporta velocidades de hasta 100 y 150 Mbps. Actualmente puede llegar a velocidades de 1 Gbps.
  - Categoría 6 y 7: son capaces de manejar señales con anchos de banda de 250 y 600 MHz, respectivamente.

5.4.2.2. Cable coaxial

Un cable coaxial consiste en un alambre de cobre rígido como núcleo, rodeado de un material aislante.

El aislante está forrado de un conductor cilíndrico que con frecuencia es una malla de tejido fuertemente trenzado. El conductor externo se cubre con una envoltura de plástico (Fig. 5.14).

En redes de área local el cable coaxial se emplea tanto en transmisiones en banda base o en banda ancha.

- Cable coaxial en banda base: hace uso de la señalización digital. Es el más utilizado en redes LAN con topología en bus, principalmente en el caso de los sistemas Ethernet.
- Cable coaxial en banda ancha: hace uso de la señalización analógica. Las redes basadas en el ya no se construyen.

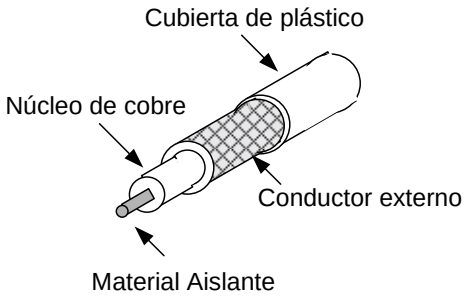


FIGURA 5.14. Cable coaxial.

Los cables coaxiales solían ser ampliamente utilizados en el sistema telefónico para las líneas de larga distancia, pero en la actualidad han sido reemplazados por la fibra óptica. Aún se utilizan en la televisión por cable y las redes de área metropolitana o WAN.

5.4.2.3. Fibra óptica

Un sistema de transmisión óptico está formado por tres componentes:

1. La fuente de luz. Convencionalmente un pulso de luz un bit a 1 y la ausencia de luz indica un bit a 0.

2. El medio de transmisión. Es una fibra de vidrio muy delgada.

3. El detector. Genera un pulso eléctrico cuando la luz incide en el.

La fibra óptica esta formada por un núcleo de vidrio, a través del cual se propaga la luz. Dicho núcleo está rodeado por un revestimiento de vidrio cuyo índice de refracción es menor que el del núcleo con el fin de mantener toda la luz en este último. Para proteger el revestimiento se utiliza una cubierta de plástico (Fig. 5.15).

La fibra óptica permite un gran ancho de banda y por tanto velocidades de transmisión mas elevadas que los cables vistos anteriormente. Su principal utilización es en redes LAN con topologías en anillo.

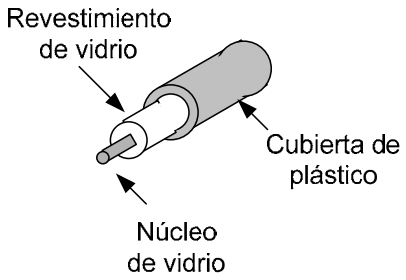


FIGURA 5.15. Fibra óptica.

Algunas de las principales desventajas es el coste de las tomas de conexión y que dichas conexiones son mas complicadas que una conexión Ethernet.

5.4.2.4. Transmisión inalámbrica

Actualmente vivimos en la era de la información, en muchas ocasiones los trabajadores deben disponer de una conexión a red aunque no existan conexiones físicas. Las comunicaciones inalámbricas son una buena solución para este tipo de problemas.

Al conectarse una antena del tamaño apropiado a un circuito eléctrico, las ondas electromagnéticas pueden ser difundidas de manera eficiente y ser captadas por un receptor a cierta distancia. Toda comunicación inalámbrica se basa en el este principio. Las redes LAN generalmente se clasifican según la técnica de transmisión utilizada. Por tanto podemos definir las siguientes tipos de redes LAN inalámbricas:

- LAN de infrarrojos.
- LAN de espectro expandido.
- LAN de banda estrecha. Estás últimas operan en el rango de las microondas, pero no hacen uso de espectro expandido.

5.5. MODELO IEEE 802, IEEE 802.3 Y ETHERNET

El Instituto Americano de Ingenieros Eléctricos y Electrónicos o IEEE (*The Institute of Electrical and Electronics Engineers*) ha estandarizado diferentes redes de área local o LAN y redes de área metropolitana o MAN (*Metropolitan Area Network*) bajo el nombre de IEEE 802. Algunas de las más significativas son:

- IEEE 802.3, Ethernet.
- IEEE 802.11, LAN inalámbrica.
- IEEE 802.15, Bluetooth.
- IEEE 802.16, MAN inalámbrica.

5.5.1. Modelo de referencia IEEE 802

El modelo de referencia IEEE 802 se centra en las 2 primeros capas del modelo OSI (*Open System Interconnection*) (Fig. 5.16).

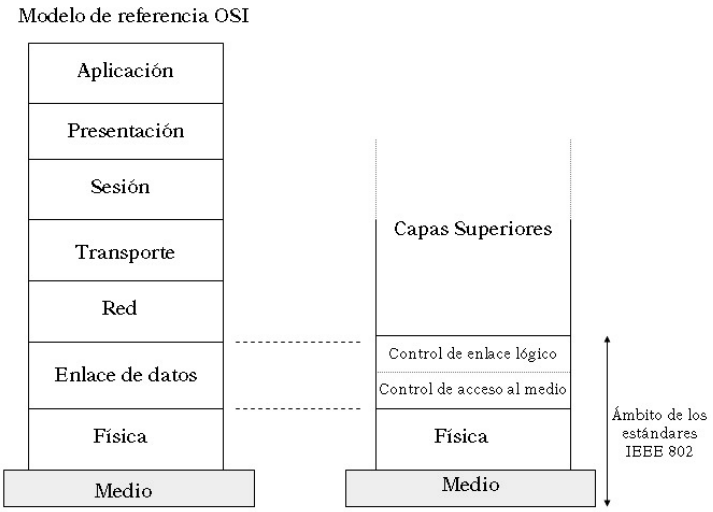


FIGURA 5.16. Ámbito de los estándares IEEE 802 con respecto al modelo OSI.

Como se puede ver en la Figura 5.16, la capa inferior del modelo de referencia IEEE 802 es la capa física del modelo OSI, donde se incluyen funciones como:

- Codificación, decodificación de señales.
- Generación o eliminación de preámbulo (sincronización).
- Transmisión y recepción de bits.

La capa física del modelo 802 también incluye una especificación del medio de transmisión y de la topología. Por encima de la capa física se encuentra la capa de enlace de datos del modelo OSI, El modelo 802 divide dicha capa en:

1. Control de Acceso al medio o MAC (*Media Access Control*), donde se encuentran las siguientes funciones:
    - En transmisión, ensamblado de datos en tramas con campos de dirección y detección de errores.
    - En recepción, desensamblado de tramas, reconocimiento de dirección y detección de errores.
    - Control de acceso al medio de transmisión LAN.
2. Control de enlace lógico o LLC (*Logical Link Control*), su función es hacer de interfaz con las capas superiores y control de errores y de flujo.

5.5.1.1. Control de enlace lógico

LLC especifica los mecanismos para direccionar estaciones a través del medio y para controlar el intercambio de datos entre dos usuarios sin la utilización de nodos intermedios.

Existen tres posibles servicios para dispositivos conectados que usan LLC.

1. Servicio no orientado a conexión sin confirmación: este servicio es de tipo datagrama. No incluye mecanismos de control de flujo ni de errores, por lo que no esta garantizada la recepción de los datos. En la mayoría de los casos existe alguna capa superior de software encargada de gestionar las cuestiones de fiabilidad.

2. Servicio en modo orientado a conexión: se establece una conexión lógica entre dos usuarios que intercambian datos. Existe control de flujo y control de errores.

3. Servicio no orientado a conexión con confirmación: Los datagramas son confirmados, pero no se establece una conexión lógica previa.

5.5.1.1.1. Protocolo de control de enlace lógico

El protocolo LLC sigue el modelo definido por HDLC (High-Level Data-Link Control). Las principales diferencias entre ambos son:

- LLC utiliza el modo de operación balanceado asíncrono de HDLC para dar soporte al servicio LLC orientado a conexión.
- LLC utiliza una unidad de datos de protocolo o PDU no numerada para dar soporte al servicio no orientado a conexión sin confirmación.
- LLC ofrece un servicio no orientado a conexión confirmado para ello utiliza dos PDU no numeradas nuevas.
- LLC permite multiplexación mediante el empleo de los puntos de acceso al servicio del subnivel LLC o LSAP.

A continuación se va a describir brevemente la estructura que tiene una PDU en el subnivel LLC (Fig. 5.17).

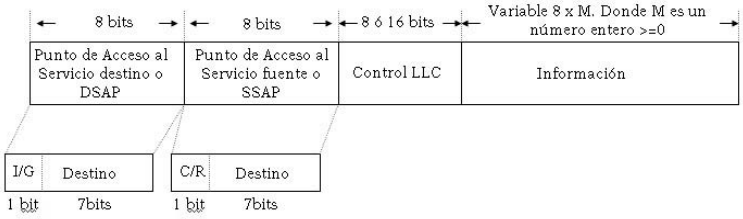


FIGURA 5.17. PDU del subnivel LLC.

- El DSAP indica el punto de acceso al servicio al que va dirigido el campo de información.
  - Bit I/G: Indica si la dirección destino es una dirección individual (I/G=0) o si es un grupo (I/G=1).

  - 7 bits para la dirección destino.
- El SSAP indica el punto de acceso al servicio en el que el campo de información fue generado.
  - Bit C/R: Indica si es una orden (C/R=0) o si es una respuesta (C/R=1).
  - 7 bits para la dirección origen.
- Control LLC. Esta formado por:
  - 16 bits para formato de información y supervisión.
  - 8 bits para formato no numerado.
- Información. Debe tener un número entero de octetos.

En el estándar se definen 3 tipos de protocolos LLC, denominados tipos de operación, uno por cada uno de los servicios definidos.

- Operación de tipo 1. Soporta el servicio no orientado a conexión sin confirmación.
- Operación de tipo 2. Soporta el servicio orientado a conexión.
- Operación de tipo 3. soporta el servicio no orientado a conexión con confirmación.

5.5.1.2. Control de acceso al medio

Las estaciones o elementos de red comparten un mismo medio a través del cual viaja la información que transmiten. Es por tanto necesario, algún método de control de acceso al medio con objeto de hacer un uso eficiente de este.

Los parámetros a tener en cuenta en cualquier técnica de control de acceso al medio son:

- Si el control se realiza de forma centralizada o descentralizada.
  - Centralizado: se diseña un controlador con autoridad para conceder el acceso a la red. De esta forma cuando una estación desee transmitir información a través de la red debe esperar a tener permiso del controlador.

  - Descentralizado: Las estaciones realizan conjuntamente la función de control de acceso al medio para determinar dinámicamente el orden en que transmitirla.

- Si la asignación del canal es estática o dinámica.

5.5.1.2.1. Algoritmos de asignación de canal de acceso múltiple

Existen un gran número de algoritmos para asignar una canal de acceso múltiple. A continuación vamos a mencionar algunos de los más interesantes:

- ALOHA: la aparición en la década de los 70 Norman Abramson y sus colegas de la Universidad de Hawai desarrollaron un método novedoso y elegante para resolver el problema del acceso al medio. Este método ha sido ampliado por muchos investigadores. Se van a ver dos versiones:
  - ALOHA puro: se permite a los usuarios transmitir cuando tengan datos que enviar. Por supuesto va a ver colisiones y algunas de las tramas en colisión se dañarán. Sin embargo debido a la propiedad de retroalimentación de la difusión, un emisor siempre podrá saber si la trama fue destruida o no escuchando el canal. Si por alguna razón no es posible escuchar mientras se transmite, se necesita confirmación de recepción. Si la trama fue destruida se espera un tiempo aleatorio y la envía de nuevo. Estos sistemas que en lo que varios usuarios comparten un canal común de modo tal que puede dar pie a conflicto se conoce como sistemas de contención.
  - ALOHA ranurado. En contraste con el ALOHA puro se divide el tiempo en intervalos discretos, cada uno de los cuales correspondería a una trama. Por tanto la probabilidad de que no haya mas tráfico durante la misma ranura es mayor.
- En el algoritmo ALOHA las estaciones transmiten sin importarle lo que hacen el resto de estaciones o elementos de red.. En las redes de área local es posible que las estaciones detecten lo que están haciendo el resto de estaciones y adapten su comportamiento en base a ello, esto recibe el nombre de protocolos de detección de portadora. Estas estaciones pueden lograr una utilización del canal mucho mejor.
  - Acceso Múltiple con Detección de Portadora Persistente-1. También conocido como CSMA persistente. Cuando una estación tiene información que transmitir, lo primero que hace es escuchar el canal para saber si otra estación está transmitiendo. Si el canal está

ocupado, la estación espera hasta que se desocupa. Cuando la estación detecta un canal inactivo, transmite la trama. Si ocurre una colisión, la estación espera una cantidad aleatoria de tiempo y comienza de nuevo. Se denomina persistente-1 porque la estación transmite con una probabilidad de 1 cuando encuentra que el canal está inactivo.

- Acceso Múltiple con Detección de Portadora no Persistente. También conocido como CSMA no persistente. Antes de enviar información la estación escucha el canal. Si nadie más está transmitiendo, la estación comienza a hacerlo. Sin embargo si el canal está en uso, la estación no se queda continuamente a la escucha hasta que el canal quede libre sino que espera un periodo aleatorio de tiempo y repite el algoritmo.
- CSMA con detección de colisiones o CSMA/CD. Los algoritmos CSMA vistos hasta ahora también presentan un ineficiencia manifiesta. Cuando dos estaciones detectan el canal inactivo y transmiten datos se produce la colisión de sus tramas y por tanto el medio permanece inutilizable mientras dure la transmisión de dichas tramas dañadas. Este desaprovechamiento de la capacidad puede reducirse si una estación continua escuchando el medio mientras dure la transmisión. La inclusión de característica conduce a las siguientes reglas para CSMA/CD:

  1. Si el canal esta libre, transmite. Sino se va al paso 2.

  2. Si el canal está ocupado, continua escuchando hasta que el canal quede libre, en cuyo caso transmite inmediatamente.

  3. Si se detecta colisión durante la transmisión, se transmite una pequeña señal de interferencia para asegurarse de que todas las estaciones constaten la colisión. A continuación se deja de transmitir.

  4. Tras la emisión de la señal de interferencia, la estación espera una cantidad aleatoria de tiempo conocida como espera (*backoff*), intentando transmitir de nuevo a continuación (volviendo al paso 1).

5.5.1.2.2. La trama MAC

La capa MAC recibe un bloque de datos de la capa LLC y debe realizar funciones relacionadas con el acceso al medio y la transmisión de datos: síncrona o asíncrona. Para implementar dichas funciones se utiliza una unidad de datos de protocolo (PDU) a la que se denomina Trama MAC (Fig. 5.18).

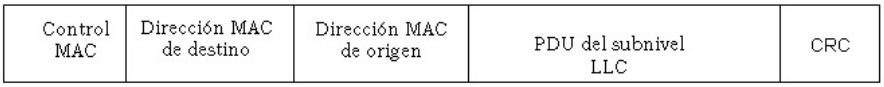


FIGURA 5.18. Trama MAC.

A continuación se describirá brevemente cada uno de los campos que forma una trama MAC:

- Control MAC: contiene información de control de protocolo MAC.
- Dirección MAC de destino: es la dirección física MAC en la LAN a la que va dirigida la trama.
- Dirección MAC de destino: es la dirección física MAC en la LAN que originó la trama.
- PDU LLC: en la trama MAC van encapsulados la unidad de datos del protocolo LLC.
- CRC: campo de comprobación de redundancia cíclica, también conocido como campo de secuencia de comprobación de trama o FCS (*Frame Check Sequence*). Este campo permite la detección de errores.

5.5.2. IEEE 802.3 y ETHERNET

EL estándar IEEE 802.3 y Ethernet son idénticos excepto por pequeñas diferencias. De tal forma que son capaces de coexistir en el mismo medio. En las redes Ethernet se utilizan comúnmente cuatro tipos de cableado (Tabla 5.1).

Cada una de estas redes recibe un nombre que utiliza la siguiente notación:

<Velocidad (Mbps)><Tipo de transmisión><Distancia (centenares de metros) (opcional)>

TABLA 5.1. Cableado Ethernet

Nombre	Cable	Segmento máximo	Nodos/ segmento	Ventajas
10Base5	Coaxial grueso	500 m	100	Cable original. Obsoleto
10Base2	Coaxial delgado	185 m	30	No se necesita concentrador
10Base-T	Par trenzado	100 m	1024	Sistema más económico
10Base-F	Fibra óptica	2000 m	1024	Mejor entre edificios

Así la red con nombre 10Base5 implica una velocidad máxima de 10 Mbps, transmisión en banda base y una longitud máxima de segmento de 500 metros. Por el contrario la red 10Base-T indica una velocidad máxima de 10 Mbps, transmisión en banda base y que el tipo de cable es trenzado. Como se verá en apartado posterior han aparecido redes con una velocidad de transmisión mucho mayor que reciben el nombre de redes de alta velocidad.

El protocolo de control de acceso al medio es el CSMA/CD. La trama MAC definida por el IEEE 802.3 esta compuesta por una serie de campos que se van a definir a continuación (Fig. 5.19)

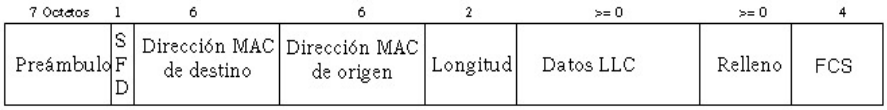


FIGURA 5.19. Trama MAC, IEEE 802.3

- Preámbulo: establece la sincronización entre el emisor y el receptor.

- Delimitador de comienzo de trama o SFD (*Stara Frame Delimiter*): consiste en la secuencia de bits 10101011, que indica el comienzo real de la trama y permite que el receptor pueda localizar el primer bit del resto de la trama.

- Dirección MAC de destino: indica la estación o estaciones a las que va dirigida la trama.

- Dirección de MAC de origen: indica la estación que envió la trama.

- Longitud/Tipo: contiene la longitud del campo de datos LLC expresado en octetos, o el campo tipo Ethernet, dependiendo de que la trama siga la norma IEEE 802.3 o la especificación primitiva de Ethernet.

- Datos LLC: unidad de datos proporcionada por el LLC.

- Relleno: octetos añadidos para asegurar que la trama sea lo suficientemente larga como para que la técnica de detección de colisiones funcione correctamente.

- Secuencia de comprobación de trama o FCS (*Frame Check Sequence*): comprobación de redundancia cíclica de 32 bits.

En el estándar IEEE 802 también se encuentran definidas otras características de la red como conectores que se utilizan para conectar los diferentes elementos de red al canal, la estructura de la trama MAC, etc.

5.5.3. Redes de área local virtuales

En muchas organizaciones o empresas los cambios organizacionales ocurren frecuentemente. Por lo que los administradores de sistemas pierden mucho tiempo quitando y poniendo conectores de un lado a otro. En algunos casos no se pueden realizar dichas modificaciones por problemas de cableado, como por ejemplo que la estación del usuario este demasiado lejos del concentrador correcto.

Para dar mayor flexibilidad a estos cambios, los fabricantes de redes empezaron a trabajar en una forma de volver a cablear edificios completos mediante software. Esto dio lugar al concepto de redes de área local virtuales o VLAN (*Virtual Local Area Network*) que fue estandarizado por el IEEE 802 y que utilizan conmutadores diseñados para proporcionar esta flexibilidad.

Así, para configurar una VLAN el administrador de la red decide cuantas redes virtuales habrá, que nombre tendrán y que equipos de la red pertenecerán a una u otra red virtual.

Algunos de los ejemplos más comunes de VLAN con Internet son:

- Posibilidad de conectar dos o más edificios de una empresa utilizando como vínculo Internet.

- Permitir, a algunas personas de la organización, la conexión a el centro de cálculo desde su casa. O que un usuario pueda acceder a su equipo de casa desde un sitio remoto, como por ejemplo un hotel.

Internet es un medio inseguro, muchos de los protocolos utilizados hoy en día para transferir datos de una máquina a otra a través de la red carecen de algún tipo de encriptación o medio de seguridad, por lo que cualquier persona con un sniffer sería capaz de recibir y analizar los paquetes enviados. El tunnelling consiste en abrir conexiones entre dos máquinas por medio de un protocolo seguro, como puede ser SSH (*Secure Shell*), a través de las cuales realizaremos las transferencias inseguras, que pasarán de este modo a ser seguras.

5.6. DISPOSITIVOS DE INTERCONEXIÓN DE REDES

Debido al crecimiento que pueden experimentar las redes, es fundamental disponer de diversos dispositivos que permitan la expansión de la red o la conexión con otras redes. En este apartado se describirán los dispositivos mas utilizados como: Repetidores, puentes, encaminadores y pasarelas.

5.6.1. Repetidores

Son dispositivos que se encargan de regenerar las señales en los dos segmentos de red que interconectan. Trabajan a nivel físico (Fig. 5.19). Por tanto son capaces de conectar diferentes medios físicos de transmisión, ya que esto solo implica modificar el formato físico de la señal y no la decodificación o estructura de las tramas de datos. Transmite a nivel de bits.

Generalmente los repetidores son dispositivos de uso limitado y se suelen emplear para interconectar redes LAN homogéneas.

5.6.2. Puentes

Mientras que los repetidores actúan sobre los bits transferidos. Los puentes operan sobre las tramas que se transfieren en los niveles de enlaces de datos, particularmente sobre el nivel de control de acceso al medio o MAC (Fig. 5.20).

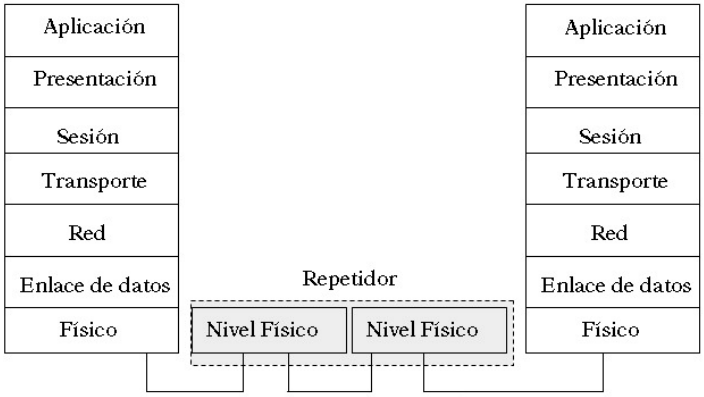


FIGURA 5.19. Esquema OSI de un repetidor.

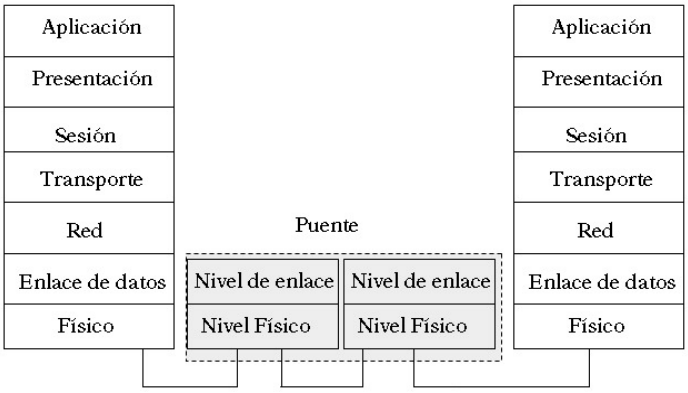


FIGURA 5.20. Esquema OSI de un puente.

Las funciones básicas de un puente son: filtrado, el puente es capaz de examinar los campos de dirección de la trama y determinar si enviar o no dicha trama. Para poder examinar estas tramas el puente debe disponer de suficiente memoria temporal para almacenar dicha trama. Y el reenvío de la trama.

A continuación se va a ver como trabaja un puente. Imagine dos redes LAN interconectadas por un puente (Fig. 5.21).

1. El puente lee las tramas transmitidas de la LAN B. Comprueba los campos destino y acepta aquellas tramas que van dirigidas a la LAN A.

2. Retransmite hacia A cada una de las tramas, haciendo uso del protocolo de control de acceso al medio de esta LAN.

3. El mismo proceso se sigue para el tráfico entre A y B.

5.6.3. Encaminadores o *Routers*

Los encaminadotes no solo incorporan la función de filtrado característica de los puentes sino que, además, determinan la ruta hacia el destino, empleándose tanto en redes locales como en redes de área extensa. Al ser capaces de indicar la ruta destino abarcan hasta el nivel 3 de OSI (Fig. 5.22).

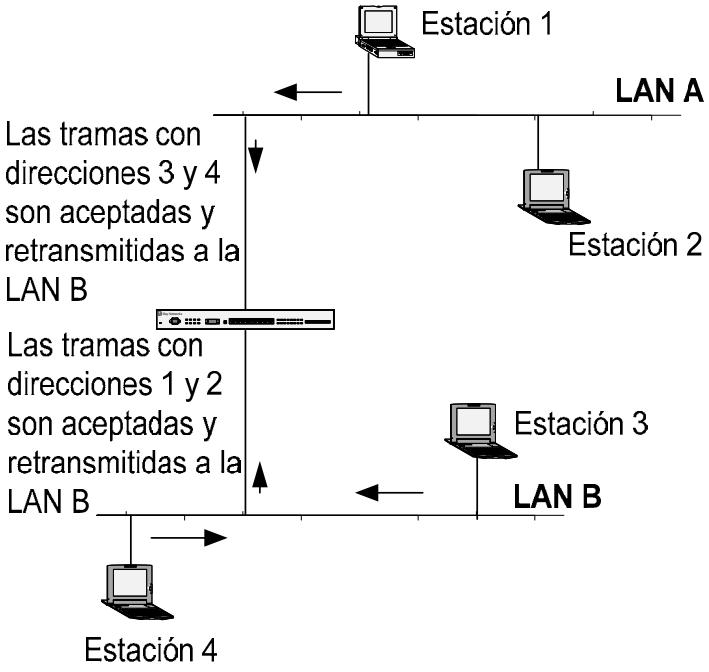


FIGURA 5.21. Puente que conecta 2 redes LAN.

Existen diferentes técnicas de encaminamiento. Algunas de ellas son: el encaminamiento estático, árbol de expansión y encaminamiento desde el origen.

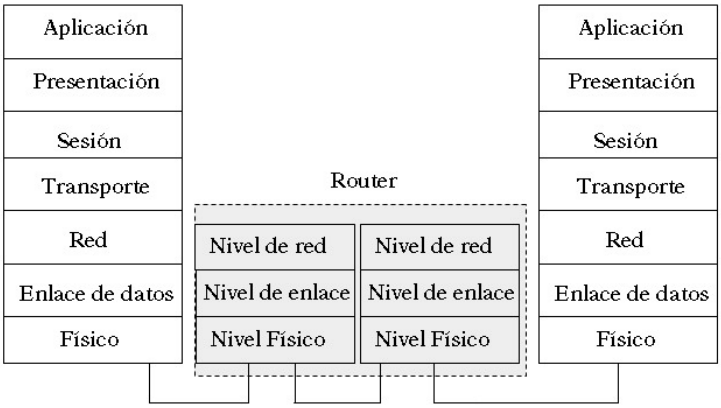


FIGURA 5.22. Esquema OSI de un encaminador o *router*.

5.6.4. Pasarela o *Gateway*

La pasarela o *gateway* realizan la traducción completa entre familias de protocolos, proporcionando conectividad extremo a extremo entre redes de distinta naturaleza (Fig. 5.23).

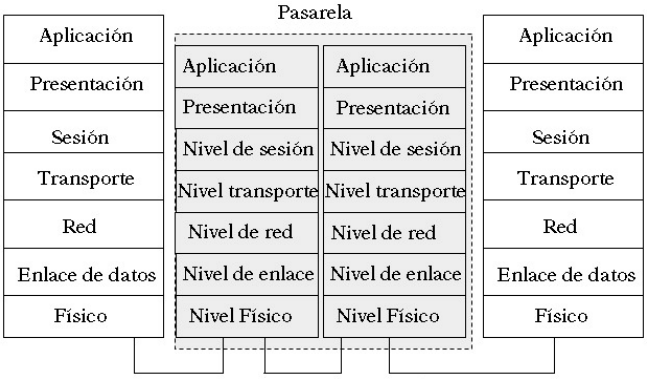


FIGURA 5.23. Esquema OSI de una pasarela o *gateway*.

5.7. REDES DE ALTA VELOCIDAD

El rápido avance de las tecnologías ha producido grandes cambios en aplicaciones, diseño y como no en las redes. En el apartado de Ethernet se

estudiaron las redes que tenía definidas en un principio el estándar 802.3. Actualmente dicho estándar comprende velocidades de transmisión de 10 Mbps, 100 Mbps, 1 Gbps y 10 Gbps.

A continuación se pasan a describir las redes de alta velocidad más importantes:

- Fast Ethernet y Gigabit Ethernet: es la extensión de la técnica de acceso múltiple con detección de portadora y detección de colisiones, o CSMA/CD, de 10 Mbps.
- Canal de fibra: este estándar proporciona una solución de bajo coste y fácilmente escalable para alcanzar tasas de datos elevadas en áreas locales.
- Redes LAN inalámbricas de alta velocidad. En el apartado 5.8. se hablará mas detenidamente de este tipo de redes.

En la Tabla 5.2 se pueden ver algunas de las características más importantes de las redes que se han mencionado hace un momento.

5.7.1. IEEE 802.3 100 Mbps (FAST ETHERNET)

Fast Ethernet es un conjunto de especificaciones desarrolladas por el comité 802.3. Con el objetivo de proporcionar una red LAN de bajo coste, compatible con Ethernet, que funcione a 100 Mbps. La definición genérica de estos estándares es 100BASE-T. El comité definió varias alternativas para los diferentes medios de transmisión (Tabla 5.3).

5.7.2. GIGABIT ETHERNET

A finales del año 1995 el comité IEEE 802.3 creó el grupo de alta velocidad con el objetivo de investigar estrategias para transmitir paquetes con formato Ethernet a velocidades del orden de Gigabits por segundo.

Gigabit Ethernet utiliza el mismo formato para las tramas y protocolos que el CSMA/CD utiliza en las versiones IEEE 802.3 a 10 Mbps y 100 Mbps. Actualmente se definen las siguientes alternativas para la capa física:

- 1000Base-SX. Utiliza longitudes de onda pequeñas, entre los 770 y los 860 nm. Proporciona enlaces dúplex de 275 m mediante fibras multimodo de 62,5 µm o hasta 550 m con fibras multimodo 50 µm.

- 1000Base-LX. Utiliza longitudes de onda mayores, entre los 1270 y los 1355 nm. Proporciona enlaces dúplex de 550 m mediante fibras multimodo de 62,5 µm o 50 µm, o de 5 km con fibras monomodo de 10 µm.

- 1000BASE-CX. Proporciona enlaces de 1 Gbps entre dispositivos localizados dentro de una habitación utilizando latiguillos de par trenzado apantallado de menos de 25 m. Cada enlace consiste en dos pares trenzados, apantallados, cada uno de los cuales se utiliza en un sentido.

- 1000BASE-T. Permite conectar dispositivos que están separados hasta 1000 m. Para ello utiliza cuatro pares no apantallados tipo 5.

TABLA 5.2. Características de algunas redes LAN de alta velocidad

	Fast Ethernet	Gigabit Ethernet	Canal de fibra	LAN inalámbrica
Velocidad de transmisión	100 Mbps	1 Gbps –10 Gbps	100 Mbps –3,2 Gbps	1 Mbps – 54 Mbps
Medio de transmisión	UTP, STP, fibra óptica	UTP, STP, fibra óptica	STP, fibra óptica, cable coaxial	Microondas 2,4 GHz, 5 GHz
Método de Acceso	CSMA/CD	Conmutado	Conmutado	CSMA/sondeo
Estándar	IEEE 802.3	IEEE 802.3	Asociación del canal de fibra	IEEE 802.11

En estos últimos años el incremento en el tráfico de Internet e intranets a motivado que la atención se haya desplazado a Ethernet con capacidad de 10 Gbps. Actualmente existen 4 opciones para su capa física:

- 10GBASE-S. Diseñada para transmisiones de 850 nm sobre fibras multimodo. Puede alcanzar distancias de hasta 300 m.

- 10GBASE-L. Diseñada para transmisiones de 1310 nm sobre fibras monomodo. Puede alcanzar distancias de hasta 10 Km.

- 10GBASE-E. Diseñada para transmisiones de 1550 nm sobre fibras monomodo. Puede alcanzar distancias de hasta 40 Km.

- 10GBASE-LX4. Diseñada para transmisiones de 1310 nm sobre fibras monomodo o multimodo. Puede alcanzar distancias de hasta 10 Km. Este medio utiliza multiplexación por división de longitud de onda (WDM *Wavelength Division Multiplexing*) para multiplexar el flujo de bits sobre cuatro ondas de luz.

TABLA 5.3. Alternativas de Fast Ethernet según el medio de transmisión

	100BASE-TX		100BASE_T4.	100BASE-FX
Medio de transmisión	2 pares STP	2 pares, UTP categoría 5	2 fibras ópticas	4 pares, UTP categoría 3, 4 o 5
Técnica de señalización	MLT-3	MLT-3	4B5B, NRZI	8B6T, NRZ
Velocidad de transmisión	100 Mbps	100 Mbps	100 Mbps	100 Mbps
Longitud máx. de seg.	100 m	100 m	100 m	100 m
Cobertura red	200 m	200 m	400 m	200 m

5.7.3. Canal de Fibra

Conforme se han ido aumentando las capacidades de un ordenador (velocidad, memoria, etc.) y las aplicaciones se han vuelto más complejas con una mayor dependencia de elementos multimedia (sonido, imágenes, etc.), la necesidad de mayor velocidad en el envió de datos al procesador ha aumentado. Este requisito afecta a dos métodos de conexión de datos con el procesador:

- El canal de entrada y salida. Manipulan por lo general transferencias entre procesadores y dispositivos periféricos, como por ejemplo CD-ROM o dispositivos de entrada y salida de audio. No se tiene en cuenta el formato o significado de los datos. El canal de entrada/salida transfiere los datos entre una memoria temporal en el dispositivo de origen y otra en el dispositivo de destino.

- Las comunicaciones de red. Gestionan las transferencias entre elementos de la red en distancias locales, metropolitanas o de área amplia.

El canal de fibra está diseñado para combinar las características más sobresalientes de estas tecnologías. Así, entre los tipos de recursos orientados a canal que se incorporan en la arquitectura de protocolos de canal de fibra se encuentran:

- Modificadores de tipos de datos para encaminar la carga útil contenida en las tramas sobre memorias temporales de interfaz especificas.

- Elementos de nivel de enlace asociados con operaciones individuales de entrada/salida.

- Especificaciones de la interfaz de un protocolo para dar soporte a arquitecturas de canal de entrada/salida existentes, como la interfaz SCSI (*Small Computer System Interface*).

Entre los tipos de recursos orientados a red incorporados en la arquitectura de protocolos de canal de fibra se encuentran:

- Multiplexación completa de tráfico entre múltiples destinos.

- Conectividad igual a igual entre cualquier par de puertos en una red de canal de fibra.

- Posibilidad de interconexión con otras tecnologías.

La asociación de industrias del canal de fibra, es el consorcio industrial encargado de promover el uso del canal de fibra. También enumera los requisitos que ambiciona conseguir:

- Enlaces full-duplex con dos fibras por enlace.

- Rendimientos desde 100 Mbps hasta 800 Mbps sobre una sola línea (de 200 Mbps a 1600 Mbps por línea full-duplex).

- Cobertura de distancia de hasta 10 Km.

- Alta capacidad de utilización independiente de la distancia.

- Mayor conectividad que los actuales canales de conexiones múltiples.
- Soporte para múltiples niveles de coste/rendimiento, desde pequeños sistemas hasta grandes computadores.
- Capacidad de transportar varios grupos de órdenes de interfaz para canales y protocolos de red ya existentes.

Por último se van a mostrar los distintos medios utilizados de canal de fibra (Tabla 5.4).

5.8. REDES INALÁMBRICAS

La aparición de los portátiles y la necesidad de obtener o enviar información a través de la red, aunque no se disponga de conexión física, dio lugar a que varios grupos empezarán a trabajar en redes inalámbricas. El método más práctico fue equipar los portátiles u ordenadores con transmisores y receptores de radio de onda corta que les permitiera comunicarse. Pronto, esto condujo a que diferentes empresas empezarán a comercializar las redes inalámbricas.

TABLA 5.4. Longitud máxima para los distintos tipos de canal de fibra

	100 Mbps	200 Mbps	400 Mbps.	800 Mbps
Fibra de modo simple	-----	10 km	10 km	10 km
Fibra multimodo de 50 µm	-----	2 km	1 km	0,5 km
Fibra multimodo de 62,5 µm	-----	1 km	1 km	175 m
Cable coaxial de video	100 m	100 m	71 m	50 m
Cable coaxial en miniatura	42 m	28 m	19 m	14 m
Par trenzado apantallado	80 m	57 m	46 m	28 m

El problema fue que no había compatibilidad entre ellas. Así un ordenador equipado con una radio de marca X no funcionaba con una estación base de marca Y. Finalmente, El IEEE se encargo de diseñar un estándar para redes de área local inalámbricas. Este estándar recibió el nombre de 802.11, aunque en jerga común se conoce como Wi-Fi.

El estándar propuesto tenía que trabajar en 2 modos:

- En presencia de una estación base (Fig. 5.24). Toda la comunicación se hace a través de la estación base. En la terminología del 802.11 se conoce como punto de acceso.

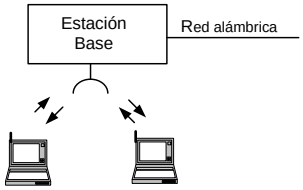


FIGURA 5.24. Red inalámbrica con estación base.

- En ausencia de una estación base (Fig. 5.25). Los ordenadores se pueden comunicarse directamente entre si. Este modo se llama a veces red ad hoc.

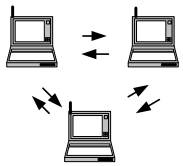


FIGURA 5.25. Red ad hoc.

5.8.1. Requisitos de las redes de área local inalámbricas

Una red LAN inalámbrica debe cumplir los mismos requisitos que cualquier otra red LAN, como alta velocidad, cobertura de pequeñas distancias, conectividad total entre las estaciones pertenecientes a la red y capacidad de difusión. Además de las mencionadas, existe un conjunto de necesidades específicas para entornos de redes inalámbricas. Una de las más importantes son:

- Rendimiento. El protocolo de acceso al medio debe hacer una utilización eficiente del medio inalámbrico para maximizar la capacidad.

- Número de nodos. Las redes LAN inalámbricas pueden necesitar dar soporte a cientos de nodos mediante el uso de varias celdas.

- Conexión a la LAN troncal.

- Área de servicio. Una zona de cobertura para una red LAN inalámbrica tiene un diámetro típico de entre 100 y 300 metros.

- Robustez de transmisión y seguridad. Debe permitir transmisiones fiables incluso en entornos ruidosos y debe ofrecer cierto nivel de seguridad como escuchas.

- Funcionamiento de redes adyacentes. Evitando transferencias entre ellas.

- Traspasos (*Handoff*)/Itinerancia (*Roaming*). El protocolo MAC usado en LAN inalámbricas debería permitir a las estaciones móviles desplazarse de una celda a otra.

- Configuración dinámica. El direccionamiento MAC y la gestión de la red LAN deberían permitir la inserción, eliminación y traslados dinámicos y automáticos de sistemas finales sin afectar a los demás usuarios.

5.8.2. Clasificación de LAN inalámbricas

Las redes LAN inalámbricas actuales se encuentran dentro de una de las siguientes categorías:

- LAN de infrarrojos (IR, *infrared*). Estas redes están limitadas a una sola habitación, ya que la luz infrarroja no es capaz de atravesar muros opacos. Principalmente se utiliza para el intercambio de datos de dispositivos móviles.

- LAN de espectro disperso. Hace uso de técnicas de espectro disperso como: el espectro ensanchado por salto de frecuencia (FHSS, *Frequency Hopping Spread Spectrum*) o el espectro disperso por secuencia directa (DSSS, *Direct Sequence Spread Spectrum*). En la mayoría de los casos las LAN de espectro ensanchado funcionan en bandas ISM (*Industrial, Scientific and Medical*) que no necesitan licencia.

- Microondas de banda estrecha. Operan en el rango de las microondas pero no hacen uso de espectro expandido. Muchas de estas LAN funcionan a frecuencias para la que es necesaria una licencia. Mientras que otras veces lo hacen en alguna de las bandas ISM.

Por último, se va a realizar una comparativa de cada una de estás LAN inalámbricas (Tabla 5.5).

5.8.3. Arquitectura del 802.11

En este apartado se pretende describir brevemente la arquitectura del 802.11. Para ello se comentará la pila de protocolos que define IEEE 802.11 y los servicios que proporciona.

5.8.3.1. Pila de protocolos

Los protocolos utilizados en todas las variantes 802, entre ellas Ethernet, tienen ciertas similitudes estructurales. En la Figura 5.26 se muestra una vista parcial de la pila de protocolos del estándar 802.11. La capa física se corresponde con la capa física de la OSI. En dicha capa se especifican 5 técnicas de transmisión.

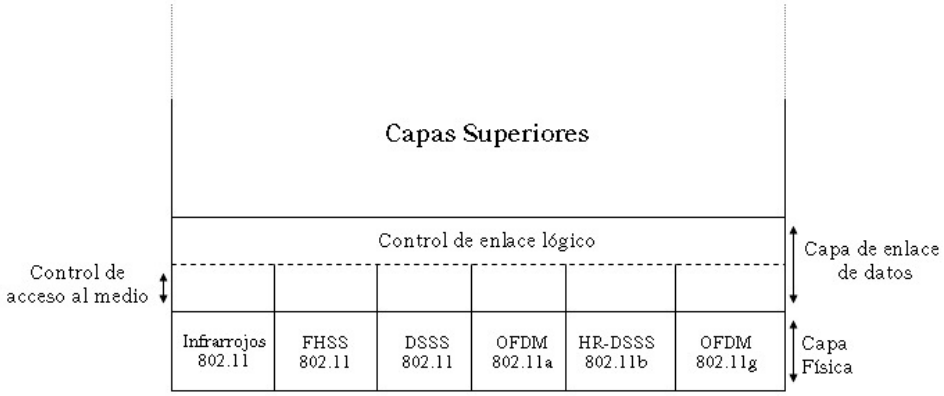


FIGURA 5.26. Pila de protocolos del 802.11.

5.8.3.1.1. Capa física

1. Infrarrojos

2. Espectro disperso con salto de frecuencia o FHSS. Esta técnica consiste en transmitir una parte de la información en una determinada frecuencia durante un intervalo de tiempo, pasado este tiempo se cambia la frecuencia de emisión y se sigue transmitiendo a otra frecuencia.

3. Espectro disperso de secuencia directa o DSSS. En esta técnica se genera un patrón de bits redundante (señal de chip) para cada uno de los bits que componen la señal. El estándar IEEE 802.11 recomienda un tamaño de 11 bits o chips, utilizando lo que se conoce como secuencia Barker. En recepción es necesario realizar el proceso inverso para obtener la información original. Utiliza modulación por desplazamiento en fase a 1Mbaudio.

4. Multiplexación por división de frecuencias ortogonales o OFDM. Esta técnica consiste en dividir la señal de radio en muchas subseñales que son transmitidas simultáneamente hacia el receptor en diferentes frecuencias. Existe una versión mejorada de esta técnica que se incluye en 802.11g.

TABLA 5.5. Comparativa de las LAN inalámbricas

	Infrarrojos		**Espectro disperso**		**Radio**
	Infrarrojos difusos	**Infrarrojos de haz directo**	**Salto frecuencia**	**Secuencia directa**	**Banda estrecha**
Velocidad (Mbps)	1 – 4	1 – 10	1 – 3	2 – 50	10 – 20
Movilidad	Estacionario/ móvil	Estacionario con LOS	Móvil	Estacionario/móvil	
Alcance (m)	15 – 60	25	30 – 100	30 – 250	10 – 40
Detectabi-lidad	Despreciable		Pequeña		Alguna
Longitud de onda/fre-cuencia	λ: 800-900 nm		902 – 928 MHz 2,4 – 2,4835 GHz 5,725 – 5,85 GHz		902 – 928 MHz 5, 2-5,775 y 18, 825-19,205 GHz
Técnica de modula-ción	ASK		FSK	QPSK	FS/QPSK
Potencia radiada	———————		< 1 W		25 mW
Método de acceso	CSMA	Anillo con paso de testigo, CSMA	CSMA		Reserva, ALOHA, CSMA
Necesidad de licencia	NO		NO		Si al menos que sea ISM

5. Espectro disperso de secuencia directa de alta seguridad o HR-DSSS. Está técnica es similar a DSSS. Utiliza 11 millones de chips/segundo para conseguir 11 Mbps en la banda de 2,4 GHz.

5.8.3.1.2. Capa de enlace de datos

La capa de enlace de datos definida en el estándar IEEE 802.11 se divide en dos subniveles o capas:

- La capa de control de acceso al medio o MAC. El estándar 802.11 soporta dos modos de funcionamiento:
  - Modo distribuido, donde la decisión de transmitir se reparte entre todos los nodos. Función de coordinación distribuida o DCF (*Distributed Coordination Function*). Esta función utiliza el protocolo de acceso múltiple con detección de portadora y evitación de colisiones o CSMA/CA (*Carrier Sense Multiple Access with Collision Avoidance*).
  - Modo centralizado, utiliza la función de coordinación puntual o PCF (*Point Coordination Function*). Se utiliza la estación base para controlar toda la actividad en su celda.

  Es importante remarcar que todas las implementaciones soportan DFC, y PCF es opcional, ubicándose encima del DFC (Fig. 5.27).

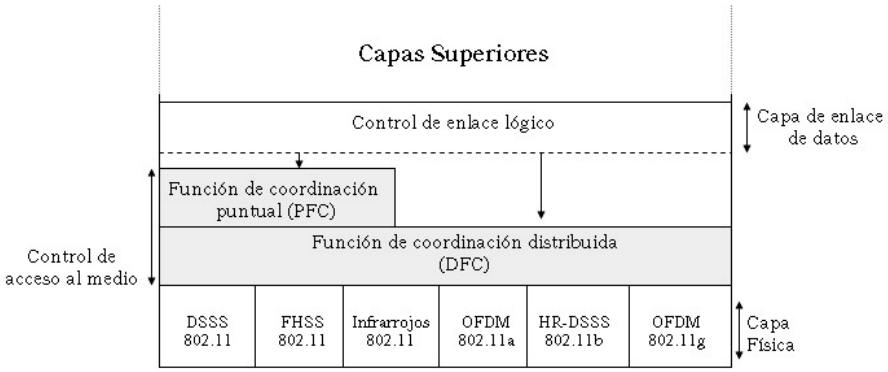


FIGURA 5.27. Arquitectura de la Capa MAC, 802.11.

- La capa de control de enlace. Se encarga de ocultar las diferencias entre las variantes 802 con el propósito de que sean imperceptibles por la capa de red. Ya se ha descrito en apartados anteriores.

5.8.3.2. Servicios

El estándar IEEE 802.11, establece que una red inalámbrica debe proporcionar nueve servicios. Estos se dividen en dos categorías:

- Servicios de distribución. Son proporcionados por las estaciones base y tienen que ver con la movilidad de la estación conforme entran y salen de las celdas, conectándose ellos mismos a la estación base y separándose ellos mismos de dichas estaciones. Define 5 servicios:

    1. Asociación: es utilizado por las estaciones móviles para conectarse ellas mismas a las estaciones base. La estación base puede aceptar o rechazar la conexión.

    2. Disociación: es posible que la estación base o la estación móvil se disocie, con lo que se rompería la relación. Por ejemplo una estación móvil podría utilizar este servicio antes de apagarse.

    3. Reasociación: permite a una estación cambiar su estación base predefinida.

    4. Distribución: determina como encaminar tramas enviadas a la estación base.

    5. Integración: traduce del formato 802.11 al requerido por la red destino. Así, imaginemos que una trama necesita enviarse a través de una red no 802.11 con un esquema de direccionamiento o formato de trama diferentes, esto traducción sería realizada por este servicio.

- Servicios de estación. Son los servicios relacionados con una sola celda de la red. Se utilizan después de que haya ocurrido la asociación y son:

    1. Autenticación: una estación debe autenticarse antes de que se le permita enviar datos.

2. Desautenticación: cuando estación previamente autenticada desea abandonar la red, se desautentica.

3. Privacidad: la información enviada a través de la LAN inalámbrica debe ser confidencial y por tanto debe codificarse.

4. Entrega de datos: se proporciona una forma de transmitir y recibir datos.

5.9. APLICACIONES

Las redes LAN tienen diversas aplicaciones. En este apartado se van a describir algunas de las más importantes.

5.9.1. Bus de campo o Fieldbus

Según la fundación de bus de campo. Un bus de campo es:

> Un enlace de comunicación, digital, bidireccional y multiacceso que permite la comunicación entre medidores inteligentes y dispositivos de control. Esto es servido como una red de área local para procesos de control avanzados, entradas y salidas remotas y aplicaciones de automatización de alta velocidad.

Su objetivo es reemplazar los sistemas de control centralizados por redes de control distribuido que permitan mejorar la calidad del producto, reducir los costos y mejorar la eficiencia. Para ello la información que envían y/o reciben los dispositivos de campo es digital, lo que resulta mucho más preciso que si se recurre a métodos analógicos.

Cada dispositivo de campo es un dispositivo inteligente y puede llevar a cabo funciones propias de control, mantenimiento y diagnóstico (Fig. 5.28).

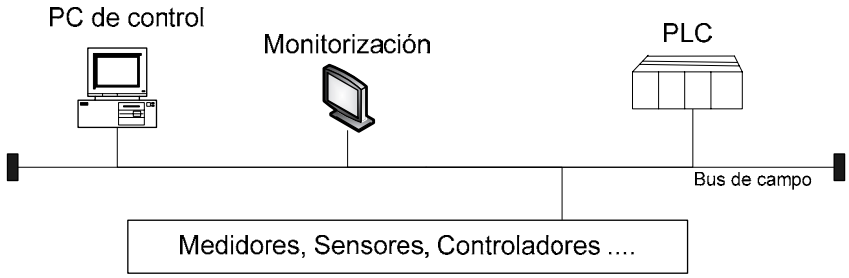


FIGURA 5.28. Ejemplo de un bus de campo.

5.9.1.1. Profibus

Es un estándar de bus de campo, desarrollado en el año 1987 por las empresas alemanas Bosch, Klöckner Möller y Siemens. Se trata de una red abierta, estándar, independiente de cualquier fabricante que cuenta con varios perfiles y que se adapta a las condiciones de las aplicaciones de automatización industrial.

Los perfiles con los que cuenta son:

- Profibus FMS (*Fieldbus Message Specification*). Es una solución orientada a resolver problemas de comunicaciones.

- Profibus DP (*Decentralized Periphery*). Es una solución de alta velocidad, diseñada y optimizada para la comunicación entre los sistemas de automatización y los dispositivos distribuidos.

- Profibus PA (*Process Automation*). Es una solución orientada a la aplicación de procesos

5.9.2. Redes LAN de ordenadores personales

Una parte importante de las empresas u organizaciones es la información. En muchos casos es necesario que dicha información sea compartida por los trabajadores de dichas organizaciones. Las redes LAN permiten compartir datos.

Actualmente no solamente se utiliza para compartir datos, sino recursos tales como aplicaciones, vídeos, etc.

5.9.3. Redes de respaldo y almacenamiento

Se utilizan para interconectar grandes sistemas como computadores centrales y dispositivos de almacenamiento masivo. Sus características fundamentales son:

- Alta velocidad.
- Interfaz de alta velocidad.
- Acceso distribuido.
- Distancia delimitada.
- Número limitado de dispositivos.

Normalmente las redes de respaldo se localizan en grandes compañías o en instalaciones con alto presupuesto para el procesamiento de datos.

5.10. CONCLUSIONES

En este capítulo se ha pretendido dar una visión global de que son y para que sirven las redes de área local. A la hora de diseñar un red LAN se deben tener en cuenta un conjunto de características como son la topología, el medio de transmisión, el acceso al medio, etc. Estas características están fuertemente relacionadas unas con otras.

El éxito de las LAN provoco que un gran número de instituciones y organizaciones definieran estándares. El instituto de ingenieros eléctricos y electrónicos o IEEE desarrollo un conjunto de estándares que definen las características de diferentes redes de área local como: IEEE 802.3 o el estándar 802.11 que habla acerca de las redes inalámbricas.

Las redes LAN suelen utilizarse principalmente para transmitir y compartir información, entre los diferentes estamentos de una empresa, a distancias pequeñas y alta velocidad. Esto no quita que sean utilizadas para otras misiones como procesos de control, de automatización o como redes de respaldo.

5.11. BIBLIOGRAFÍA

J. García Tomas. Alta velocidad y calidad de servicio en redes IP. Ed. RA-MA, 2002.

J. García Tomas. Redes para proceso distribuido: Área local, arquitecturas, rendimiento, banda ancha. Ed. RA-MA, 2001.

P. Mariño. Las comunicaciones en la empresa: Normas, redes y servicios. Ed. RA-MA, 2003.

W. Stallings. Comunicaciones y redes de computadores. Ed. Pearson, 2004.

A.S. Tanenbaum. Redes de ordenador. Ed. Prentice-Hall, 2003.

5.12. EVALUACIÓN

5.12.1. Evaluación objetiva

1. En la topología en bus:
   A) La comunicación entre el nodo y la toma de conexión es unidireccional
   B) Los terminadores tienen como misión absorber las señales que viajan por el bus
   C) La información se propaga a través del bus en un único sentido
   D) Cuando un nodo recibe información destinada a el, la elimina del bus

2. ¿Cual de las siguientes normas hace referencia a las redes LAN inalámbricas?
   A) IEEE 802.3
   B) IEEE 1394
   C) IEEE 802.11
   D) IEEE 802.14

3. LLC define 3 servicios. Cada servicio lleva asociada un tipo de operación. Indique cual de las siguientes afirmaciones es verdadera.
   A) El servicio no orientado a conexión sin confirmación tiene asociado el tipo operación 1
   B) El tipo de operación 3 tiene asociado el servicio orientado a conexión
   C) El servicio no orientado a conexión con confirmación tiene asociado el tipo operación 1
   D) El servicio no orientado a conexión sin confirmación tiene asociado el tipo operación 2

4. Se desea unir dos redes totalmente diferentes. Indique que dispositivo de conexión utilizaría.

   A) Repetidor

   B) Pasarela

   C) Puente

   D) Repetidor

5. Indique que protocolo de acceso al medio utiliza las redes Gigabit Ethernet.

   A) ALOHA puro

   B) Acceso Múltiple con Detección de Portadora no Persistente

   C) ALOHA ranurado

   D) Acceso Múltiple con Detección de Colisiones

6. Indique cual de las siguientes respuestas no es una técnica de transmisión definida por el IEEE 802.11.

   A) DSSS

   B) FHSS

   C) CSMA/CD

   D) OFDM

7. ¿Cuál de estos servicios no es de estación?

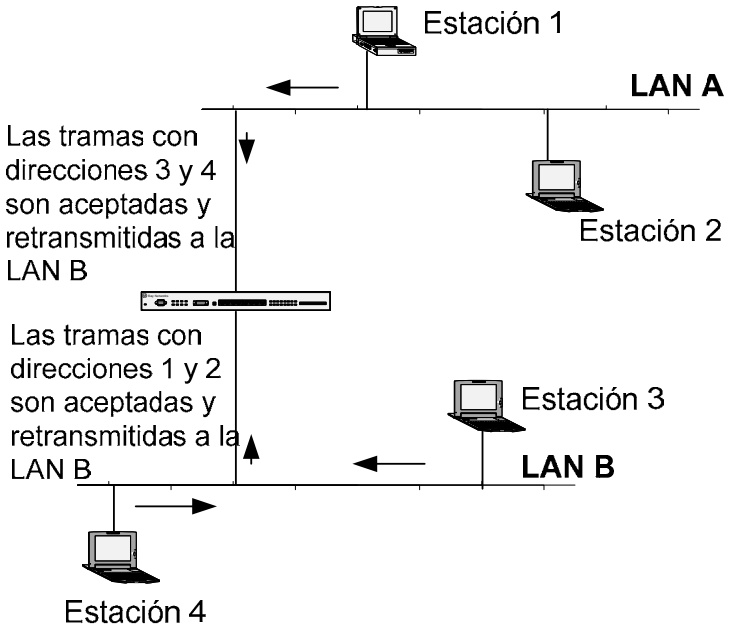
   A) Autenticación

   B) Entrega de datos

   C) Privacidad

   D) Asociación

8. En la siguiente figura se pueden observar 3 redes enlazadas por puentes.



Teniendo en cuenta que la estación 4 quiere enviar información a la estación 5. ¿Cuál de las siguientes afirmaciones es falsa?

A) La información enviada por la estación 4 se envía a las redes A y C por los puentes 1 y 2

B) La estación 3 comprueba que la información por el bus no va a ella y la deja pasar

C) El puente 2 comprueba la dirección destino de la información y la envía a la red C

D) La información que viaja por el bus es absorbida por los terminadores del bus

9. Indique cual de las siguientes opciones no es un perfil de Profibus:

A) Profibus FMS

B) Profibus DP

C) Profibus FM

D) Profibus PA

10. Los nodos o elementos de red de una topología en anillo están conectados a este mediante.
    A) Repetidor
    B) Pasarela
    C) Puente
    D) Repetidor

TEMA 6
REDES DE ÁREA AMPLIA. COMPONENTES. REDES IP

6.1. Introducción y objetivos
6.2. Características generales de las redes de área amplia
6.3. Protocolos de redes de área amplia: HDLC, PPP, FR, etc.
6.4. Las redes IP: redes WAN públicas
6.5. Direccionamiento IP y encaminamiento en IP
6.6. Los encaminadores como componentes típicos de WAN
6.7. Comunicación IP cliente/servidor mediante *sockets*
6.8. Operación de aplicaciones TCP, orientadas a conexión
6.9. Operación de aplicaciones UDP
6.10. Operación de aplicaciones del nivel de red
6.11. Conclusiones
6.12. Bibliografía
6.13. Evaluación

6.1. INTRODUCCIÓN Y OBJETIVOS

En este capítulo se hará una descripción de las características fundamentales de las redes de área amplia, tanto de las redes tradicionales basadas en protocolos de niveles bajos de la arquitectura OSI como de las redes IP públicas como Internet.

La pila de protocolos IP está tan ampliamente aceptada que puede afirmarse, sin lugar a dudas, que la mayor parte de protocolos de aplicación en el mundo de las comunicaciones se basa en ella. Por tanto es éste el capítulo apropiado para tratar, más allá de las características generales de cualquier red IP amplia, de los conceptos subyacentes al modelo de "red de redes", desde la estructura y funciones de las direcciones IP, verdaderos "carnés de identidad" de los equipos en las redes, del encaminamiento de mensajes en las redes IP, de cómo este proceso es gestionado fundamentalmente por los encaminadores y de qué protocolos de encaminamiento suelen utilizarse.

Asimismo se tratará de otros conceptos fundamentales de las redes IP, como los *sockets*, abstracción creada para poder diferenciar distintos servicios IP en los equipos de la red y cuya comprensión resulta esencial para entender el funcionamiento de cualquier aplicación cliente/servidor. Se hace énfasis en las diferencias importantes de aplicaciones con transporte TCP (orientadas a conexión) y de aplicaciones con transporte UDP (sin conexión previa). Finalmente se describen algunos protocolos importantes (como el ICMP) que funcionan directamente sobre transporte IP, en el nivel de red.

6.2. CARACTERÍSTICAS GENERALES DE LAS REDES DE ÁREA AMPLIA

Una red de área amplia sirve para crear conexiones de datos a través de un área geográfica extensa. Típicamente las empresas usan las redes de área amplia (WAN) para conectar diferentes redes locales y que la información se pueda distribuir entre oficinas geográficamente distantes (Fig. 6.1).

Debido al alto coste de construcción de una red global privada que conecte multitud de sitios remotos los servicios de una WAN son generalmente alquilados a proveedores de servicios.

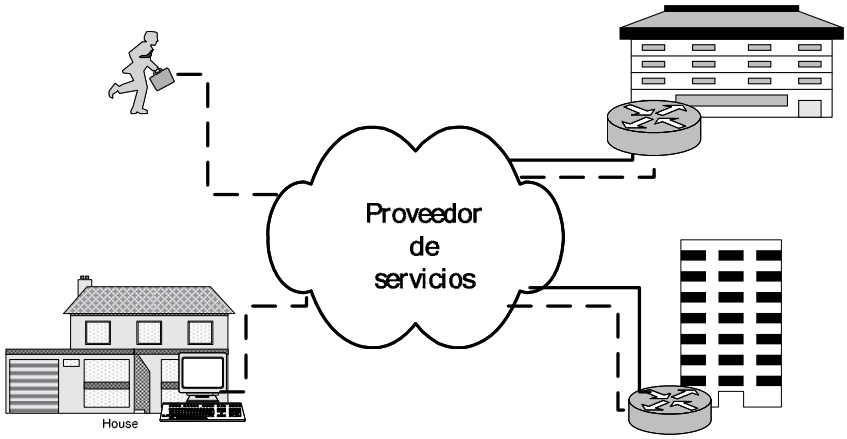


FIGURA 6.1. Servicios WAN.

La empresa que quiere usar la WAN contrata el servicio de un proveedor externo y puede usar sus recursos de red. Dentro de la nube (es decir, en la conexión WAN usada por la empresa), el proveedor de servicios usa una parte de su red para transportar la información de la empresa. Las necesidades de la conexión varían según precio y necesidades del tomador de los servicios.

Los tres tipos de conexión WAN disponibles son:

- **Líneas alquiladas**. Se conocen también como conexiones dedicadas o punto a punto. Proporcionan un camino único preestablecido de comunicación desde una red local, a través de la red del proveedor de servicios, hasta otra red local. Esta conexión es de uso privado del cliente y está reservada para el por el proveedor de servicios. Eliminan los problemas asociados con la compartición de la conexión entre varios clientes, pero son caras. Habitualmente son conexiones serie síncronas de hasta velocidades de E3, con disponibilidad garantizada del ancho de banda.

- **Conmutación de circuitos.** En este caso debe existir un camino formado por circuitos, dedicado, entre el emisor y el receptor durante la duración de la llamada. Se usa típicamente para proporcionar servicios telefónicos básicos o RDSI, Red Digital de Servicios Integrados. Estas conexiones son comunes en situaciones en las que se hace un uso sólo esporádico de la WAN.

- **Conmutación de paquetes.** En esta técnica los dispositivos de la red comparten un enlace punto a punto único para transportar la información en forma de paquetes de la fuente al destino a través de la red del proveedor. Estas redes usan circuitos virtuales para dotar al sistema de conectividad de extremo a extremo. Las conexiones físicas se consiguen mediante dispositivos de conmutación. En las cabeceras de los paquetes se identifica su destino. Los servicios ofrecidos son parecidos a los de la línea alquilada, pero la línea es compartida y el coste del servicio es menor.

6.3. PROTOCOLOS DE REDES DE ÁREA AMPLIA

El formato de la trama de nivel 2 de OSI en redes amplias depende del tipo de servicio elegido. A su vez, la selección del protocolo dependerá de la tecnología WAN seleccionada y del dispositivo de comunicaciones. En el caso de seleccionar una línea alquilada los protocolos más habituales son PPP (*Point to Point Protocol*), SLIP (*Serial Line Internet Protocol*) o HDLC (*High-level Data Link Control*). Si se ha optado por conmutación de circuitos la elección es semejante a la anterior y para el caso de conmutación de paquetes se puede seleccionar comúnmente entre X.25, Frame Relay o ATM (*Asynchronous Transfer Mode*).

A continuación se describen las características más significativas de algunos de estos protocolos, de uso común hoy en día en muchas redes de área amplia.

6.3.1. Protocolo HDLC

El protocolo HDLC (*High-level Data Link Control*) no soporta varios protocolos de nivel 3 en un enlace único, ya que no dispone de una manera estándar para indicar qué protocolo está transportando. Esta limitación ha sido vencida por distintos fabricantes que han modificado ligeramente la norma para hacerlo más flexible. Tal es el caso de *Cisco Systems* y su versión propietaria de HDLC, que usa en la cabecera un campo propio que hace posible que en el mismo enlace serie viajen tramas de diferentes protocolos de nivel 3. La Figura 6.2 muestra las tramas del HDLC estándar y de la versión propietaria de Cisco.

Es un protocolo de enlace de datos estándar ISO, derivado del antiguo SDLC de IBM, que encapsula datos en líneas serie síncronas.

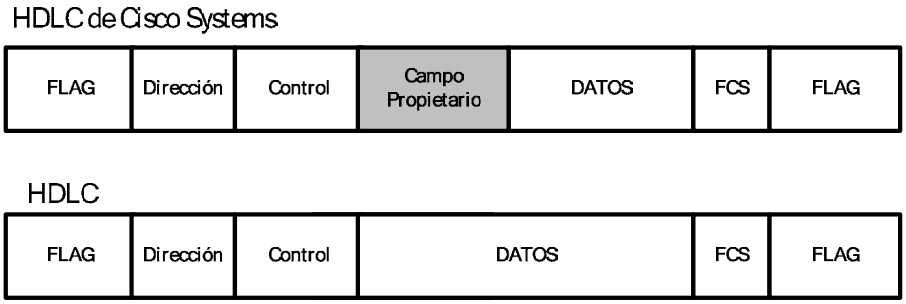


FIGURA 6.2. Formatos de las tramas HDLC.

6.3.2. Protocolo PPP

Los desarrolladores de Internet diseñaron PPP (*Point to Point Protocol*) para poder crear conexiones en enlaces punto a punto. El protocolo, descrito en los RFC 1332, 1661 y 2153 permite encapsular información de diferentes protocolos de nivel de red en enlaces punto a punto.

PPP puede ser el protocolo de selección en líneas serie asíncronas, síncronas y en líneas RDSI.

Desde un punto de vista funcional se puede decir que PPP es un protocolo de enlace de datos con servicios de nivel de red. Se puede ver PPP como dividido en dos subniveles. Estos subniveles aumentan la funcionalidad de PPP (Fig. 6.3).

PPP utiliza su componente NCP para encapsular diferentes protocolos de nivel 3 y su componente LCP para negociar y establecer las opciones de control sobre el enlace WAN.

Las opciones que ofrece PPP para establecer el enlace de datos son opciones de LCP y se usan principalmente para negociar y chequear las tramas. Tales opciones son:

- **Autenticación**. Mediante estas opciones el lado llamador de la comunicación informa de quien es para demostrar que tiene permiso administrativo para hacer la llamada. Desde los extremos se intercambian mensajes de autenticación, que pueden usar uno de dos

protocolos posibles: PAP (*Password Authentication Protocol*) o CHAP (*Challenge Handshake Authentication Protocol*).

- **Opciones de compresión.** Estas opciones permiten mejorar el rendimiento efectivo de las conexiones PPP al reducir la cantidad de datos en la trama que deben atravesar el enlace.

- **Detección de errores.** Permiten que cualquiera de los dos extremos identifiquen condiciones de error.

Opciones de PPP multi-enlace. Mediante estas opciones, en algunas plataformas, se puede conseguir un balanceo de carga en las interfases usadas por PPP. Se fragmentan los paquetes y se secuencian (siguiendo el RFC 1990) para dividir la carga de PPP y enviar fragmentos por circuitos PPP paralelos.

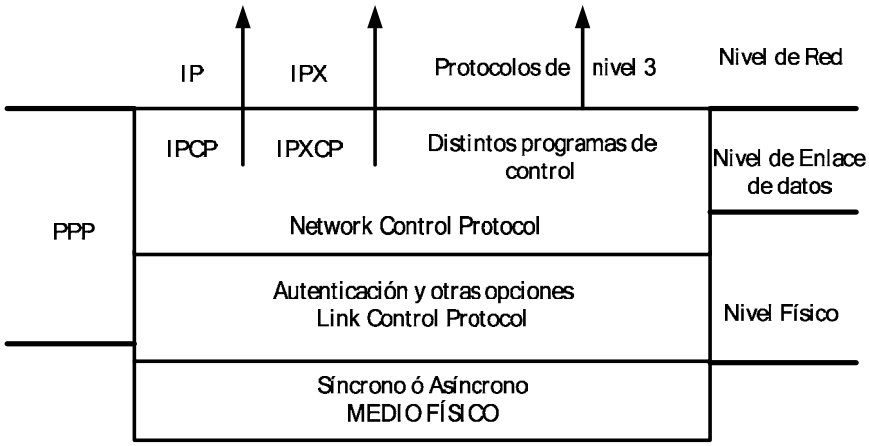


FIGURA 6.3. Subniveles del protocolo PPP.

6.3.3. Protocolo *Frame Relay*

Frame Relay es un estándar de ANSI y de la ITU-T que define el proceso de envío de datos por una red pública de datos. Es una tecnología de enlace de datos orientada a conexión, diseñada para conseguir alto rendimiento y eficacia. Depende de los protocolos de niveles superiores para la corrección de errores y su confianza se basa asimismo en las redes digitales y de fibras actuales, mucho más fiables que las menos modernas.

Frame Relay define el proceso de interconexión entre el encaminador del cliente y el conmutador de acceso local del proveedor de servicio (Fig. 6.4). No define cómo debe ser la transmisión de datos dentro de la nube del proveedor de servicios.

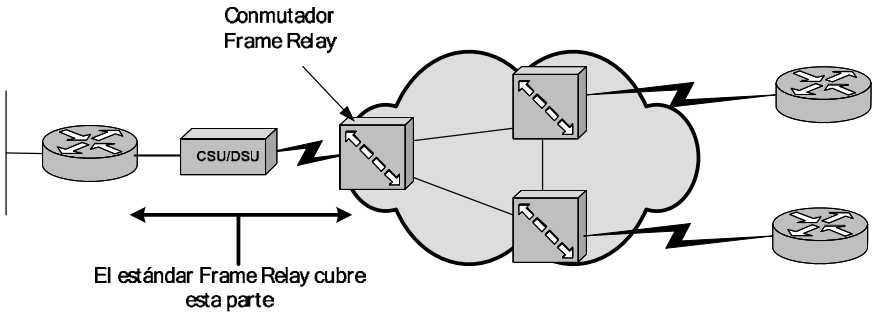


FIGURA 6.4. Introducción de *Frame Relay*.

Frame Relay constituye la evolución natural de las redes X.25 existentes en Empresas que han pasado de trabajar con terminales sin capacidad de proceso, en modo texto, realizando transacciones que generan bajo tráfico, a instalar ordenadores personales conectados en Redes de Área Local en sus dependencias, incorporando nuevas aplicaciones gráficas y multimedia, en un entorno distribuido, lo cual genera necesidades de grandes cargas de tráfico en periodos de tiempo muy cortos.

El protocolo proporciona un método de multiplexación estadística de muchas conversaciones lógicas de datos (circuitos virtuales) por un único enlace físico de transmisión. Para ello asigna identificadores de conexión a pareja de equipos terminales de datos (DTE). El conmutador del proveedor crea una tabla que asocia tales identificadores de conexión con sus puertos de salida de datos. Al recibir una trama, el conmutador analiza el identificador destino y reenvía la trama al puerto de salida asociado con el identificador. Antes de enviar la primera trama de la comunicación se establece el camino completo al destino.

Los circuitos virtuales de *Frame Relay* pueden ser de dos tipos: permanentes o conmutados. Los permanentes permiten ahorrar ancho de banda asociado con el establecimiento y la destrucción del circuito virtual y están asociados a circunstancias en las que debe existir comunicación permanente.

6.3.4. Dos aproximaciones al mismo problema: RDSI y ADSL

El CCITT (Comité Consultivo Internacional para Telegrafía y Telefonía) define RDSI (Red Digital de Servicios Integrados) (Fig. 6.5), como:

> Red que procede por evolución de la Red Digital Integrada y que facilita conexiones digitales extremo a extremo para proporcionar una amplia gama de servicios, tanto de voz como de otros tipos, y a la que los clientes acceden a través de un conjunto definido de interfaces normalizados.

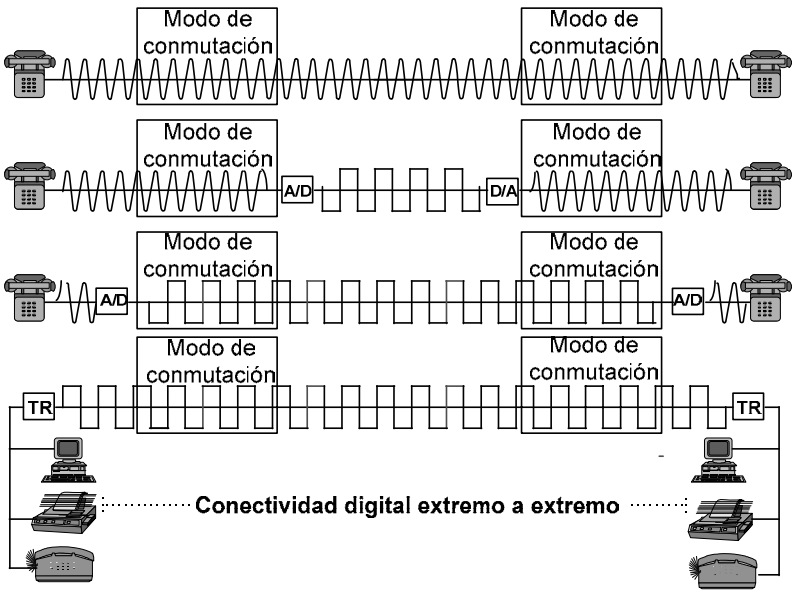


FIGURA 6.5. Esquema básico de una red RDSI.

La RDSI ha sido definida con un conjunto de normas, interfaces y configuraciones armonizadas y comunes a los principales países, lo que permite que los terminales y aplicaciones desarrolladas en un país puedan ser utilizados directamente en cualquier otro. La RDSI considerada hasta ahora se denomina de banda estrecha (RDSI-BE) puesto que trabaja básicamente con conexiones conmutadas de 64 kbit/s, pudiéndose llegar hasta los 2 Mbit/s.

Antes de describir los accesos de cliente a la red es necesario definir, porque así lo hacen las normas internacionales, las agrupaciones funcionales y puntos de referencia que configuran la RDSI, así como los canales de acceso a la red.

Las agrupaciones funcionales son el conjunto de equipos y elementos que configuran la RDSI. Se definen las siguientes agrupaciones funcionales:

- **Terminación de Red 1 (TR1):** Constituye la separación física entre la instalación del cliente y la red exterior. Realiza funciones de transmisión.

- **Terminación de Red 2 (TR2):** Realiza funciones tales como conmutación, concentración, control, mantenimiento, etc., en el interior de las instalaciones de cliente. Un ejemplo de TR2 puede ser una centralita privada o una red de área local. En el caso más sencillo, desaparece, coincidiendo entonces físicamente los puntos de referencia S y T.

- **Equipo Terminal 1 (ET1):** Son terminales que están diseñados para conectarse directamente a la RDSI. Como ejemplo pueden citarse los teléfonos digitales RDSI. Fax grupo 4, Videoconferencia, Videotex RDSI, etc.

- **Equipo Terminal 2 (ET2):** Representa cualquier terminal que no puede conectarse directamente a la RDSI. Para hacerlo necesitan de los Adaptadores de Terminal.

- **Adaptador de Terminal (AT):** A este equipo se le conectan, por un lado, los terminales que no pueden acceder directamente a la RDSI, y por otro se conecta él mismo a la RDSI. Permite por tanto la conexión de los ET2 a la RDSI.

Por otro lado, los puntos de referencia de RDSI son los puntos que separan las agrupaciones funcionales. En las normas internacionales se definen los siguientes:

- **Punto de Referencia S:** Constituye el punto de conexión física de los terminales con la RDSI.

- **Punto de Referencia T:** Representa la separación entre las instalaciones del cliente y los equipos de transmisión de línea (TR1).

- **Punto de Referencia U:** Define la línea de transmisión entre las instalaciones del cliente y la central RDSI. Se corresponde físicamente,

para el acceso básico RDSI, con el actual bucle de cliente analógico a dos hilos.

- **Punto de Referencia R:** Es el punto de conexión de cualquier terminal no RDSI, como los terminales modo paquete X.25, terminales con interfaz V.24 ó terminales con interfaz analógico a dos hilos.

En RDSI se han definido también los canales de acceso para la transferencia de información y señalización:

- **Canal B:** Es un canal de 64 kbps que transporta la información generada por el terminal del cliente.

- **Canal D:** Es un canal de 16 kbps o 64 kbps, dependiendo de la estructura de acceso del cliente, que se utiliza para transportar la señalización para el establecimiento y control de llamadas. También puede utilizarse para transmitir información de cliente a baja velocidad.

Los canales de información pueden combinarse para formar diversos tipos de acceso de cliente. Se han normalizado dos estructuras de acceso diferentes, atendiendo al número y tipo de canales de información y señalización que contengan, y se denominan **Acceso Básico y Acceso Primario**.

El acceso básico, o acceso 2B + D, proporciona dos canales de 64 kbps (canales B) y un canal de 16 kbps (Canal D) para la señalización y control de los canales B. En las instalaciones de cliente, la velocidad de transmisión de información es de 192 kbps (B+B+D+Control+Sincronismo+Mantenimiento) y está soportado físicamente por una instalación a cuatro hilos, dos para transmisión y dos para recepción, en configuración de bus de datos (denominado Bus Pasivo), que permite la conexión de hasta ocho equipos terminales. En el lado de red, desde la central RDSI hasta la instalación del cliente, se utiliza como soporte para la transmisión el mismo bucle de cliente analógico existente, con transmisión full-duplex y técnicas de cancelación de eco.

El acceso primario, o acceso 30B + D, proporciona 30 canales de 64 kbps (canales B) y un canal de señalización (canal D) de 64 kbps. En las instalaciones de cliente se dispone de 2048 kbps que, a través de una agrupación funcional TR2 (centralita digital multiservicio o equipo equivalente), se puede estructurar esta capacidad de transferencia de información en 30B + D.

Para enlazar las instalaciones del cliente con la central RDSI, se utiliza un sistema MIC a 2 Mbps que puede estar soportado en cable, radioenlace, etc.

La tecnología DSL (*Digital Subscriber Line*) es una tecnología de módem que utiliza las líneas telefónicas de par trenzado ya existentes para transportar datos de gran ancho de banda, como video u otros datos multimedia, hasta el abonado del servicio.

Los servicios xDSL son conexiones dedicadas punto a punto para acceso a la red pública sobre par trenzado en el bucle local entre el proveedor del servicio y el abonado.

La tecnología ADSL es asimétrica, lo que significa que permite mayor ancho de banda desde el proveedor hacia el cliente que en el otro sentido. Esta asimetría, unida a la característica de ser un servicio permanente (sin necesidad de establecimiento de llamada), hace a ADSL la selección más popular para cualquier tipo de servicio de acceso remoto en el que es mucho más importante la velocidad de "bajada" que la de "subida", es decir, sobre todo usuarios personales o pequeñas empresas sin servidores con muchas visitas.

ADSL permite transmitir más de 6 Mbps al abonado y hasta 640 kbps más en ambas direcciones.

Un circuito ADSL conecta un módem ADSL a cada extremo de una línea telefónica de par trenzado, creando tres canales de información:

- Un canal de alta velocidad de "bajada".
- Un canal duplex de velocidad media.
- Un canal de servicio telefónico básico.

Este último está separado del módem digital mediante filtros, garantizando así el servicio telefónico básico ininterrumpido, incluso en el caso de que ADSL falle.

El canal de alta velocidad varia entre 1.5 y 6.1 Mbps y el duplex entre 16 y 640 kbps. Además cada canal se puede multiplexar para formar múltiples canales de menor velocidad. Los módems ADSL proporcionan datos consistentes con los 1.544 Mbps del T1 de los EE.UU. y con el 2048 Mbps del E1 europeo y pueden adquirirse con distintas capacidades. En su configuración mínima se tiene 1.5 o 2.0 Mbps de "bajada" y un canal duplex de 16 kbps.

En un par de cobre la atenuación por unidad de longitud aumenta a medida que se incrementa la frecuencia de las señales transmitidas. Y cuanto mayor es la longitud del bucle, tanto mayor es la atenuación total que sufren las señales transmitidas. Ambas cosas explican que el caudal máximo que se puede conseguir mediante los módems ADSL varíe en función de la longitud del bucle de abonado.

Hasta una distancia de 2,6 km de la central, en presencia de ruido (caso peor), se obtiene un caudal de 2 Mbps en sentido descendente y 0,9 Mbps en sentido ascendente. Esto supone que en la práctica, teniendo en cuenta la longitud media del bucle de abonado en las zonas urbanas, la mayor parte de los usuarios están en condiciones de recibir por medio del ADSL un caudal superior a los 2 Mbps. Este caudal es suficiente para muchos servicios de banda ancha, y desde luego puede satisfacer las necesidades de cualquier internauta, teletrabajador así como de muchas empresas pequeñas y medianas.

Como antes se ha explicado, el ADSL necesita una pareja de módems por cada usuario: uno en el domicilio del usuario y otro en la central local a la que llega el bucle de ese usuario. Esto complica el despliegue de esta tecnología de acceso en las centrales. Para solucionar esto surgió el DSLAM ("*Digital Subscriber Line Access Multiplexer*"): un chasis que agrupa gran número de tarjetas, cada una de las cuales consta de varios módems, y que además concentra el tráfico de todos los enlaces ADSL hacia una red WAN.

La integración de varios módems en un equipo, el DSLAM, es un factor fundamental que ha hecho posible el despliegue masivo del ADSL. De no ser así, esta tecnología de acceso no hubiese pasado nunca del estado de prototipo dada la dificultad de su despliegue, tal y como se constató con la primera generación de módems ADSL.

6.4. LAS REDES IP: REDES WAN PÚBLICAS

La red Internet no es más que un ejemplo de una red de redes basada en protocolos IP. Es la más popular y la que más sirve como ejemplo del poder de este conjunto de protocolos que ha demostrado su valía cumpliendo, sin apenas modificaciones, 22 años con su versión actual, IPv4, y tiene ya hace años preparada y completamente desarrollada su nueva versión, IPv6. Ha demostrado ser un motor potente de comunicaciones para cualquier tipo de organizaciones, empresas o individuos que pretenden una comunicación fiable, a cualquier distancia, y de calidad.

Pero, además, ese modelo de red de redes, una única red creada a partir de la conexión inteligente de una multitud de redes, se ha impuesto como modelo de protocolos de comunicaciones para redes de empresa o de organizaciones, siendo así completo el dominio de tales protocolos, permitiendo tal nivel de normalización el desarrollo y puesta en marcha de múltiples tipos de aplicaciones y de programas en el mundo de las comunicaciones.

Parece, por tanto, muy normal que las redes IP se estén imponiendo también como redes WAN públicas, al ser simplemente una extensión del modelo de comunicaciones internas o de cada organización y favorecer la cooperación e integración consistente de comunicaciones entre redes de la misma o de distintas organizaciones.

Es lógico, entonces, que se deba entender los fundamentos básicos de tales redes para poder entender muchos de los modelos de protocolos de comunicaciones industriales, que se basan en este tipo de redes.

6.5. DIRECCIONAMIENTO Y ENCAMINAMIENTO EN REDES IP

El protocolo IP usa direcciones IP para identificar cada una de las interfaces de comunicación de cualquier equipo participante de una red IP. Cada interfaz se identifica en la red mediante, al menos, una dirección IP. En la mayor parte de los casos basta con una única dirección IP para cada equipo, y esta dirección es necesaria para poder encaminar los mensajes hacia el equipo. Si al hacer referencia a un equipo se utiliza su nombre de *host*, éste es traducido siempre a su dirección IP pues es sólo ésta la que viaja como dirección destino en cada mensaje generado con esa dirección destino. La traducción se hace mediante una pequeña base de datos local o configurando el equipo emisor como cliente de la aplicación DNS (*Domain Name System*) y, en este caso, el equipo debe tener accesible un servidor DNS.

Una **dirección IP** es un número binario de 32 bits o 4 octetos. Este número suele representarse en la notación “punto”, más fácil de leer y de escribir, en la que cada octeto de una dirección se convierte a su número decimal correspondiente. Por ejemplo, la dirección de www.uned.es es un número binario de 32 bits:

00111110 11001100 11000000 00010000

correspondiente al número:

62.204.192.16

en notación “punto”.

Conviene recordar que el número máximo representado para cada octeto es 255, correspondiente al número binario 11111111.

En la Figura 6.6 se muestra el formato general de una dirección IP, en el que se señalan las dos partes de la dirección: la **dirección de red y la dirección local.**

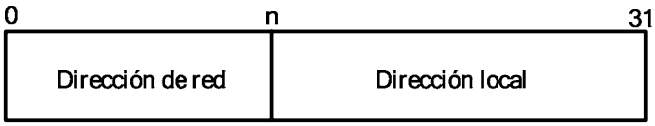


FIGURA 6.6. Formato básico de una dirección IP.

La dirección de red identifica la red a la que está conectado el equipo y la dirección local identifica un nodo particular y específico dentro de la red que puede ser, por ejemplo, la de una empresa particular. En esa red todos los equipos deben tener una dirección local única en el rango de la red para poder establecer comunicación con éxito.

Atendiendo al tamaño de las redes citadas se puede afirmar que dentro de una “red de redes” cada red particular pertenece a uno de tres formatos de direcciones de red. En la Figura 6.7 se muestran tales formatos, conocidos como formatos de redes de clase A, B o C y en la Tabla 6.1 se indican las características más importantes de cada una de ellas.

En los inicios de Internet a las organizaciones de gran tamaño, empresas como IBM o *Digital Equipment Corporation*, se les concedía direcciones de red de clase A, con una longitud de un octeto para la parte de red. El resto de los bits se emplean para dar direcciones únicas a los equipos de tal red. En la actualidad, y en Internet, no hay ya direcciones de red de clase A libres en la red Internet, lo cual no prejuzga su uso posible en una red interna de una organización.

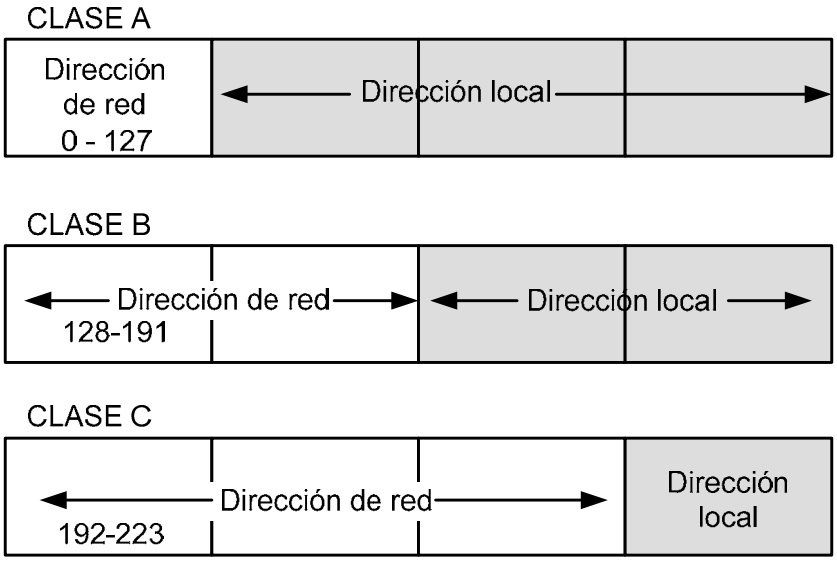


FIGURA 6.7. Clases de direcciones de una red IP.

TABLA 6.1. Características de las clases de direcciones IP

Clase	Nº octetos parte de red	Primeros bits	Primer octeto	Número de direcciones locales
A	1	0	0 - 127	16.777.216
B	2	10	128 -191	65.536
C	3	110	192 - 223	256

Las direcciones de red de clase B dedican los dos primeros octetos a la parte de red y los dos restantes a la parte local y las direcciones de red de clase C dedican los tres primeros octetos a parte de red y el último octeto a la parte local.

Una mirada a la Figura 6.7 basta para descubrir que, además de lo citado, cada clase de dirección está indicada por los primeros bits de cada dirección, tal como se indica en la Tabla 6.1. Ya sea en formato “punto” o en binario es fácil identificar la clase de una dirección IP mirando el primer número de la dirección o los primeros bits de la misma.

Además de los formatos citados hay otras dos clases de direcciones IP:

- Las direcciones de clase D, que empiezan con un número entre 224 y 239. Se utilizan para tráfico de multidifusión (*multicast*), que permite enviar un mensaje a un grupo de equipos distribuido en la red de redes sin tener que enviarlo una vez a cada uno.

- Las direcciones de clase E, que empiezan con un número entre 240 y 255, que están reservadas para uso experimental.

Además se han reservado (RFC 1918) varios bloques de direcciones para poder ser usadas en redes que no se van a conectar a Internet y que no necesitarán conectividad con otra organización. Estas direcciones son:

10.0.0.0-10.255.255.255

172.16.0.0-172.31.255.255

192.168.0.0-192.168.255.255

Estas direcciones de red no son encaminables en la red Internet. Ningún encaminador trata mensajes cuya dirección IP destino pertenezca a una de ellas.

Además hay otras dos direcciones en cada red cuyo uso está reservado para operaciones especiales:

- La propia dirección de red, que corresponde a poner a 0 todos los bits de la parte local de la dirección IP. Para la red 144.21 es la correspondiente 144.21.0.0 y, con este formato, aparecen en la tabla de rutas de cualquier equipo que use protocolos IP.

- La dirección de difusión completa (*broadcast*), que corresponde a poner a 1 todos los bits de la parte local de la dirección IP. En la red 144.21 sería la dirección 144.21.255.255. Estas direcciones se usan para enviar un mensaje a todos los equipos de la red simultáneamente.

Con esta arquitectura de direcciones es muy normal tener que dividir el espacio de direcciones del que se dispone de manera que coincida con la estructura física de la red. Para ello se utilizan las **subredes** o redes dentro de la propia red, un segundo nivel lógico, en buena parte independiente de lo comentado hasta ahora. Para ello, la parte local de la dirección IP se divide en una parte de subred y una parte de equipo, tal como puede verse en la Figura 6.8.

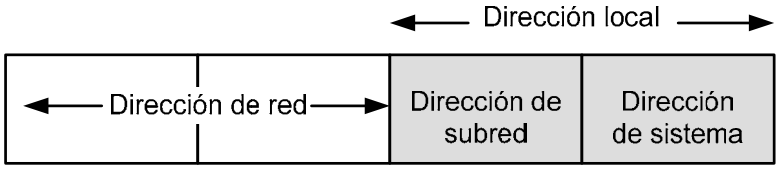


FIGURA 6.8. División de la dirección local en parte de subred y parte de equipo.

Las direcciones de subred, aún no siendo la única manera de gestión, suelen dividirse en octetos. Así, por ejemplo, una organización con direcciones de clase B, por ejemplo la 144.21.0.0, usará el tercer octeto para identificar la subred. Por ejemplo:

144.21.1

144.21.2

144.21.3

En este caso el cuarto octeto es el que se usa para identificar los equipos particulares de una subred.

Para poder decir que un equipo está correctamente preparado desde el punto de vista de su comunicación en una red IP debe haber sido configurado con:

- La dirección IP.
- La **máscara de red y subred**, que es otro número de 32 bits que tiene tantos bits a 1, desde el inicio del número, como bits hay dedicados en la dirección IP a la parte de red. Por ejemplo, en nuestro ejemplo anterior, un equipo con la dirección IP 144.21.4.44 tendría una máscara de red de 255.255.255.0.
- La dirección IP de, al menos, un encaminador IP. Si sólo hay uno se habla del **encaminador por defecto**.
- La dirección IP de, al menos, un servidor DNS, para poder trabajar en las aplicaciones utilizando nombres de equipos en lugar de direcciones IP.

El tráfico de mensajes IP se encamina hacia una dirección IP mediante el proceso de encaminamiento IP. En este proceso, el equipo emisor del mensaje consulta las partes de red y subred de la dirección destino, lo compara con su propia configuración de dirección de red y encaminará el mensaje usando su **tabla de rutas**, en la cual aparece la información de a dónde enviar los mensajes dependiendo de la dirección destino.

En el caso de un equipo son una sola tarjeta de red y una sola dirección IP es muy normal que el proceso de encaminamiento se traduzca en:

- Envío directo del mensaje al destino si éste pertenece a la misma subred del origen.
- Envío al encaminador correspondiente a la red de destino, obtenido de la tabla de rutas.
- Envío al encaminador por defecto en caso de que la red destino no tenga una entrada específica en la tabla de rutas.

Con esta arquitectura de una red de redes, cada una de las cuales con posibles divisiones en subredes, y haciéndose el encaminamiento entre redes IP mediante encaminadores, parece claro el papel esencial que juegan los encaminadores.

6.6. LOS ENCAMINADORES COMO COMPONENTES TÍPICOS DE REDES DE ÁREA AMPLIA

Un mensaje IP sigue un camino formado por una serie de saltos hasta su destino. Si el destino está en la red local el número de saltos es 1, pero si está más lejos el número de saltos es igual al número de encaminadores que tiene que atravesar.

Un encaminador tiene en general n interfases de comunicación, lo que significa que “conoce” n redes IP diferentes. Su tabla de rutas es más rica que la de un equipo normal y su objetivo de diseño es simple: encaminar mensajes IP hacia su destino. El proceso que lleva a cabo es también sencillo de entender. Para cada mensaje IP que entra por una de sus interfases:

1. Determina la dirección IP destino del mensaje, obteniéndola de la cabecera IP.

2. Busca en su tabla de rutas la entrada más parecida a la dirección destino.

3. Determina la interfaz de salida del mensaje, basándose en la información obtenida en el paso anterior.

4. Reenvía el mensaje por la interfaz determinada.

Parece pues evidente que cada encaminador debe disponer de una tabla de rutas lo más completa y bien configurada posible. Esta tabla de rutas tiene entradas con un formato muy sencillo. Los campos más relevantes de estas entradas son:

- Dirección de red destino.

- Interfaz de salida para esta dirección y/o dirección IP del siguiente encaminador en el camino del mensaje.

- Forma en la cual se obtuvo esta información.

La tabla de rutas tiene una entrada por cada interfaz configurada del encaminador. El resto de las entradas pueden ser estáticas o dinámicas.

Las entradas **estáticas** se llaman así por ser incluidas mediante la interfaz de configuración por el administrador del encaminador, que será el único que podrá asimismo eliminarlas o modificarlas.

En una red pequeña y que sufra pocas modificaciones en el transcurso del tiempo estas entradas suelen ser suficientes. Sin embargo para redes que cambian mucho al sufrir altas, bajas y modificaciones de nuevas redes y subredes hace falta otro tipo de configuración adicional, que sea capaz de cambiar dinámicamente conforme va cambiando la propia red. Estas entradas **dinámicas** de la tabla de rutas son configuradas dinámicamente mediante protocolos de encaminamiento.

Tales protocolos de encaminamiento son los responsables de mantener actualizadas las tablas de encaminamiento en los encaminadores y saber cuál es el mejor encaminador para el siguiente salto.

Dependiendo de los algoritmos que utilicen, de las métricas que usen (la métrica es una forma de decidir que ruta es mejor que otra) y de los consumos de recursos (ancho de banda, memoria, capacidad de cálculo) se dispone de distintos protocolos de encaminamiento.

Toda una serie de ellos se basan en el **algoritmo de "vector distancia"**. Suelen usar como única métrica el número de saltos entre origen y destino por cada uno de los caminos posibles. Entre los más utilizados se pueden citar:

- RIP (*Routing Internet Protocol*). Es quizás el más antiguo de los aún en uso. Funciona como una aplicación con transporte UDP. Tiene un límite máximo de 16 saltos. Como problemas típicos envía sus mensajes en difusión completa, no tiene mecanismos de autenticación y envía las tablas de rutas de cada encaminador cada 30 seg. Algunos de estos problemas se han corregido en la versión RIPv2.

- IGRP (*Internet Generic Routing Protocol*) y EIGRP (*Enhanced* IGRP). Funcionan como aplicaciones del nivel 3 de OSI, el mismo nivel en el que reside el protocolo IP. Permiten modificar las métricas clásicas, ajustando los valores no solo al número de saltos sino a características como el ancho de banda de cada uno de los posibles caminos. Son protocolos propietarios de *Cisco Systems*. El EIGRP está diseñado para grandes redes, envía mensajes en multidifusión, tiene capacidad de autenticación y hace un uso muy eficaz de recursos.

Los otros protocolos comunes de encaminamiento usan diferentes versiones del algoritmo de **"estado del enlace"**. Con estos protocolos cada encaminador crea una especie de mapa de la red completa y descubre todos los caminos desde el mismo a cualquier otro destino en la red. A cada enlace del mapa se le asigna una métrica y se calcula después el coste total combinando los costes de los enlaces y, finalmente, se elige el mejor camino para transmitir el tráfico. El protocolo más típico de este estilo es el OSPF (*Open Shortest Path First*) que funciona también en el nivel 3 como EIGRP por ejemplo.

Finalmente hay que decir que si la red es tremendamente extensa hay que acudir a otro tipo de protocolos de encaminamiento que utilicen conceptos que aparecen sólo para ese tipo de redes. Tal es el caso, por ejemplo, del protocolo BGP (*Border Galway Protocol*).

6.7. COMUNICACIÓN IP CLIENTE/SERVIDOR MEDIANTE *SOCKETS*

En el caso más general, cualquier aplicación de la familia IP (también conocida como servicio IP) está compuesta de la siguiente forma:

- Un servidor de la aplicación. Es un proceso del sistema operativo de un equipo, que atiende peticiones de conexión de clientes.

- Uno o varios clientes de la aplicación, que suelen residir en otros sistemas en la red.

Todos los servicios IP del nivel de aplicaciones (servidor Web, TELNET, FTP, RIP, TFTP, DHCP, correo electrónico SMTP, etc.) siguen esta estructura. Además, cada uno de sus correspondientes servidores crea, al arrancar, una estructura abstracta (***socket***) en la memoria del equipo, que está caracterizada, además de por otras cosas, por la dirección IP, el protocolo de transporte que usa (TCP o UDP) y el número de puerto en memoria asociado. Este número de puerto es, además, un estándar para muchas de estas aplicaciones. De hecho hasta el número de puerto 1023 se les conoce como los "puertos bien conocidos" (Tabla 6.2).

TABLA 6.2. Algunos números de puerto bien conocidos.

Protocolo de aplicación	Número de puerto
FTP-DATA	20
FTP-CONTROL	21
TELNET	23
SMTP	25
DNS	53
TFTP	69
HTTP	80

Cuando un cliente en la red quiere conectarse a un servidor envía un mensaje a la dirección IP del servidor indicando el número de puerto del mismo. A su vez, en su mensaje viaja la dirección IP origen del mensaje y el número de puerto del cliente. Este número no tiene por que ser siempre el mismo y se suele

usar uno cualquiera por encima de 1024. El servidor contesta dando servicio o no (como se ve en las siguientes secciones) dependiendo del protocolo de transporte. Si se establece el flujo de mensajes cliente/servidor, este flujo se puede caracterizar siempre casi completamente por 5 parámetros:

- Números de puerto del cliente y del servidor.
- Direcciones IP del cliente y del servidor.
- Protocolo de transporte utilizado, TCP o UDP.

Todos estos parámetros se pueden obtener, para cada flujo, mediante el comando *netstat*, que aparece en cualquier implementación de los protocolos IP, independientemente del sistema operativo.

Dependiendo de cuál sea el protocolo de transporte se usará una aproximación u otra para el control, gestión y eficacia de la transmisión. Así se pueden dividir las aplicaciones IP en:

- Aplicaciones **orientadas a conexión**. Son las que usan el protocolo TCP como protocolo de transporte. Sus diseñadores pueden dedicarse a la estructura y funcionamiento del protocolo de aplicación, ya que la fiabilidad de la transmisión, el control del flujo y el carácter de control del tráfico lo gestiona el protocolo TCP. Las aplicaciones más típicas son Telnet y sus variantes seguras, http, el correo electrónico SMTP y, en general, todas las aplicaciones de usuario final.

- Aplicaciones **sin conexión previa**. Son las que usan el protocolo UDP como protocolo de transporte. UDP sólo garantiza velocidad de transmisión. Todos los aspectos de control de flujo, fiabilidad, etc. deben ser tenidos en cuenta en el diseño del protocolo de aplicación. Las aplicaciones más típicas suelen ser las de gestión de red como TFTP, DHCP, SNMP, RIP y otras.

- Aplicaciones más sofisticadas, que combinan las dos anteriores. Entre ellas cabe destacar el DNS o servicio de nombres de dominio, el NFS para acceso a sistemas de ficheros remotos o muchas de las aplicaciones multimedia comunes en nuestras redes actuales.

6.8. OPERACIÓN DE APLICACIONES TCP ORIENTADAS A CONEXIÓN

El protocolo TCP (*Transmisión Control* Protocol) se caracteriza por ofrecer un comportamiento de "máquina finita" desde el primer momento, que permite que el propio protocolo realice funciones de control de flujo, control de congestión y fiabilidad de la transmisión. Para ello dispone de una serie de campos dinámicos en la cabecera TCP de cada mensaje que le permiten llegar a retransmitir, si fuera necesario, mensajes perdidos.

Cuando un cliente envía el mensaje citado de petición de conexión al servidor de una aplicación que funcione sobre TCP da comienzo el **establecimiento de una conexión TCP**, proceso que se realiza en tres pasos.

En el primer paso, este mensaje lleva en la cabecera TCP el bit de SYN a 1, el campo de ACK a 0 (este campo indica el número de octeto esperable desde el otro lado de la conexión y, en este mensaje debe ser evidentemente 0 pues aún no existe una conexión) y el campo de número de secuencia SEC a un valor cualquiera dado, por ejemplo, SEC-CLIENTE (Fig. 6.9).

Al llegar al servidor de la aplicación, TCP crea una entrada en la tabla de conexiones "embriónicas" (a medio hacer) en el servidor y contesta con un segundo mensaje al cliente. En este mensaje, en la cabecera TCP (que tendrá los números de puerto origen y destino invertidos con respecto al primer mensaje) los bits de SYN y de ACK tienen un valor de 1, el número de secuencia es otro, por ejemplo, SEC-SERVIDOR y el número de ACK es igual a SEC-CLIENTE+1, indicando que se ha recibido el mensaje anterior.

Cuando el cliente recibe este tercer mensaje contesta con otro mensaje al servidor, en el que, en su cabecera TCP, el bit de SYN está a 0, el bit de ACK está a 1, el número de secuencia es SEC-CLIENTE+1 y el número de ACK es SEC-SERVIDOR+1. Al llegar este mensaje al servidor, éste da por acabado el proceso, borra la entrada en la tabla de conexiones "embriónicas" y crea una nueva entrada en la tabla de conexiones activas. Desde este momento se puede ya comprobar la conexión mediante el comando *netstat* citado.

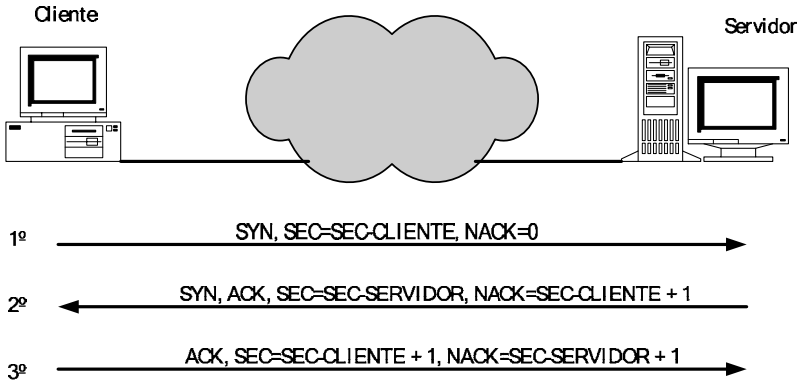


FIGURA 6.9. Establecimiento de una conexión TCP.

Este proceso puede iterarse para cada cliente nuevo y dependerá de la aplicación y de los recursos del equipo servidor la cantidad de conexiones simultáneas que aguante el servidor citado. Además, mediante el conocimiento de cliente y servidor del número inicial de octeto de envío del contrario y del campo de "tamaño de la ventana", TCP controla cuándo y cuántos octetos se pueden enviar en cada mensaje. Esto es así porque el tamaño de la ventana va cambiando dinámicamente, indicando el número de octetos admisibles para la siguiente transmisión, dependiendo del espacio de memoria asignado en cada momento para la comunicación específica en esa conexión. Este mecanismo, dinámico en ambos extremos, hace que cualquier comunicación TCP sea *full-duplex*, haciendo aún más interesante el método.

Entre los bits de señalización de la cabecera TCP, y además de los ya citados SYN (petición de sincronización) y ACK (indicación de haber recibido el mensaje anterior del otro lado de la conexión) se dispone también del bit URG para indicar que el mensaje lleva datos de control de la conexión TCP, del bit FIN para indicar que se quiere finalizar la conexión (este proceso es también muy concreto, haciendo intervenir 4 mensajes en un intercambio parecido al ya analizado) y del bit RST para poder romper una conexión de manera forzada.

Estas características son las más significativas de TCP, que le hacen especialmente fiable como protocolo de transporte para aplicaciones de usuario final en las que se desee tener un transporte fiable, ordenado y con control de flujo. Este es el caso, como ya se ha descrito, de la mayor parte de aplicaciones generadoras de tráfico en cualquier red IP.

6.9. OPERACIÓN DE APLICACIONES UDP

El protocolo UDP (*User Datagram* Protocol) es un protocolo de transporte diseñado para eficaz para transmisiones de aplicaciones cliente/servidor que cumplan dos características típicas:

1. Que la aplicación sea de respuesta única y rápida. EL procedimiento debe seguir la norma de que el cliente pregunta, el servidor responde y ya no hay más.

2. Que la cantidad de datos necesaria para contestar a cada petición sea corta. Si no puede ser así, la aplicación que use UDP debe estructurarse para repetir varias veces el procedimiento básico descrito.

Este protocolo no crea conexiones y, como consecuencia, no cumple nada de lo descrito en el apartado anterior, es decir, no garantiza la entrega de mensajes, no es capaz de secuenciar, no hace control de flujo ni de congestión.

La cabecera UDP es realmente pequeña, como consecuencia de todo lo citado, indicándose en ella sólo los números de puerto origen y destino, lo que hace al protocolo por otro lado un gestor eficaz del ancho de banda en cada mensaje de una aplicación que funcione sobre UDP.

Las aplicaciones típicas de UDP no son de usuario final sino de gestión. Entre ellas cabe destacar protocolos de encaminamiento como RIP o BGP, protocolos de asignación dinámica de direcciones IP, como DHCP, protocolos de transferencia de ficheros sin autenticación, como TFTP o protocolos de gestión de red como SNMP.

6.10. OPERACIÓN DE APLICACIONES DEL NIVEL DE RED

La pila de protocolos IP es anterior a la arquitectura OSI y, debido a ello, no es extraño encontrar otro tipo diferente de aplicaciones a las citadas hasta ahora, que tienen su transporte en el propio protocolo IP y no hacen, por tanto, uso ninguno del concepto de *socket* ya analizado. Son las aplicaciones del nivel de red. Entre ellas cabe destacar las aplicaciones de protocolos de encaminamiento, como los ya citados OSPF, IGRP y EIGRP.

Se debe destacar también, como parte de los citados protocolos del nivel de red, el protocolo ICMP (*Internet Control Message* Protocol). Este protocolo es el encargado de tratar de resolver una serie de problemas comunes en cualquier red IP. Entre ellos los más significativos son:

- Que un encaminador no funcione correctamente y se extravíen mensajes.
- Que se intente enviar mensajes a servidores que estén inalcanzables, ya sea por no existir ningún equipo con la dirección pedida, por no existir el proceso servidor o por ser filtrado el mensaje por un cortafuegos.
- Que se envíe un mensaje a través de un encaminador habiendo otro que utiliza una ruta mejor.
- Que un encaminador no puede procesar mensajes a la velocidad a la que le están llegando, pudiéndose haber un problema de congestión en la red.

ICMP trata de solucionar todos estos problemas, además de ofrecer otros servicios, mediante mensajes sencillos compuestos por la cabecera IP (con número de protocolo 1, para ICMP) y una cabecera de ICMP. En esta última (Fig. 6.10) hay dos campos (tipo y código) que indican el tipo de mensaje ICMP.

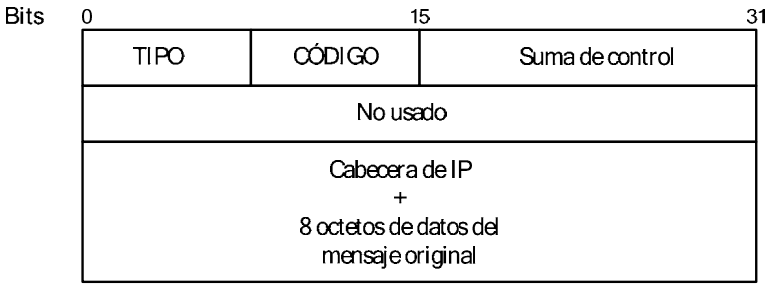


FIGURA 6.10. Cabecera del protocolo ICMP.

Quizás el caso más conocido del uso de ICMP sea el generado por el uso del comando *ping*, que permite saber si un equipo de dirección IP determinada está accesible en la red. El equipo que genera el ping envía mensajes de datos de tipo 8 y código 0, siendo la cantidad de datos configurable en el origen. Si el equipo destino está activo contesta con los mismos datos en otro mensaje ICMP de tipo 0 código 0. De esta familia de mensajes de petición ICMP tiene también los de petición de máscara de dirección de un equipo (tipos 17 y 18) o de petición de marca de tiempo de un equipo (tipos 13 y 14).

Pero donde mejor se ve el uso inteligente de ICMP es en los encaminadores. Si en un encaminador intermedio entre el origen de un mensaje y su destino se determina que hay algún tipo de problema asociado con el mensaje, el encaminador enviará un mensaje ICMP al origen indicando que problema ha habido.

Entre estos los más comunes son:

- Mensajes de destino inalcanzable. El encaminador indicará al equipo origen del mensaje que genera un error cuál es la causa mediante un mensaje ICMP de tipo 3. Como es fácil de suponer las razones pueden ser muchas y variadas, indicándose por el campo código de la cabecera de ICMP. Varían entre valor 0, red inalcanzable, hasta valor 12, no se puede llegar al *hos*t por el tipo de servicio requerido, pasando por que el destino no tenga el protocolo solicitado, hay que fragmentar el mensaje pero no está permitido, la ruta no es correcta, etc.

- Mensaje de tiempo de vida expirado. Cada mensaje IP lleva, en su cabecera IP, un campo TTL que se decrementa en una unidad al pasar por un encaminador. Si al llegar a un encaminador el campo vale 1, el encaminador lo convierte en 0, descarta el mensaje y envía un mensaje ICMP a la dirección IP origen indicando tiempo de vida expirado, con el tipo 11.

- Mensaje de petición de control de congestión, indicando una petición de desaceleración de envío de mensajes al encaminador, con el tipo 4.

- Mensaje de redirección de los mensajes, indicando otra ruta mejor, con el tipo 5.

Además, para que el equipo origen pueda discernir qué mensaje ha tenido cada uno de los posibles problemas, cada mensaje ICMP lleva adjunto la cabecera de IP y los 8 primeros octetos del mensaje original que provocó el mensaje de error de ICMP (Fig. 6.10).

Finalmente cabe destacar también el uso inteligente que se hace de este protocolo en el desarrollo de aplicaciones de control de caminos alternativos para redes extensas mediante aplicaciones como el *traceroute* que es capaz de descubrir el camino concreto de los mensajes IP de origen a destino en cualquier red usando mensajes UDP con diferentes tiempos de vida TTL para desencadenar mensajes ICMP por cada encaminador visitado.

6.11. CONCLUSIONES

A fin de completar las ideas básicas de protocolos de comunicación genéricos se ha hecho una descripción de algunos de los más utilizados en conexiones WAN, tanto para servicios punto a punto como para redes de conmutación de paquetes, para pasar a tratar a las redes IP como lo que verdaderamente son hoy en día: la arquitectura de comunicaciones de transporte de datos prácticamente universal.

En este sentido se ha hecho una exposición de los puntos fundamentales de la citada pila de protocolos como el esquema de direccionamiento IP, las distintas clases de direcciones, las direcciones prohibidas, las subredes como herramientas de gestión para administrar mejor las direcciones IP, etc.

Para cualquier técnico de comunicaciones que deba gestionar aplicaciones en redes IP es esencial conocer cómo es el proceso de encaminamiento en tales redes. Con ese objeto se ha explicado el proceso de encaminamiento de mensajes de cualquier equipo que resida en una red IP, especialmente en el caso de los encaminadores, verdaderos puntos clave de cualquier red extensa. Se ha analizado después el concepto de *socket*, fundamental para entender el proceso de comunicaciones cliente/servidor de cualquier servicio IP. Se ha descrito también los distintos servicios que ofrecen los protocolos de transporte, TCP y UDP, para los distintos tipos de aplicaciones IP, exponiendo claramente cómo TCP ofrece un servicio más lento pero más fiable, justo al revés que UDP.

Finalmente se ha analizado también la estructura de funcionamiento de aquellas aplicaciones IP, como ICMP o EIGRP, que utilizan como transporte el propio protocolo de red IP.

6.12. BIBLIOGRAFÍA

P. Mariño. Las comunicaciones en la empresa. Normas, redes y servicios. Ed. RA-MA, 2003.

J. García Tomás, S. Ferrando y M. Piattini. Redes para proceso distribuido. Ed. RA-MA, 2001.

A.S. Tanenbaum. Computer networks. Ed. Prentice-Hall, 1996.

D.E. Comer. Redes globales de información con Internet y TCP/IP, principios básicos, protocolos y arquitectura. Ed. Prentice-Hall Hispánica, 1996.

S. Feit. TCP/IP. Arquitectura, protocolos, implementación y seguridad, Ed. McGraw-Hill, 1997.

6.13. EVALUACIÓN

6.13.1. Evaluación objetiva

1. ¿A qué dirección de red y dirección local pertenece un equipo cuya configuración IP está dada por los números 22.1.2.5 255.255.0.0?

   A) Dirección de red 22, dirección local 1.2.5

   B) Dirección de red y subred 22.1.2, dirección local 5

   C) Dirección de red y subred 22.1, dirección local 2.5

   D) Falta información para poder contestar a la pregunta

2. En un mensaje TCP entre dos equipos IP se observa su cabecera obteniéndose que SYN=1, ACK=1 y el número de secuencia SEC=123. ¿Cuál será la estructura correspondiente en la cabecera TCP del mensaje de contestación a este mensaje?

   A) SYN=1, ACK=0, SEC-OTRO=124

   B) SYN=0, ACK=1, SEC-OTRO=124

   C) SYN=1, ACK=1, SEC-OTRO=123

   D) SYN=0, ACK=0, SEC.OTRO=124

3. ¿A qué clase de direcciones IP pertenece la dirección 172.17.12.122 con máscara de red 255.255.255.0?

   A) A una red de clase A

   B) A una red de clase B

   C) A una red de clase C

   D) A una red de clase D

4. Si hay una línea punto a punto entre un encaminador de Cisco y uno de Nortel y se usa en ambos el protocolo HDLC, ¿es esperable algún problema en la comunicación?

   A) Si, que no lleguen a afectarla correctamente pues la cabecera de HDLC de *Cisco Systems* no cumple la norma estándar de HDLC

   B) Si, que haya problemas intermitentes de pérdida de comunicación, típicos de una línea punto a punto

   C) No se debe esperar ningún tipo de problema

   D) Si, que las transmisiones TCP tengan un retardo inaceptable, debido a la naturaleza lenta del protocolo HDLC

5. ¿Qué pieza básica para el encaminamiento de mensajes IP comparten todos los equipos, del tipo que sean, que forman parte de una red IP?

   A) Que el software de comunicaciones permita la asignación de varias direcciones IP a la misma tarjeta de red.

   B) Que los ficheros de configuración de números de puerto normalizados coincidan en todos los equipos

   C) Que sigan el mismo formato de máscaras de red en sus direcciones

   D) La tabla de rutas, verdadera brújula del encaminamiento IP, que indica hacia donde enviar cada mensaje para que llegue a su destino

6. ¿Cuál de las siguientes es una descripción adecuada del protocolo CHAP?

   A) Es un protocolo de aplicación de IP, con transporte UDP

   B) Es un protocolo de recuperación de errores que funciona en colaboración con *Frame Relay* en redes amplias

   C) Es un protocolo opcional de autenticación para la fase de establecimiento de conexión, dentro del protocolo PPP

   D) CHAP no es un verdadero protocolo sino parte de la implementación de los *sockets* IP dentro de los sistemas *Windows* de *Microsoft*

7. ¿Cuál es el número de puerto asociado con el protocolo de encaminamiento EIGRP?

   A) El número de puerto es el 69

   B) El número depende del transporte. Para UDP es el 69

   C) No tiene número de puerto, al ser IP su transporte

   D) No tiene número de puerto al no necesitar transporte. Funciona directamente sobre protocolos de nivel 2, como Ethernet

8. Cuando se intenta una conexión web con un servidor utilizando su nombre, ¿en qué cabecera del mensaje IP generado viaja tal nombre?

   A) En la cabecera propia de DNS

   B) En la cabecera de TCP, junto con la dirección IP

   C) En la cabecera de IP

   D) En ninguna cabecera de este mensaje viaja el nombre.

9. ¿Cómo tratará un encaminador un mensaje entrante cuya dirección IP destino sea la 192.168.1.222?

   A) Lo reenviará a su encaminador por defecto.

   B) Lo tratará en el proceso de protocolo de encaminamiento con el que trabaje.

   C) Lo descartará completamente y no lo tratará.

   D) Lo reenviará a la dirección 192.168.255.255.

10. En una red RDSI, ¿es verdadero o falso que el canal D siempre dispone de 64 kbps?

    A) Verdadero, gracias a ello se puede conseguir el establecimiento de conexión en digital.

    B) Falso, depende del tipo de servicio RDSI.

    C) Verdadero, de esta manera se pueden manejar diferentes conexiones TCP, además de telefónicas, simultáneamente.

    D) Falso, en ningún servicio se llega más allá de 16 kbps.

TEMA 7

SISTEMAS DE COMUNICACIONES INALÁMBRICOS Y MÓVILES

7.1. Introducción y orientaciones para el estudio
7.2. Objetivos
7.3. Conceptos básicos de la propagación inalámbrica
7.4. Diseño de redes inalámbricas y móviles
7.5. Redes de área local móviles (WLAN)
7.6. Redes de área personal (WPAN)
7.7. Redes fijas de acceso inalámbrico (WMAN)
7.8. Redes de acceso celular (WWAN)
7.9. Redes de satélites espaciales
7.10. Las redes inalámbricas y los buses de campo
7.11. Conclusiones
7.12. Bibliografía
7.13. Evaluación

7.1. INTRODUCCIÓN Y ORIENTACIÓN EN EL ESTUDIO

Este capítulo presenta los tipos de redes no cableadas disponibles para el acceso de la información necesaria para la prestación de los servicios de las Tecnologías de la Información y las Comunicaciones (TIC), desde terminales fijos, móviles y transportables. Se comienza con el estudio de la problemática que presenta la propagación inalámbrica, las contramedidas que mitigan sus efectos en las comunicaciones, y las métricas que permiten evaluar las prestaciones de las redes inalámbricas y móviles existentes. Desde un punto genérico se plantea el diseño de dichas redes a partir de una configuración general para simplificar su estudio. A continuación se describen las diferentes normas agrupadas por su área de cobertura en WLAN, WPAN, WMAN y WWAN, además de algunos servicios de las redes de satélites que tienen interés en las aplicaciones inalámbricas y móviles. Finalmente se analiza el interés de las redes de radio en el contexto de las aplicaciones industriales y su impacto en los buses de campo, y se plantean algunas alternativas de configuración dada la inexistencia de normas o protocolos inalámbricos en el ámbito industrial.

7.2. OBJETIVOS

La comprensión de las diferentes normas de redes de radio implica una interpretación correcta de los protocolos que se definen en las diferentes capas del modelo OSI. En las redes inalámbricas y móviles, el conocimiento de los protocolos específicos de las capas física y de enlace está muy relacionado con los problemas que el medio físico presenta a la propagación de las señales electromagnéticas, es decir, con las características del canal de radio, en términos de frecuencias de portadoras y ancho de banda. La cuantificación de estas características se expresa mediante la definición de métricas de prestaciones con las que se evalúan las normas internacionales, de acuerdo con las aplicaciones a las que están orientadas.

El objetivo fundamental de este capítulo consiste en proporcionar al alumno los conceptos fundamentales que le permitan comprender las evaluaciones cualitativas y cuantitativas, que justifiquen la utilización de cualquier norma inalámbrica y móvil en las aplicaciones industriales.

7.3. CONCEPTOS BÁSICOS DE LA PROPAGACIÓN INALÁMBRICA

7.3.1. Efectos de la propagación

La propagación inalámbrica (*wireless*) presenta características específicas respecto a la cableada, debido a que la interacción de las señales electromagnéticas en el espacio libre (espacio exterior, atmósfera, zonas rurales, zonas urbanas e interior de instalaciones) se traduce en los canales de radio en perjudiciales fenómenos de desvanecimiento, interferencia y distorsión, que afectan negativamente al correcto intercambio de información entre los equipos de comunicaciones estáticos, que al contrario de lo que sucede en los cables, han de compartir el medio indiscriminadamente con todo tipo de sistemas públicos y privados. Si además los equipos de comunicaciones están en movimiento, a los efectos anteriores hay que añadir los que introduce la movilidad (*mobile*) por los cambios continuos de la trayectoria y el entorno geográfico (configuración del terreno, obstáculos de todo tipo, arbolado, climatología, distancia, orientación, etc.) a los que se somete el canal de radio con diferentes velocidades de desplazamiento relativo.

De la misma manera que en los medios cableados, estos efectos perniciosos se manifiestan en pérdidas de señal que reducen la relación S/N en el receptor, haciendo la información ininteligible, en errores en los bits en las comunicaciones digitales (aumento del BER), y en retardos que afectan a servicios en tiempo real, como son los casos de la telefonía, videoconferencia, sistemas de control en bucle cerrado, y cualesquiera otros servicios interactivos. Es por ello que buena parte de la complejidad de los equipos de comunicación por radio es debida a la necesidad de introducir circuitos electrónicos que mitiguen en lo posible todos los efectos negativos que afectan a los canales de información.

A continuación se describen estos efectos perjudiciales y posteriormente algunos métodos para reducirlos, ya que es imposible su total eliminación. En este sentido hay que considerar la existencia de diferentes normas de redes inalámbricas y comunicaciones móviles, como conjuntos particulares de soluciones a los problemas de propagación en un entorno geográfico de aplicación definido y para unos servicios concretos, aparte de las selecciones de márgenes de frecuencias para las portadoras (espectro electromagnético) y de anchos de banda de los canales de radio (velocidades binarias alcanzables).

- Pérdidas por distancia (cobertura)

  La separación entre emisor y receptor hace que la potencia de la señal en el receptor sea inversamente proporcional al cuadrado de la distancia entre ambos. El término cuadrático es debido a que la potencia emitida en el espacio libre por el emisor se reparte sobre una superficie esférica cuya área es proporcional al cuadrado del radio de la misma. El límite impuesto a la comunicación por esta pérdida determina el área de cobertura del emisor en las comunicaciones terrestres, medida sobre la superficie del suelo sobre el que están situados los potenciales receptores, que normalmente se define por el radio máximo permitido ("radio de cobertura" o "alcance máximo" desde el emisor). En las redes móviles suelen utilizarse los términos de "Picocelda" para coberturas del orden de 50m de radio, y "Microcelda" y "Nanocelda" para distancias respectivamente mayores y menores de aquélla. En ubicaciones de interior (*indoor*) de edificios es donde suelen utilizarse las nanoceldas, mientras que en espacios exteriores (*outdoor*) tienen mayor interés las microceldas y macroceldas (radio de 50 km). Las picoceldas suelen emplearse en ambos tipos de ubicación dependiendo de la densidad de terminales móviles por unidad de superficie (m^2).

- Pérdidas por configuración del terreno (obstáculos)

  Tanto los obstáculos naturales (terreno, arbolado, climatología, etc.) como los artificiales (edificios, obras civiles, camiones, etc.) de grandes dimensiones que impiden la visión directa (LOS*, Line-of-Sight*) entre emisor y receptor, dan lugar a pérdidas de potencia en el receptor denominadas "desvanecimiento lento". En el caso de las comunicaciones móviles la duración real de este desvanecimiento lento está sujeta a la velocidad de desplazamiento del terminal móvil. En las ondas milimétricas, es decir, con señales de microondas en bandas por encima de los 10 GHz, la comunicación ha de establecerse únicamente por visión directa.

- Pérdidas por multicamino (*multipath*)

  Aún en el caso de que no haya visión directa entre emisor y receptor (NLOS, *Non Line-of-Sight*), la señal en el receptor es la suma de muchas señales procedentes del emisor como consecuencia de rebotes (reflexiones) producidos sobre el entorno y sus obstáculos. Las pérdidas son debidas a que dichas señales presentan diferencias de fase entre

ellas y con el rayo directo (si es el caso), y como consecuencia la suma de señales en el receptor es destructiva por la reducida diferencia entre los caminos recorridos (multitrayecto).

Cuando la señal recibida sufre atenuaciones diferentes a frecuencias distintas, los ecos introducen una dispersión temporal sobre la señal transmitida denominada "ampliación del retardo" (*delay spread*). El carácter selectivo del canal en frecuencia depende del ancho de banda de la señal transmitida comparado con el inverso del *delay spread*, denominado "ancho de banda de coherencia del canal". Las señales con ancho de banda reducido en comparación con el "ancho de banda de coherencia del canal" presentan una atenuación plana en frecuencia. Si es del mismo orden sufren el efecto de interferencia intersimbólica (superposición de bits en el receptor) y precisan equipos de ecualización. En las transmisiones por espectro ensanchado (*spread spectrum*) las señales presentan un ancho de banda superior al "ancho de banda de coherencia del canal", y el receptor es capaz de distinguir los sucesivos ecos de señal recibidos y puede utilizarlos para combinarlos coherentemente antes de la decisión de la señal recibida.

Cuando el receptor es móvil se produce una variación en el tiempo de propagación del canal de transmisión denominada "ampliación Doppler" (*Doppler spread*), como consecuencia del desplazamiento del terminal durante la comunicación. El efecto resultante es la aparición de desvanecimientos selectivos en el tiempo que distorsionan la señal recibida. La dispersión frecuencial consiste en la alteración del ancho de banda de la señal transmitida que produce ampliaciones o reducciones del mismo.

La duración temporal de la transmisión afecta también a la distorsión de la señal recibida. Si la duración es corta el canal es invariante en el tiempo, pero a medida que aumenta la duración de la señal el canal tiene más tiempo para variar y distorsionar la señal transmitida, de tal forma que como en el caso del *delay spread* la distorsión temporal está relacionada con el ancho de banda de la señal transmitida. Se define un "tiempo de coherencia" del canal, durante el cual su comportamiento es invariante, y se cuantifica como el inverso de la "máxima frecuencia Doppler" que depende de la velocidad de desplazamiento del terminal móvil cuando éste se mueve en la misma dirección que la de propagación de la señal transmitida entre emisor y receptor. La cuantificación del "tiempo de coherencia" del canal es muy útil para dimensionar la profundidad del entrelazado cuya aplicación elimina los errores a ráfagas que se producen durante un desvanecimiento profundo.

En los sistemas de telefonía celular se utiliza un ancho de banda muy superior a la "máxima frecuencia Doppler" para no alterar el ancho de banda de la señal recibida en el terminal. No obstante existe otro efecto debido a la velocidad excesiva del móvil que consiste en los desplazamientos espectrales que conlleva (efecto Doppler), y que pueden bloquear el seguimiento frecuencial del canal utilizado en el enlace por parte del terminal receptor. Este desplazamiento en frecuencia (Fd) depende de la velocidad del móvil (V) y de la longitud de onda (λ) de la portadora (Fp) según la ecuación:

$$Fd = V/\lambda = V.Fp/C \tag{7.1}$$

siendo C la velocidad de la luz en el vacío (300.000 km/s).

7.3.2. Reducción de los efectos de la propagación

El avance experimentado en las comunicaciones inalámbricas y móviles, permite el diseño de equipos que incorporan contramedidas para reducir los efectos mencionados. A continuación se exponen algunos conceptos en los que se basan dichas contramedidas, y que se han incorporado a las normas internacionales de redes inalámbricas y móviles (Fig. 7.1).

- Codificación de fuente

  Procedimientos de compresión de datos para reducir el flujo binario tanto en almacenamiento como en transmisión. Por ejemplo, las normas de compresión de ficheros de imágenes (JPEG), vídeo (MPEG) y sonidos (MP3) tratan de reducir su tamaño sin pérdida de información, consiguiendo factores de compresión entre 20 y 50 (por ejemplo, reducción de velocidad de 166 Mbps a 3 y 8 Mbps).

- Codificación de canal

  Procedimientos de protección frente a errores de transmisión en el canal, aumentando la redundancia de la información, es decir, al contrario que en los códigos de fuente, se añaden más bits de los necesarios para poder corregir los errores en el receptor sin necesidad de retransmisiones, que aumenten el retardo. Estos códigos se denominan FEC (*Forward Error Control*) o de "control de errores hacia adelante" para distinguirlos de las retransmisiones (ARQ, *Automatic Repeat reQuest*). Algunos de los códigos FEC más utilizados son: Viterbi (convolucional), Fano, Reed-Solomon, Walsh, Secuencial, Turbo

Códigos, etc.). Se suele cuantificar la "tasa de codificación de canal" mediante el cociente 1/k, siendo k el número bits de información útil por bit agregado de código (redundante) para FEC (tasas típicas son: 1/2, 1/3, etc.).

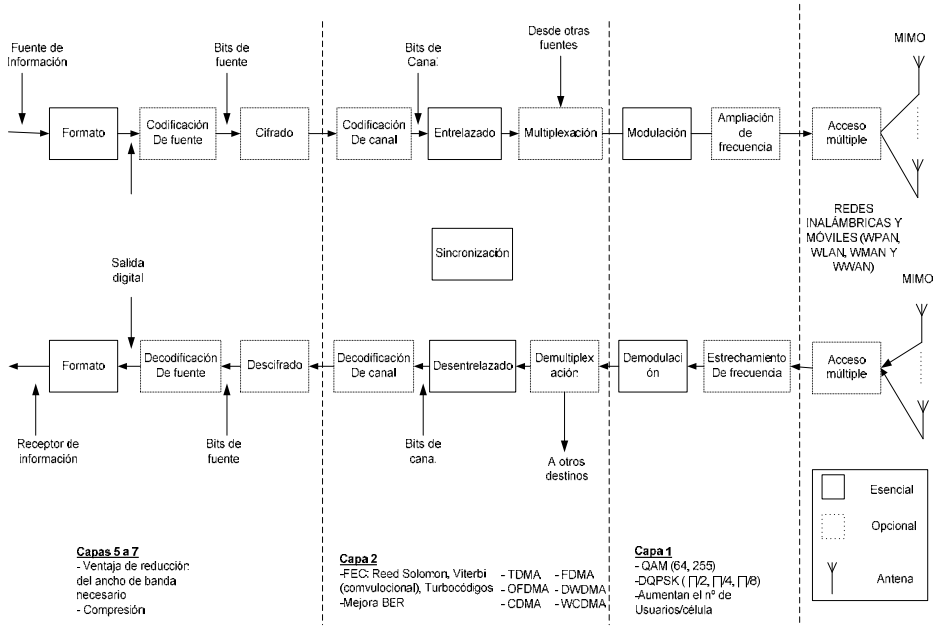


FIGURA 7.1. Contramedidas en las capas del modelo OSI para redes inalámbricas y móviles (emisores y receptores).

- Entrelazado

Procedimiento de reordenación de la información antes de ser transmitida por el canal de radio. Se basa en que para optimizar la codificación y decodificación de canal los errores introducidos han de ser aleatorios, es decir, no deben guardar relación entre ellos como si el canal no tuviese "memoria". Si el canal introduce una ráfaga de errores, por ejemplo, por desvanecimientos en comunicaciones móviles, al realizar la reordenación inversa en recepción los errores quedan separados y con ello se consigue transformar una ráfaga de errores en el mismo número de errores pero situados de forma aleatoria, eliminado la "memoria" del canal. Su inconveniente es el retardo que introduce la reordenación de la información en emisión y recepción, y por ello debe

ser acotado para asegurar la calidad de la las comunicaciones interactivas y de tiempo real. Un valor típico para el tiempo de entrelazado en transmisión de voz es 20 ms.

- Diversidad

  Métodos de compensación de las constantes variaciones de la señal recibida (desvanecimientos lentos y rápidos) por combinación de las distintas señales recibidas para recuperar una única señal menos sujeta a variaciones, en el sentido de mantener o aumentar su relación S/N en el receptor y mantener o incrementar así su velocidad binaria (bps) según la ecuación de Shannon (apartado 3.3). Existen diferentes técnicas de diversidad para conseguir varias señales en un mismo terminal: espacial (dos o más antenas separadas), temporal (la misma transmisión en tiempos separados), polarización (misma señal radiada con distinta polarización), angular (diagramas de radiación o haces direccionales distintos) y frecuencial (emisión en diferentes bandas).

  De todas las técnicas mencionadas la más utilizada es la diversidad espacial. En las comunicaciones móviles no hay dificultad en situar dos antenas suficientemente separadas en la estación base, mientras que las pequeñas dimensiones del terminal móvil hacen más difícil esta solución. Las técnicas restantes presentan mayores problemas en la práctica, tanto por la necesidad de emplear recursos adicionales como por la mayor complejidad de realización. En un terminal móvil la incorrelación de caminos se considera aceptable con separaciones del orden de $\lambda/2$, lo cual puede ser factible con portadoras de frecuencias superiores a 900 MHz. En cuanto a la estación de base se impone una separación dada por la mayor de las dos relaciones:

  - Debe ser superior a 10 el cociente entre la altura de las antenas sobre el suelo y su separación.

  - La separación debe ser mayor que $10.\lambda$ (de la portadora).

- Control de potencia de emisión

  Procedimiento de regulación de potencia en los extremos de transmisión para garantizar la calidad de la comunicación en recepción. Se basa en la reducción del efecto “fiesta de cóctel” (*cocktail party*) que se produce dentro de una zona de cobertura con varios terminales transmitiendo simultáneamente. Debido a este efecto, si una comunicación detecta en

recepción una relación S/N por debajo del umbral incrementa la potencia de emisión para mejorarla, lo que a su vez aumenta la interferencia con el resto de las comunicaciones presentes que la perciben como una disminución en su relación S/N. Si en alguna de ellas la nueva relación S/N queda por debajo del umbral, esta comunicación actuará del mismo modo, pudiendo alcanzarse una situación de aumento paulatino del nivel de interferencia global que lleve a la saturación del mecanismo de control de potencia de todas las comunicaciones presentes.

En las comunicaciones móviles existe otro efecto denominado "cerca-lejos" (*near-far*) relacionado también con el control de potencia de emisión. Es debido a las diferentes ubicaciones de los terminales móviles que dialogan con la estación base, de tal forma que en el enlace ascendente (móvil a estación base) la señal recibida en la estación base será más débil cuanto más alejado (*far*) esté el terminal móvil. El resultado es que dicha señal podría ser enmascarada por las señales procedentes de los terminales móviles más próximos a la estación base (*near*), debido a la imperfección de los filtros recortadores de señal de las etapas de potencia de los transmisores móviles, que permite que parte de la señal emitida se introduzca en las bandas adyacentes, así como en otras bandas, como consecuencia de los productos de intermodulación debidos a la no linealidad de las etapas de potencia clase C utilizadas normalmente en los equipos móviles. Si no se utilizase un control de potencia sobre el enlace ascendente, el efecto "cerca-lejos" degradaría de tal forma las comunicaciones, que obligaría a utilizar en una misma celda frecuencias de operación más separadas espectralmente entre sí, con la consiguiente dificultad de planeamiento frecuencial y disminución de capacidad del sistema de comunicaciones móviles (caída de la eficiencia espectral).

En GSM se aplica un control de potencia independiente tanto para el enlace ascendente como el descendente debido a su naturaleza FDD (frecuencias distintas en cada tipo de enlace). En el enlace ascendente el control de potencia es obligatorio pero en el descendente es opcional. Los incrementos o decrementos de potencia se hacen por múltiplos de 2dB, a partir de un nivel mínimo de 20 mW (13 dBm) para todo terminal móvil (apartado 7.3.3). Para el enlace descendente, aunque es opcional, las estaciones base presentan 8 clases distintas según su potencia de emisión (320-2,5 W) definiendo 16 pasos de 2 dB con objeto de limitar su cobertura para evitar solapamientos indeseados

entre celdas. En la práctica el algoritmo de control de potencia y el de decisión del traspaso (*handover*) de llamada están interrelacionados (apartado 7.8).

El control de potencia además de mejorar la calidad de las comunicaciones, disminuye la interferencia total y en consecuencia incrementa la capacidad del sistema. Se consigue así que al minimizar en cada emisor la potencia de transmisión necesaria, se obtenga una reducción colectiva del nivel de interferencia, incrementando la capacidad del sistema al poder añadir más comunicaciones (más terminales por área de cobertura). Cabe señalar también el ahorro de energía que supone la utilización del control de potencia, al prolongar la vida de las baterías en los terminales móviles y portátiles.

7.3.3. Métricas de prestaciones y factor de mérito

Para la evaluación de las tecnologías de acceso inalámbricas y móviles se establece una serie de conceptos de mérito, que pueden ser convenientemente cuantificados en métricas para tomar decisiones respecto a su idoneidad en los casos concretos de aplicación según los requerimientos. A continuación se definen algunas de las métricas más utilizadas y un factor de mérito.

- Densidad espectral de potencia (dBm/MHz)

  Suele especificarse como el cociente "Potencia radiada/Ancho de banda", y representa el reparto de la potencia emitida por todo el ancho de banda del canal. Tiene interés en la evaluación de la interferencia que puede introducirse en otras comunicaciones que comparten el mismo canal (interferencia cocanal). En la norma UWB (apartado 7.6.4) la Comisión Federal de Comunicaciones en USA (FCC) impone un límite de densidad espectral de potencia de 75 nW/MHz entre 3,1 GHz y 10,6 GHz. Para la norma IEEE 802.11a (apartado 7.5.2) su valor máximo es de 16 dBm/MHz en la banda licenciada de 5,725-5,825 GHz. El dBm toma como referencia de potencia 1mW, y se calcula según:

  $$\text{dBm} = 10.\log (P/1\text{mW}) \qquad (7.2)$$

  siendo P la potencia radiada expresada en miliwatios.

- Relación Portadora/Interferencia (CIR)

  La CIR (*Carrier-to-Interference Ratio*) relaciona en dBs (apartado 3.5.1) la potencia de la señal portadora en el canal con la potencia de la señal interferente. Esta métrica suele utilizarse para analizar la coexistencia de diferentes normas inalámbricas y móviles en una misma área de cobertura. Por ejemplo el CIR para la coexistencia de Bluetooth (apartado 7.6.2) con IEEE 802.11b (apartado 7.5.2) debe ser de 10 dB.

- Energía por bit (J, Julios)

  Representa la cantidad de energía (J) que ha de ser consumida para transmitir un bit de información. Suele utilizarse para evaluar tecnologías de radiofrecuencia que emplean baterías muy pequeñas, y por ello es fundamental la restricción del consumo para alargar la vida útil de la red inalámbrica (habitualmente estándares WLAN y WPAN) en un contexto de operación en donde no es posible la recarga. Valores típicos son desde 0,5 μJ hasta 3 μJ.

- Ancho de banda de información (*Information Bandwidth*)

  Indica la cantidad de bits útiles que pueden ser transmitidos cuando se usa 1 Hz del ancho de banda, y se expresa en "bit/Hz".

- Eficiencia espectral (bit/seg/Hz)

  Está relacionada con las técnicas de modulación jerárquica (HM, *Hierarchical Modulation*) también denominadas de modulación adaptativa, cuando la relación S/N decrece en la entrada del demodulador por efectos de propagación y se desea mantener la tasa de error de bit (por ejemplo: BER de 10^{-4}) reduciendo la eficiencia espectral. Así puede pasarse de una modulación de alta eficiencia QAM64 (5 b/s/Hz) a otra de más baja eficiencia como QPSK (1,5 b/s/Hz) manteniendo la misma BER aunque haya disminuido la relación S/N.

- Capacidad espacial (Cs ≡ bps/m²)

  En comunicaciones móviles representa la tasa de datos de celdas agregadas por unidad de superficie. Cuanto más alta, mayor es la posibilidad de aumentar el número de terminales operando

simultáneamente en la misma superficie. También se le suele denominar "intensidad de datos", por comparación con la intensidad de iluminación de una lámpara en lúmenes/m². Algunos fabricantes de equipos para comunicaciones móviles (apartado 7.8) introducen una métrica (bps/Hz/km²) que combina la eficiencia espectral y la capacidad espacial, para evaluar las prestaciones de los diferentes estándares de cuarta generación (4G) en lugares de elevada concentración de usuarios, estimando aumentos desde 1-20 bps/Hz/km² en macroceldas (3 G), hasta 100-500 bps/Hz/km² en picoceldas, alcanzándose valores superiores a 1000 bps/Hz/km² en nanoceldas.

- Potencia consumida por la batería (Pdc ≡ Watios)

  Cuanto menor sea, más autonomía tiene el terminal en horas de funcionamiento.

- Coste (P€ ≡ Euros)

  Las posibilidades de aplicación por los potenciales usuarios aumentan cuanto menor es el coste.

- Tamaño o factor de forma del terminal (V ≡ m³)

  En sistemas portátiles la facilidad de uso es inversa al volumen y peso asociados.

- Factor de mérito (M)

  Algunos de los conceptos anteriores pueden ser considerados como variables que forman parte del factor de mérito (M) de una tecnología inalámbrica concreta, que podría expresarse según la ecuación:

  $$M = Cs / (Pdc.P€.V) \qquad (7.3)$$

- Calidad de servicio (QoS, *Quality of Service*)

  En una red inalámbrica interactiva y en tiempo real representa un conjunto de características cuantitativas y cualitativas necesarias para satisfacer la funcionalidad requerida en una comunicación concreta. Entre estas características están: el retardo (nseg), la variación del retardo (*jitter*), la tasa de error (BER), la pérdida de paquetes, la asignación de ancho de banda, la velocidad binaria (bps), etc. Una

referencia típica para evaluar la QoS es la definición de clases de servicio utilizada en las redes cableadas con transmisión ATM, considerando los siguientes conceptos: temporización entre emisor y receptor, la velocidad (bps) y el modo de conexión (conmutado/datagrama).

- Clase 1: relacionada, constante y conmutado
- Clase 2: relacionada, variable y conmutado
- Clase 3: no relacionada, variable y conmutado
- Clase 4: no relacionada, variable y datagrama

Para comparar los modos conmutado y datagrama de las redes cableadas con sus equivalentes en las redes inalámbricas, puede suponerse que en el modo conmutado siempre existe un "canal de radio" permanente entre emisor y receptor, mientras que en el modo datagrama el "canal de radio" ha de establecerse en cada salto de la ruta inalámbrica, con los retardos, errores y posibles pérdidas de información que tal procedimiento implica.

7.4. DISEÑO DE REDES INALÁMBRICAS Y MÓVILES

7.4.1. Elementos básicos

En general una red de comunicaciones inalámbricas y móviles presenta en su forma más compleja, como es el caso de las redes celulares de telefonía móvil, una estructura compuesta por los siguientes elementos: Estaciones Móviles (EM), Estaciones Base (EB), Estaciones de Control (EC) y Centros de Conmutación (CC). A continuación se describen las funciones típicas de cada uno de ellos (Fig. 7.2).

Las estaciones móviles (EM) son los equipos que tienen la función de suministrar el servicio concreto a los usuarios en el lugar, instante y formato (voz, datos, gráficos, imágenes fijas y móviles, textos, etc.) adecuados. Según la complejidad del servicio cada estación móvil puede actuar en modo emisor, receptor o en ambos modos (transceptor).

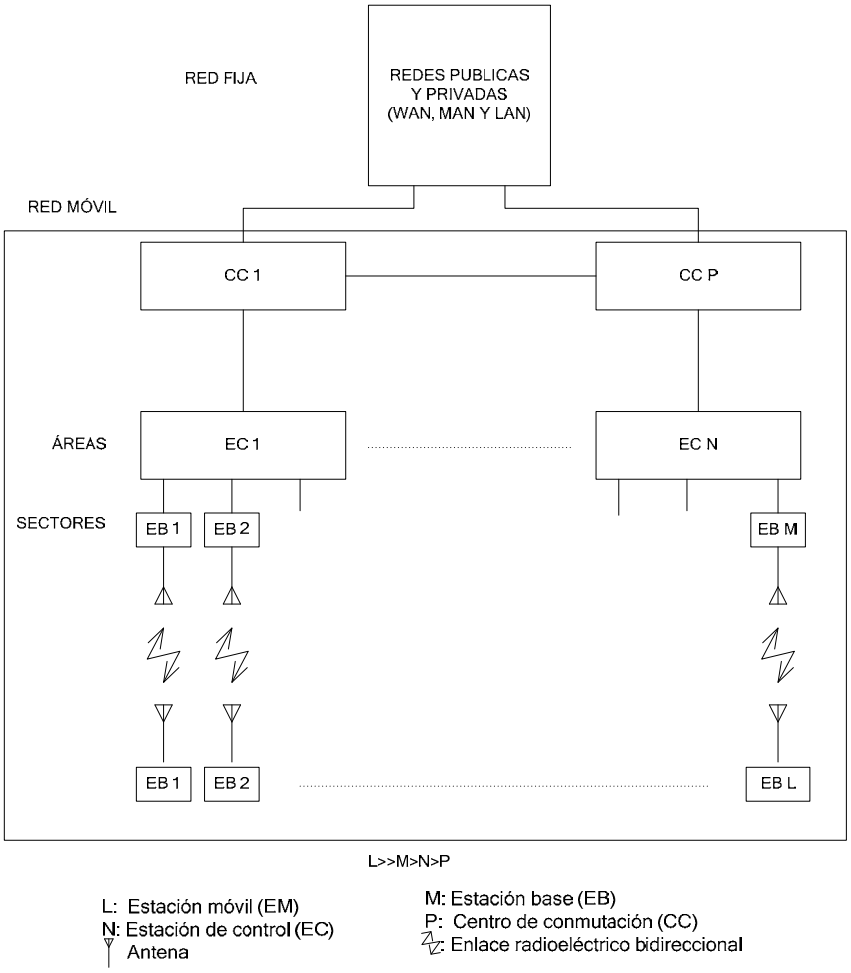


FIGURA 7.2. Estructura genérica de las redes inalámbricas y móviles.

En los dos primeros modos el acceso a la red ó al móvil es unidireccional, mientras que en el tercero es bidireccional o unidireccional, dependiendo de que la estación móvil pueda ser o no localizada por la red. Desde el punto de vista de su realización física lo más simple es que el terminal de usuario y la estación móvil formen una unidad indivisible. En los casos en que el móvil transporta a varios usuarios (autobuses, trenes, barcos, aviones, etc.), los terminales móviles y la estación móvil pueden ser físicamente distintos, así como estar conectados por cable (cabinas) o mediante un enlace radioeléctrico (por ejemplo, equipos telefónicos con manos libres vía Bluetooth).

La reducción de potencia de las estaciones móviles permite disminuir la interferencia entre las estaciones móviles asignadas a canales idénticos, así como el tamaño y peso de los componentes suministradores de energía (baterías), lo que conlleva una mejor calidad del servicio y la mayor comodidad de uso y autonomía de la estación móvil. La complejidad y coste de la estación móvil para un servicio concreto dependerá de las técnicas empleadas en el aprovechamiento del espectro y de reducción de potencia, dentro de un margen de calidad aceptable.

Las estaciones base (EB) se encargan de mantener el enlace radioeléctrico entre la estación móvil y el centro de control del servicio durante la comunicación. Una estación base atiende a una o varias estaciones móviles, y según el número de éstas (L) y el tipo de servicio se dimensiona el número adecuado de estaciones base (M), para proporcionar una cobertura total de servicio en el área geográfica concreta (apartado 7.3.1). Los condicionamientos técnicos del enlace radioeléctrico para dar un buen servicio simultáneo a las estaciones móviles de una estación base, son fundamentalmente la utilización eficiente del espectro y la reducción de potencia de los terminales. Para aprovechar el espectro de frecuencia se emplean técnicas de multiplexación y modulación de señales que proporcionan un número elevado de canales en un ancho de banda reducido. Asimismo, mediante técnicas adecuadas de asignación de canales a estaciones móviles, se consigue reutilizar los mismos canales entre distintas estaciones base.

Las estaciones de control (EC) realizan las funciones de gestión y mantenimiento del servicio. Una tarea específica consiste en la asignación de estaciones base en un sector, dentro del área de cobertura, a las estaciones móviles que se desplazan por el mismo. En el caso más general de un servicio móvil cualquier estación móvil puede desplazarse de un sector a otro dentro del área de cobertura, sin pérdida de la comunicación durante su movimiento independientemente de la ruta seguida. Asimismo cuando una estación móvil es solicitada por otra para establecer una comunicación, aquélla debe ser localizada en el sector (M) en que se encuentre en un instante dado. La función de conmutación entre estaciones base (*handoff* o *handover*), permite cambiar el canal ocupado por la EM en la EB anterior por otro libre en la EB próxima (apartado 7.8.1). La función de localización (*roaming*) de una EM fuera de su sector habitual, implica que cada estación base debe conocer las EM "residentes" y las "visitantes" (en tránsito) para que las EC puedan determinar la posición de una EM. El coste del servicio dependerá del número y complejidad de las funciones de las EC, así como de las áreas geográficas cubiertas por su incidencia en el número de estaciones base (M) y de control (N).

Los centros de conmutación (CC) permiten la conexión entre las redes públicas y privadas (WAN, MAN y LAN) con la red de comunicaciones móviles, así como la interconexión entre estaciones de control localizadas en las distintas áreas geográficas de la red móvil. Según la complejidad del servicio y cobertura de la red, los centros de conmutación pueden formar parte de redes fijas al margen de la red móvil, o bien sus funciones ser asumidas por estaciones de control.

Dentro de cada ámbito geográfico el diseño de una red inalámbrica y móvil consistirá en el despliegue de parte o todos los elementos mencionados en función del número de estaciones móviles previstas (L), las áreas de cobertura solicitadas (N) y los sectores en que se dividen (M), los centros de conmutación necesarios (P), y el tipo de servicio o servicios ofrecidos. En general todo sistema inalámbrico y móvil necesita al menos un enlace radioeléctrico entre las estaciones móviles y la estación o estaciones base.

7.4.2. Planificación y dimensionamiento de una red de radiocomunicación

Si bien los sistemas de comunicación inalámbricos y móviles presentan distintas áreas de cobertura, desde las redes de área personal (WPAN) con alcance máximo de 10 m, pasando por las de área local (WLAN), las celulares (metropolitanas WMAN e internacionales WWAN) hasta las redes de satélites, todas ellas con diversa densidad de terminales por kilómetro cuadrado, distintos márgenes espectrales, diferente movilidad y capacidad de servicio, es ilustrativo señalar las pautas a seguir en su planificación y dimensionamiento. Para ello se elige un caso de los más complejos como es el de una red celular, dado que presenta una amplia variedad de problemas y soluciones en cobertura y movilidad que pueden ser convenientemente extrapolados a otras redes particulares más sencillas.

El diseño de una red celular comprende por un lado aspectos de campo como la búsqueda de los emplazamientos de las estaciones base, la estimación del área de cobertura de cada una, el reparto de frecuencias entre celdas, el ajuste de las potencias de emisión, etc., y por otro, aspectos de dimensionamiento de red fija como el cálculo del número de controladores de estación base, su ubicación, la interconexión de los elementos de red fija, el tamaño de las bases de datos, etc.

La planificación de una red celular es un proceso complejo que comprende diversas disciplinas tales como, planificación frecuencial, cartografía (SIG, Sistemas de Información Geográfica), estadísticas de movilidad de los potenciales usuarios, teoría de teletráfico (número de llamadas en la hora cargada), mercadotecnia (ofertas de servicios en tiempos de baja utilización de la red), etc., y que generalmente involucra a varios departamentos dentro de una organización. Las herramientas de simulación y los laboratorios de evaluación mediante emuladores empleados en los departamentos, representan unos activos muy importantes de la organización que encarecen enormemente la planificación en términos de coste de equipamientos físicos (*hardware*) y lógicos (*software*), y muchas horas de trabajo del personal especializado.

Por todo ello la forma de abordar su resolución necesita un proceso de diseño secuencial, siguiendo unos pasos bien definidos e iterativos, de tal forma que la solución de configuraciones parciales en una etapa se realimenta hacia atrás para su rediseño desde etapas anteriores. La planificación debe tener en cuenta el crecimiento gradual de la demanda de servicio debido tanto al aumento de terminales en las diferentes celdas, como a la oferta de nuevos servicios que eleven el tráfico de los terminales. A continuación se esquematizan las etapas del proceso de diseño:

- Previsión de la demanda y estimación de la situación de los emplazamientos de las estaciones base. Cálculo de las áreas de cobertura por estación base. Estimación del tráfico ofrecido por célula y cálculo del número de portadoras necesarias por sector y estación base para satisfacer un grado de servicio determinado.

- Estimación de la movilidad de los usuarios y cálculo aproximado del tránsito entre células de los potenciales terminales móviles. Uso de patrones de movilidad para la estimación del número de cruces entre células, evaluación del número de traspasos de llamada y del número de actualizaciones de posición.

- Dimensionamiento de las áreas de localización. Estimación inicial de áreas de localización con minimización de la carga de señalización por actualizaciones de posición. Refinamiento de la solución mediante subdivisiones sucesivas de las áreas de localización de la configuración inicial hasta encontrar un subóptimo que minimice el conjunto de señalización por búsquedas y actualizaciones de posición.

- Dimensionamiento de los canales de señalización. Conocido el tráfico ofrecido por celda y el patrón de movilidad de los terminales se puede estimar el número de canales de señalización por estación base en función de la solución subóptima de áreas de localización de la etapa anterior, considerando los procedimientos de señalización más característicos.

- Dimensionamiento en número y ubicación del resto de los equipos de la red celular, además de las estaciones base y estaciones de control, considerando las especificaciones técnicas de capacidad de dichos equipos proporcionadas por los diferentes suministradores y las necesidades de tráfico de la red de estaciones base estimadas en la primera etapa de planificación.

- Diseño de la topología de la red, estudiando las diferentes soluciones de interconexión de estaciones de base con los equipos de la red fija, además de la evaluación de los costes de compra o alquiler de las líneas de interconexión entre equipos, y las alternativas de su reubicación geográfica para optimizar costes.

- Diseño de la red de interconexión con otras redes fijas y móviles, evaluando diferentes soluciones de conmutación siguiendo criterios de minimización de costes y fiabilidad de infraestructuras. Finalmente se dimensiona el número de enlaces necesarios para la interconexión con otras redes de voz y datos para las conexiones con el exterior, ya sea de móvil a fijo o de móvil a móvil de otro operador.

7.4.3. Evolución tecnológica

La experiencia acumulada por los investigadores, fabricantes y operadores de sistemas de radio sobre todo en lo referente a las tres generaciones anteriores de telefonía móvil, que ha fundamentado el próximo reto de alcanzar en la siguiente generación (4G) las mismas prestaciones al usuario móvil que las redes cableadas en términos de, velocidad (bps), acceso a redes IP, y calidad de servicio (QoS), se ha extendido a otros ámbitos de aplicación de tal manera que surgen continuamente nuevas necesidades de mejora en todo tipo de redes inalámbricas y móviles. Estos nuevos requerimientos en cobertura universal, bajo coste, bajo consumo, mínima infraestructura, conexión ubicua independiente de la tecnología, aplicaciones ilimitadas en todos los sectores (industria, ocio, investigación, servicios, sanidad, hogar, transporte, etc.), entre

otros muchos, someten a las normas de redes inalámbricas y móviles actuales a la presión continua de nuevas revisiones con perfiles mejorados, o a la aparición de nuevos estándares. A continuación se mencionan algunas de las tendencias innovadoras más destacadas.

7.4.3.1. Sistemas MIMO

Los sistemas MIMO (*Multiple Input-Multiple Output*) constituyen una alternativa de diversidad que utiliza múltiples antenas (más de dos por extremo) tanto para recibir como para transmitir, permitiendo que una transmisión de datos a velocidad elevada pueda dividirse en múltiples tramas que se modulan y transmiten por antenas diferentes en un instante dado, utilizando la misma frecuencia de canal que el resto de las antenas. Las reflexiones por multicamino (apartado 7.3.1) convierten a la señal de salida de cada antena receptora en una combinación lineal de múltiples tramas de datos procedentes de todas y cada una de las antenas emisoras.

Las tramas de datos se separan en el receptor mediante algoritmos que se basan en estimaciones de todos los canales entre el emisor y el receptor, lo que presenta las ventajas de multiplicar la velocidad de transmisión por tener más antenas y de aumentar el alcance (cobertura) al disponer de antenas en diversidad.

Los estudios teóricos acreditan la superación del límite representado por la ecuación de Shannon en velocidad binaria (bps), ya que al elevar el número de antenas y manteniendo constante el ancho de banda se aumenta la relación S/N en el receptor, lo que supone un incremento de la eficiencia espectral (apartado 7.3.3) favorable a estos sistemas. Se requiere para ello la igualdad en el número de antenas en ambos extremos de transmisión (emisor y receptor), por lo que en caso de desigualdad la mejora estará limitada por el número de antenas del extremo que menos tenga.

Se considera que las previsiones de la UIT (Unión Internacional de Telecomunicaciones) para las redes móviles de la cuarta generación (4G), de alcanzar 1 Gbps sin movimiento y 100 Mbps con desplazamiento de 150 km/h serán posibles con tres técnicas: MIMO, modulación y codificación adaptativas (en función de las condiciones del canal), y la codificación avanzada para el control de errores.

7.4.3.2. Sistemas de radio definido por programación

El término SDR (*Software Defined Radio*) describe los sistemas de comunicación por radio que proporcionan un control programado de una variedad de técnicas de modulación, operaciones en banda ancha o estrecha, contramedidas para asegurar las comunicaciones (control de potencia, entrelazado, codificación de errores, etc.), seguridad de la información y requerimientos de formas de onda, tanto para estándares actuales como futuros en un rango de frecuencias elevado. Esta tecnología es útil en todas las áreas de las comunicaciones inalámbricas y móviles, permitiendo soluciones eficientes y de bajo coste ante las diferentes limitaciones de los sistemas actuales, dado que pueden ser reprogramados para conseguir un mejor rendimiento, un mayor conjunto de características, servicios nuevos que proporcionen alternativas al usuario final y crecientes oportunidades de negocio a los proveedores de servicios.

La arquitectura de un sistema SDR parte de un modelo genérico de alto nivel con bloques de funcionalidades específicas conectados mediante recomendaciones de estándares abiertos. Dichos bloques son los siguientes: procesamiento para el medio (conversión de frecuencias, modulación/demodulación, etc.), seguridad de la información (privacidad, autenticación, cifrado, etc.) y procesamiento de la información (encaminamiento, multiplexación, codificación, señalización, etc.).

7.5. REDES DE ÁREA LOCAL MÓVILES (WLAN)

7.5.1. Introducción

Las alternativas de redes de acceso inalámbrico y móviles se dividen según su cobertura en redes locales (WLAN), metropolitanas (WMAN) y de área extensa (WWAN). Todas ellas están destinadas a proporcionar servicios semejantes a las redes cableadas de cobertura equivalente utilizando el espectro radioeléctrico, lo que permite la movilidad del usuario cuando utiliza terminales portátiles y móviles. Las redes de mayor cobertura (WMAN y WWAN) están soportadas por radioenlaces (punto a punto), que presentan enlaces punto-multipunto en los extremos para el acceso de los usuarios provistos de terminales fijos, portátiles y móviles, y por estar operadas por empresas de comunicaciones son de tipo público.

Bajo la denominación de WLAN (*Wireless Local Area Network*) se agrupan las redes de acceso inalámbrico con una cobertura inferior a la de las LAN (10 m a 300 m) y sus realizaciones prácticas presentan las siguientes características:

- Medio físico: márgenes del espectro electromagnético utilizados en el espacio libre
  - Radiofrecuencia (RF) y microondas: 300 MHz a 300 GHz
  - Infrarrojos (IR): 800 nm a 900 nm
- Topología
  - Permanente: con puntos de acceso a otras LAN
  - Provisional (*Ad hoc*): sin puntos de acceso a otras LAN
    - Semipermanente (meses)
    - Temporal (días u horas)
- Encaminamiento
  - RF y microondas (difusión)
  - Infrarrojo (IR)
    - IR directo (DBIR)
    - IR difuso (DFIR)
    - Totalmente difuso
    - Parcialmente difuso (QD)
- Multiplexación
  - Por división de código (CDM)
  - Por división de frecuencia (FDM)
- Método de acceso al medio
  - Probabilista
  - Determinista por asignación fija TDM

Si bien el contenido de estas características es suficientemente explícito para una información básica sobre los aspectos específicos de las WLAN, a continuación se detallan algunas de ellas:

- Desde el punto de vista de los márgenes del espectro electromagnético utilizados en el medio físico (espacio libre), cabe destacar el amplio uso de las bandas ISM por las WLAN: 902 MHz a 928 MHz (UHF), 2,4 GHz a 2,483 GHz (UHF) que es el más utilizado, y 5,725 GHz a 5,875 GHz (SHF) en la banda de las microondas. El uso de los infrarrojos es una característica particular de las WLAN (longitudes de onda desde 760 nm a 1 mm).

- En la topología de las WLAN se presentan realizaciones de tipo permanente (con infraestructura), y de tipo provisional (sin infraestructura). Las realizaciones de tipo "permanente", como su nombre indica tienen carácter indefinido en el tiempo y están destinadas a instalaciones en donde no se pueden tender cables, como en los edificios históricos o en entornos peligrosos, y en aquellas en las que existe mucho personal en movimiento permanente como: aeropuertos, hospitales, grandes almacenes, instalaciones deportivas, recintos feriales, plantas de fabricación, plantas químicas, etc. (*Hot Spots*). Las instalaciones permanentes de WLAN suelen tener "puntos de acceso" a las redes LAN cableadas, que representan su infraestructura, permitiendo de 3 a 15 puntos de acceso, con una cobertura de 30 m a 300 m por punto de acceso. Las realizaciones de tipo provisional (*Ad hoc*) según su duración pueden ser de tipo "semipermanente" (meses), para empresas que se desplazan con frecuencia o donde no sea posible instalar puntos de acceso (edificios singulares, ambientes agresivos, obras civiles, construcción, etc.), o de tipo "temporal" (días u horas) para eventos especiales como: ferias itinerantes, congresos, grupos de trabajo, catástrofes, etc. (Fig. 7.3).

- El encaminamiento en las WLAN es del tipo de "difusión" (punto-multipunto) para radiaciones electromagnéticas en RF y microondas. Con radiaciones infrarrojas (IR) la difusión presenta diferentes opciones según la alineación que puede existir entre los diferentes terminales de usuario. En la opción "IR directo" (DBIR, *Directed Beam Infrared*) los terminales emisor y receptor están alineados para su visión directa (LOS, *Line of Sight*), con un ángulo de apertura del haz de la radiación infrarroja muy estrecho, lo que permite hacer conexiones como en una topología "punto a punto" tales como: entre dos terminales, con un

punto de acceso a una LAN, o entre varios terminales y un concentrador óptico de una LAN. Las ventajas de DBIR son: alta velocidad (10 Mbps), mayor alcance, menor potencia óptica, menor riesgo para los ojos, ausencia de interferencias por multitrayecto, por no haber señales reflejadas, y mejor comunicación bidireccional. Entre sus desventajas cabe destacar: necesidad de alineación de los terminales, inmovilidad de los mismos e interrupción de la transmisión por obstáculos.

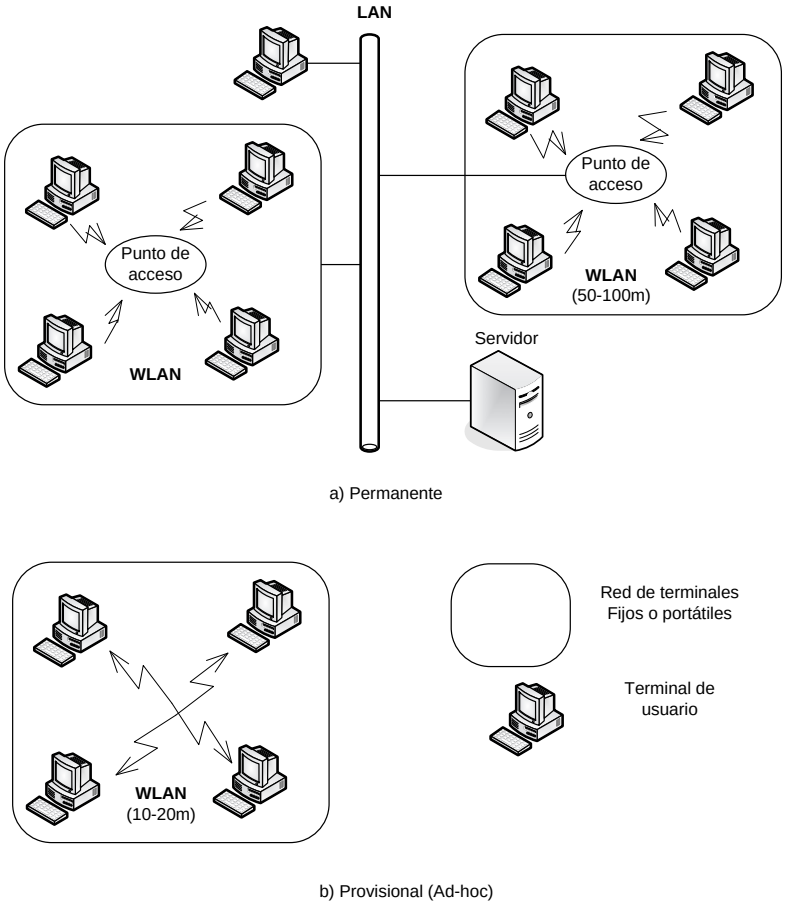


FIGURA 7.3. Topologías de una red de área local inalámbrica (WLAN).

En "IR difuso" (DFIR, *Diffused Infrared*) existen dos opciones de encaminamiento para WLAN: totalmente difuso (llamado "difuso") y parcialmente difuso (QD, *Quasi diffuse*) también denominado "dirigido". En el

primer caso (difuso) cada terminal apunta hacia el techo y las paredes con un gran ángulo de apertura (divergente) sin visión directa (NLOS, *Non LOS*), para obtener múltiples reflexiones con objeto de emitir su señal hacia los demás terminales y recibir las de éstos, lo cual tiene el inconveniente del multitrayecto que limita la velocidad de transmisión. En el segundo caso (QD) tampoco hay visión directa (NLOS), pero los haces de los terminales tienen un menor ángulo de apertura (menor divergencia) y apuntan hacia un reflector activo o pasivo situado en un punto equidistante (por ejemplo, el techo), con objeto de igualar las trayectorias de las radiaciones infrarrojas reflejadas y reducir el efecto de variación de longitud (multitrayecto), lo que mejora la velocidad. Las desventajas de DFIR (difuso y dirigido) son: mayor potencia óptica, menor alcance, velocidad limitada por multitrayecto (1 Mbps a 4 Mbps), mayor riesgo para los ojos, y que en comunicación simultánea las reflexiones de la señal saliente son más fuertes que la señal entrante. Entre sus ventajas están: ausencia de alineación de los terminales, mayor movilidad, y no interrupción de la transmisión por obstáculos. Las WLAN basadas en IR difuso, como las de IR directo, también pueden conectarse a puntos de acceso de redes LAN.

Aparte de las limitaciones mencionadas en velocidad, cobertura y movilidad que experimentan las WLAN basadas en IR por su configuración, sobre todo IR difuso debido al multitrayecto, existen otras causas de limitación debidas a fuentes de luz interferentes y a los estados transitorios de los dispositivos ópticos. Entre las fuentes de luz interferentes están las artificiales como las lámparas fluorescentes e incandescentes, y las naturales como la luz diurna (*skylight*) y la luz del sol (*sunlight*). En este sentido se desaconseja en exteriores la alineación este-oeste para enlaces de IR directo (*sunlight*).

Los dispositivos ópticos empleados como emisores son los diodos electroluminiscentes (LED) y los diodos láser (LD). Los diodos LED se consideran más seguros para los ojos y son más económicos, aunque su eficacia de conversión fotón/electrón (10% a 20%) y ancho de banda (decenas de kHz) son menores. Por el contrario los diodos LD son más caros y peligrosos para la vista (por debajo de los 1.400 nm), pero su eficacia de conversión (30% a 70%) y ancho de banda (decenas de kHz a decenas de GHz) son mayores. Los dispositivos ópticos receptores son los fotodiodos PIN y de avalancha (APD). Los fotodiodos PIN de Silicio son los más utilizados en las aplicaciones comerciales frente a los de Germanio y Arseniuro de Galio (AsGa), y los APD son más caros.

- Los tipos de multiplexación FDM y CDM ya han sido mencionados en el apartado 3.7. No obstante por su interés en el desarrollo de la tercera

generación (3G) de redes digitales inalámbricas (apartado 7.8.2) y sus correspondientes equipos portátiles de bajo consumo, peso y dimensiones, se describen someramente las técnicas de espectro ampliado. Las técnicas de comunicación basadas en espectro ampliado (SS *Spread Spectrum*) fueron desarrolladas en la segunda guerra mundial, para contrarrestar los efectos de captación e interferencia de la información producidos sobre las emisiones electromagnéticas en las zonas de combate. Su fundamento consiste en combinar la señal de información con una señal auxiliar de modulación, que extiende (*spread*) la energía de la señal transmitida mediante un código secreto, sobre un ancho de banda (*spectrum*) mayor que el correspondiente a la señal de información. El receptor de espectro ampliado realiza la operación inversa volviendo a recomponer la señal de información, con el código secreto del emisor, en su ancho de banda original. El desconocimiento del código de recuperación de la señal imposibilita su detección e intercepción a un escucha no autorizado. En efecto, las acciones de interferencia se hacen difíciles porque el receptor disminuye la potencia espectral de la señal interferente, al expandirla por todo el ancho de banda de transmisión y extraer la señal de información en un ancho de banda reducido, lo que atenúa su impacto en las prestaciones del receptor. Existen dos procedimientos de modulación en espectro ampliado, denominados de "salto de frecuencia" (FHSS, *Frequency Hopping SS*) y de "secuencia directa" (DSSS, *Direct Sequence SS*).

En FHSS el emisor pasa de un canal a otro (salto) en el tiempo, dentro del ancho de banda de transmisión, de acuerdo con un código de secuencia, enviando varios bits de información a cada canal. En sistemas militares un mismo bit puede ser enviado a varios canales en saltos muy rápidos, lo que proporciona una gran robustez a la comunicación pero encarece y complica enormemente los equipos. Un conjunto de códigos de secuencia se considera "ortogonal" (no interferentes entre sí para aplicaciones de acceso múltiple) si nunca se usa la misma frecuencia al mismo tiempo. La FCC de USA en sus bandas ISM (*Industrial, Scientific and Medical*) impone un mínimo de 50 canales de salto en la banda de 902 MHz y 75 en la de 2,4 GHz. El tiempo de permanencia en un canal de la banda de 902 MHz no debe superar los 400 ms cada 20 s, mientras que en la banda de 2,4 GHz ese mismo tiempo debe producirse cada 30 s. Sobre este método tuvieron una patente en 1940 la actriz de cine Hedy Lamarr y el compositor de música George Antheil.

En DSSS el emisor reemplaza cada bit de información por una secuencia de bits (denominada "chip") que ocupa el mismo intervalo de tiempo.

El factor por el cual el ancho de banda de la señal de información es ampliado en la transmisión se llama "ganancia de procesamiento" (*processing gain*), y algunos autores tienden a referirse indistintamente a la "longitud de la secuencia de bits" como a la "ganancia de procesamiento", identificando a ambas con el "número de bits por chip". Como ejemplos de valores de la "ganancia de procesamiento", el FCC en las bandas ISM impone un mínimo de 10 bits/chip, algunos sistemas comerciales utilizan 20 bits/chip, mientras que en sistemas militares este número puede alcanzar valores de potencias enteras de 10. Cuanto más elevado es este parámetro mayor es la inmunidad del sistema contra interferencias, y más grande la complejidad y coste del equipamiento. Como resumen de ambas técnicas de modulación puede decirse que FHSS experimenta interferencia sólo una fracción del tiempo (útil en entornos densos con muchas células adyacentes), mientras que DSSS sufre una fracción de la interferencia todo el tiempo (útil si la interferencia es baja).

La técnica de acceso múltiple por división de código (CDMA) basada en FHSS y DSSS ha introducido en los sistemas inalámbricos digitales las ventajas de: mayor tolerancia a la interferencia (*jamming*), reducción del efecto pernicioso de multitrayectoria (*multipath*), y la capacidad de compartir un ancho de banda común con otras señales, lo que ha multiplicado la densidad de usuarios celulares (capacidad) en los sistemas inalámbricos analógicos de primera generación por factores de 3 a 10. La compartición de ancho de banda se consigue asignando a cada usuario de una célula un código de modulación distinto dentro de un conjunto ortogonal, de ahí el nombre de división de código para esta técnica de multiplexación.

Un hito histórico en el despliegue comercial de la WLAN se produjo en 1985 cuando el FCC liberó las bandas ISM (902-928 MHz, 2400-2483,5 MHz y 5,725-5,85 GHz) para su utilización general sin licencia (además de USA también Canadá y Méjico), con limitación de potencia a 1 W y técnicas de modulación de espectro ampliado, con objeto de que pudieran coexistir productos de distintos fabricantes. Esta liberación fue seguida por un rápido desarrollo de productos comerciales que utilizan estas bandas, pero incompatibles entre sí por la ausencia de normas comunes. Fuera de USA el ejemplo liberalizador del espectro se concretó en 1995 asignándose las bandas 2,445-2,475 GHz (UK), 2,4-2,5 GHz (UE por el CEPT, y en Japón), también para su uso con técnicas de modulación de espectro ampliado.

Otro margen de frecuencias no regulado, además de las bandas ISM mencionadas dentro del espectro de radiofrecuencia (RF) y en concreto en el margen de las microondas (3 GHz a 300 GHz, o longitudes de onda entre 10 cm

y 1 mm), es el de las radiaciones infrarrojas (IR, *InfraRed*) con valores de longitud de onda entre los 800 nm y 900 nm (infrarrojo próximo). Debido a su característica de no necesitar licencia, una de las primeras realizaciones inalámbricas (1979) en comunicaciones industriales, destinada a plantas de fabricación (*wireless indoor network*), se desarrolló en un laboratorio suizo de IBM (Ruschlikon, Zurich) con tecnología IR difuso (125 kbps).

Los primeros productos WLAN de la década de los 80 en el siglo XX se caracterizaron por su falta de seguridad, lentitud, alto coste y ausencia de normalización, frente a las redes LAN existentes. Como elementos de comparación se pueden citar, la velocidad (1 Mbps frente a 10 Mbps de Ethernet), la tasa de error de bit (BER) de 1 en 1000 (10^{-3}) frente a 1 en 1000 millones (10^{-9}), y el número de terminales (decenas frente a centenares). Para superar estas limitaciones en el siguiente decenio se iniciaron tareas de normalización, que incorporan los avances introducidos en técnicas de modulación, que aumentan la eficiencia espectral (velocidad) y la integridad de la señal (BER digital y relación S/N analógica), así como los avances experimentados en métodos eficientes de acceso al medio. Los resultados más representativos en el desarrollo de normas para WLAN se enumeran a continuación: IEEE 802.11, HIPERLAN, FSO, HomeRF y OpenAir. Seguidamente se detallan algunas de dichas normas.

7.5.2. La Norma IEEE 802.11

Una de las primeras normas existentes para WLAN es la IEEE 802.11, que iniciada en 1990 fue finalmente publicada en 1997. En esta norma, al igual que en las LAN cableadas se definen los dos primeros niveles del modelo OSI, y presenta como características principales: el uso de las bandas IR y RF (ISM), las dos modulaciones SS en RF, la técnica de acceso CSMA/CA, la gestión de potencia, la itinerancia de terminales (*roaming*) y los mecanismos de seguridad. En IR utiliza la técnica de modulación OOK (*On Off Keying*), en la que un "1" lógico está representado por la ausencia de la señal, mientras que su presencia es un "0" lógico. La configuración definida es IR difuso para permitir la conectividad de terminales fijos, portátiles y móviles, el ancho de banda se sitúa en el infrarrojo próximo (800 nm a 900 nm) con un máximo de 2 W de potencia.

En RF utiliza la banda de ISM de 2400 MHz a 2483,5 MHz con objeto de hacerla compatible con las bandas homólogas no sujetas a regulación (liberadas) definidas internacionalmente (USA, Canadá, Méjico, Japón, UE y UK). La potencia transmitida por cada terminal se restringe a 100 mW, para cumplir con

la limitación impuesta en ISM por el FCC (máximo 1 W). Las modulaciones de espectro ampliado (SS) son las ya mencionadas a las que se añaden otras modulaciones para aumentar la eficiencia espectral y el BER: DSSS con BPSK (*Binary Phase Shift Keying*), QPSK (*Quadrature Phase Shift Keying*), y FHSS con GFSK (*Gaussian Frequency Shift Keying*). Con objeto de mejorar la fiabilidad de la transmisión en condiciones de severa interferencia y atenuación por multitrayecto (apartado 7.3.1), en DSSS se definen 5 subbandas de 26 MHz cada una (2412-2442-2470-2427-2457 MHz) cuyas dos últimas están solapadas con las anteriores, lo que permite a un terminal disponer de 5 órdenes de selectividad de frecuencia. En DSSS se utiliza una señal de expansión predefinida (a diferencia de CDMA que tiene varias), con una ganancia de procesamiento de 11 bits/chip para superar el mínimo impuesto por el FCC en IMS (10 bits/chip). En FHSS se definen 79 canales de salto con 1 MHz de ancho de banda, y 3 conjuntos de 22 secuencias de salto (patrones) con la máxima ortogonalidad para permitir la coexistencia con otras redes. Los tiempos de permanencia y frecuencia de salto para cada canal son los especificados por el FCC para la banda ISM utilizada (400 ms y 30 s respectivamente en 2,4 GHz).

Una de las características principales de la norma IEEE 802.11 en el nivel de enlace es el método de acceso CSMA/CA (*Carrier Sense Multiple Access/Collision Avoidance*), denominado de acceso múltiple por detección de portadora y prevención de colisión, que al igual que Ethernet obliga a que todos los terminales escuchen el medio antes de transmitir, y no lo ocupen hasta que esté libre, pero introduce un protocolo que minimiza la posibilidad de colisiones una vez ocupado. Bajo este protocolo (*Collision Avoidance*), un terminal antes de transmitir la información al medio (mensaje) envía una corta trama de petición RTS (*Request-To-Send*) que contiene la dirección de destino (terminal receptor) y la longitud del mensaje deseado NAV (*Network Allocation Vector*), y espera la trama de disponibilidad del terminal receptor CTS (*Clear-To-Send*) con el mismo contenido de dirección y NAV de la trama RTS enviada. Una vez recibida la trama CTS el terminal emisor transmite el mensaje y espera la trama de reconocimiento ACK (*ACKnowledge*) del nodo receptor, que le indicará la correcta recepción del mensaje, y para el resto de los terminales inalámbricos supondrá el fin de la transmisión. Este protocolo de prevención de colisiones basado en la secuencia RTS-CTS-ACK evita el problema del "terminal oculto" (causa de la aparición de las colisiones en las redes Ethernet), ya que todos los terminales que desean transmitir y detectan una trama RTS esperan por la trama CTS, y al repetirse el contenido, otros posibles terminales fuera del alcance del terminal emisor (terminales ocultos), pero dentro de la cobertura del terminal receptor, también son avisados.

En el esquema de la Figura 7.4 cuando el terminal A va a transmitir hacia el B envía previamente la trama RTS, que incluye: la dirección del receptor, la duración de la transmisión y el reconocimiento correspondiente. El terminal B recibe dicha trama, al igual que los terminales D y E, y contesta con una trama CTS que contiene su dirección y la misma duración de la transmisión. Esta respuesta de B es recibida por D, E y también por los terminales C, F y G, "terminales ocultos" por no estar en el área de cobertura de A, los cuales, advertidos de la posible colisión por la trama CTS, retienen sus datos el adecuado período tiempo para impedir la colisión. Los puntos de acceso (P) también participan en la secuencia RTS-CTS-ACK para evitar colisiones.

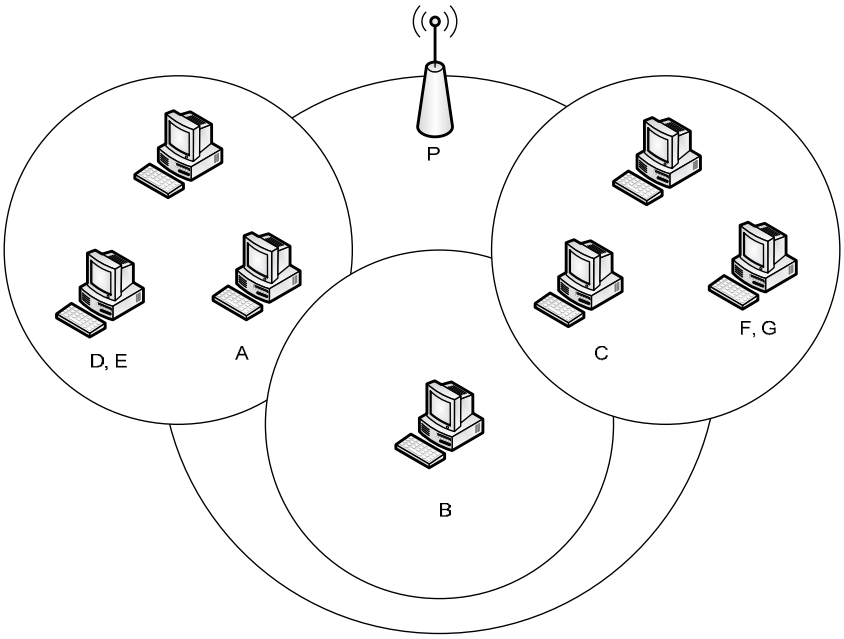


FIGURA 7.4. Representación del efecto de "terminal oculto".

Un algoritmo de reintento aleatorio (*random backoff algorithm*) se pone en marcha cuando una transmisión falla (por ejemplo: no se recibe ACK) o cuando se produce una colisión, haciendo que se aumente de forma aleatoria el tiempo de espera del terminal fallido o de los nodos contendientes, hasta un valor máximo especificado, cuanto mayor sea el número de reintentos. La ventaja de este algoritmo de reintento reside en que impide el bloqueo del medio por una transmisión fallida o por la existencia simultánea de varios terminales colisionantes, aumentando sus períodos de silencio y dando así oportunidad de transmitir a otros terminales.

La gestión de potencia en IEEE 802.11 (apartado 7.3.2) es fundamental para equipos portátiles con baterías, y presenta el modo activo continuo (CAM, *Continuously Active Mode*) y el modo de ahorro energético por sondeo (PSP, *Power Save Polling*). En modo PSP el punto de acceso mantiene los paquetes dirigidos a los nodos inalámbricos durmientes en buzones individuales, hasta que se activan de nuevo. El punto de acceso emite una corta trama de señalización (*beacon*) cada 20 ms, que contiene información de sincronismo y una lista de los terminales con paquetes en espera. Los terminales durmientes mantienen algunos temporizadores en operación, y pueden activarse a cada trama de señalización para reducir la espera al mínimo y consultar inmediatamente si están en la lista, o hacerlo menos frecuentemente aumentando la espera (*latency*) para reducir el consumo de potencia.

El procedimiento de itinerancia de terminales inalámbricos entre puntos de acceso (*roaming*) en la norma 802.11, es muy similar al de las redes móviles celulares en el sentido de seleccionar la señal más fuerte, y la única diferencia reside en que es el terminal inalámbrico el que tiene la decisión respecto del punto de acceso al que ha de conectarse, aunque la norma deja abierta la forma en que se toma esa decisión (apartado 7.4.1). Los mecanismos de seguridad definidos en IEEE 802.11 están orientados a la privacidad y la autenticación de usuario-terminal. La privacidad está basada en el algoritmo de cifrado WEP (*Wired Equivalence Privacy*) y la autenticación en un algoritmo de clave secreta (simétrico).

Las nuevas versiones de IEEE 802.11 “a” y “b” mantienen el método de acceso CSMA/CA de la primera, pero modifican sus modulaciones para aumentar la velocidad. Si bien la versión “a” fue iniciada antes de la “b” (1999), su especificación final es posterior. La norma IEEE 802.11 “g” es una extensión de la IEEE 802.11 “b” pero con multiplexación OFDM. La versión de 5GHz es la “h” e incluye el control de potencia y la selección dinámica de frecuencia. El foro WECA (*Wireless Ethernet Capability Alliance*) también denominado WiFi (*Wireless Fidelity*), está orientado a la certificación de la interoperabilidad de productos y su compatibilidad con las normas IEEE 802.11.

7.5.3. HIPERLAN

Dentro del proyecto BRAN (*Broadband Radio Access Networks*) del ETSI (Eu*ropean Telecommunications Standards Institute*), las normas para WLAN son las denominadas HIPERLAN (*High Performance Radio* LAN) en versiones 1 y 2, que se distinguen en la modulación y método de acceso al medio.

HIPERLAN 2 es compatible en la capa física con IEEE 802.11 a, dado que ambas utilizan la modulación denominada "multiplexación por división en frecuencia ortogonal (OFDM, *Orthogonal Frequency Division Multiplexing*).

La ventaja de OFDM se basa en la eliminación de la interferencia intersímbolo que se produce, como consecuencia del retardo temporal variable debido a los multitrayectos (apartado 7.3.1) en los enlaces sin visión directa (NLOS). Para ello divide el canal en varias subportadoras y transmite sobre ellas datos en paralelo (apartado 3.7.2). Cada símbolo es más largo y elimina retrasos temporales variables (52 subportadoras por canal en HIPERLAN 2). Hay que tener en cuenta la codificación de los datos, secuenciación en serie de los datos y FEC (*Forward Error Control*), y el tipo de modulación de cada subportadora. HIPERLAN 2 adapta dinámicamente el FEC (codificación de canal: Reed Solomon, Viterbi o convolucional y turbocódigos) y el tipo de modulación de cada subportadora, para conseguir altas tasas de bits (Mbps) con buena relación señal-ruido (S/N).

7.5.4. Óptica del espacio libre (FSO)

Este tipo de redes puede considerarse una alternativa a las LAN basadas en fibras ópticas, debido a que también utiliza el margen de los infrarrojos (850 nm y 1550 nm), para la transmisión punto a punto en visión directa (LOS), pero que como su nombre indica el medio empleado es la atmósfera (FSO, *Free-Space Optics*). Por esta razón, la transmisión puede alcanzar distancias desde 100 m a varios kilómetros, y esta limitada por absorción (gases, polvo, obstáculos, etc.) y dispersión debida a cambios en el índice de refracción del medio, que depende de condiciones atmosféricas tales como: lluvia, niebla (dispersión de Mie), nieve, turbulencias por gradientes de viento y temperatura (bolsas de aire con densidad rápidamente variable), etc. Las velocidades binarias alcanzadas están entre los 10 Mbps y 1,25 Gbps.

Si bien la forma de transmisión es punto a punto, pueden realizarse topologías tipo estrella, malla (conexión de todos los terminales con todos) para enlaces redundantes y anillo. Su principal ventaja frente a las LAN de fibra óptica consiste en la facilidad de instalación, que puede efectuarse en pocos días sobre las terrazas o incluso a través de las ventanas de edificios próximos. Asimismo, frente a las WLAN basadas en radiofrecuencia (RF) fuera de las bandas ISM, los infrarrojos no requieren licencia, y su única limitación está en la potencia emitida que debe permanecer dentro de los límites de la IEC (*International Electrotechnical Comission*), según la norma IEC 60825-I.

Debido a que la radiación infrarroja a 1550 nm es absorbida totalmente por la córnea, y por ello no alcanza la retina del ojo, puede utilizarse con mayores potencias que la de 850 nm.

Por esta razón cuando se necesitan altas velocidades en largas distancias, en condiciones atmosféricas difíciles, se utiliza preferentemente la emisión en 1550 nm que permite mayores potencias.

De entre las aplicaciones de las redes FSO cabe destacar las siguientes:

- Ampliación de redes MAN: las compañías operadoras de redes pueden desplegar la tecnología FSO para extender los anillos de fibra metropolitanos hacia nuevas redes, e incluso sus redes troncales para completar anillos SONET (apartado 3.5.3).

- Acceso a la última milla: las redes FSO pueden ser utilizadas como enlaces de alta velocidad para conectar usuarios finales a ISPs y hacia otras redes. También pueden eliminar los bucles locales de los operadores de redes de acceso (CATV y otros) para crear redes corporativas de alta velocidad (apartados 3.9.1 y 3.9.2).

- Redes de empresa: la facilidad de instalación de enlaces FSO permite interconectar segmentos de LAN, ubicados en edificios separados por calles u otro tipo de dominio público con visión directa (LOS).

- Redes de reserva: los enlaces FSO pueden actuar como redes de reserva, para proporcionar una infraestructura redundante a enlaces de fibra óptica.

- Redes troncales: las FSO pueden ser utilizadas para soportar tráfico de telefonía celular, desde las antenas de las estaciones base (apartado 7.4.1) hacia la infraestructura de la RTC (Red Telefónica Conmutada).

- Implantación rápida de servicios: las redes FSO permiten ofrecer todos los servicios de una red de acceso de fibra (FTTC, *Fiber To The Curve*; FTTH, *Fiber To The Home*) de forma provisional, a todos los usuarios solicitantes, mientras está siendo tendida la infraestructura definitiva de fibra óptica.

7.6. REDES DE ÁREA PERSONAL (WPAN)

7.6.1. Introducción

Son redes de comunicación de uso privado que si bien presentan en general unas características técnicas similares a las WLAN, su rango de cobertura es inferior (10 m), con velocidades más limitadas (hasta 1 Mbps). Su nombre (WPAN, *Wireless Personal Area Network*) proviene de que están orientadas a la conexión inalámbrica de terminales portátiles (PDA, *Personal Digital Assistants*), y en general de equipos próximos entre sí, ya sean periféricos de ordenador (impresoras, escáneres, ratones, teclados, etc.) o equipos de electrónica de consumo (cámaras de vídeo, lectores de DVD, auriculares, etc.). En otras aplicaciones, como la monitorización de pacientes, también se denominan "redes vestibles" (*wearable*) al poder ser instaladas sobre la ropa de una persona quieta o en movimiento. Una característica de las WPAN es la "capacidad espacial" (apartado 7.3.3), calculada en términos transporte de bits en la unidad de tiempo por la unidad de superficie de cobertura (bps/m²). Las normas para WPAN son: Bluetooth, UWB, Zigbee e IrDA.

7.6.2. Bluetooth (IEEE 802.15.1)

La norma Bluetooth, inicialmente propuesta por la firma Ericsson, ha sido promovida por varios fabricantes de equipos de comunicaciones en 1998, y presenta una modulación FHSS dentro de la banda ISM (2400 MHz a 2483,5 MHz), con método de acceso TDMA y velocidad hasta 1 Mbps. Dentro del área de cobertura de 10m puede soportar hasta 10 picorredes (*piconet*) para formar lo que se denomina una red de "dispersión" (*scatternet*), con un dispositivo principal y hasta 7 subordinados en cada picorred.

Su nombre proviene de un rey danés del siglo X que unió a varios reinos escandinavos. Si bien fue concebida inicialmente para eliminar cables entre un PC y sus periféricos, posteriormente evolucionó a la conexión con teléfonos celulares y PCs portátiles, hasta pasar finalmente a conectar cualquier producto electrónico portátil o fijo. Los terminales presentan diversos modos de control de potencia (apartado: 7.3.2): en el de "mantenimiento" (*hold*) no hay intercambio de datos, en el de "sondeo" (*sniff*) se escucha a nivel bajo, en el de "soporte" (*standby*) los dispositivos que no están en una picorred consultan mensajes en un tiempo de 1,5 s, mientras que en el "activo" se produce el intercambio de información.

El grupo de trabajo IEEE 802.15.1 creado en 1999 desarrolla seis proyectos relacionados con WPAN. El TG1 (*Task Group*) trata de especificar una norma basada en Bluetooth, para conseguir la interoperabilidad entre terminales WPAN y redes IEEE 802.11. El TG2 se orienta al análisis de la coexistencia entre WLAN y WPAN, estudiando las interferencias que se producen en la banda ISM de 2,4 GHz. El TG3 investiga en WPAN de alta velocidad (20 Mbps). El MC (*Marketing Commitee*) orienta las WPAN desde el punto de vista del mercado, para alcanzar consensos con grupos industriales respecto a los requisitos técnicos entre los diferentes proyectos. El R2SG (*Radio 2 Study Group*) observa la evolución de Bluetooth para establecer recomendaciones al grupo de trabajo. Finalmente el LRSG (*Low Rate Study Group*) analiza la viabilidad de una solución de baja velocidad con baterías sencillas de larga duración, para terminales como: sensores, equipos de control remoto, domótica, juguetes interactivos, etc.

7.6.3. UBW (IEEE 802.15.3)

Propuesta en 1998 por una asociación de miembros de la industria y del sector académico en USA, la "ultra banda ancha" (UWB, *Ultra Wide Band*) representa una tecnología emergente (apartado 3.3), basada en la transmisión de información mediante impulsos muy cortos, como en el radar, inferiores al nanosegundo (ns), muy rápidos y de muy baja potencia (0,25 mW). Esta baja potencia es posible gracias a que UWB no presenta la atenuación de Rayleigh que sufren las ondas continuas en la propagación convencional de radio, que ha de compensarse con un aumento de potencia. El margen de frecuencias utilizado está entre 3,1 GHz a 10,6 GHz, y aunque cubre bandas de frecuencia ya asignadas en el rango de 1 GHz a 6 GHz, su baja potencia de emisión hace imperceptible la interferencia.

El elevado ancho de banda, del orden de 500 MHz e incluso de GHz, y las extremadamente bajas potencias, posibilitan la transmisión de imágenes (multimedia) y en general el acceso a Internet de banda ancha en distancias cortas. Una gran ventaja es su capacidad de alcanzar velocidades superiores al Gbps. Desde 2003 existen dos grupos de trabajo, el IEEE 802.15.3 a (USA) y el ETSI ERM TG31 cuyo objetivo es generar una norma sencilla, económica, de bajo consumo y elevada velocidad (110 Mbps con 100 mW y 200 Mbps con 250 mW) con la calidad de servicio necesaria para las transmisiones multimedia para WPAN. Las realizaciones prácticas han sido dirigidas por WiMedia Alliance de Intel y por UWB Forum de Motorola, ambas incompatibles por sus diferencias en los protocolos de la capa física.

La WiMedia Alliance utiliza la multiplexación MB-OFDM (*MultiBand-OFDM*) que divide el espectro de radio en 14 canales con velocidades superiores a los 400 Mbps. La tecnología *Wireless USB* de Intel trata de sustituir la versión cableada de USB (480 Mbps en la versión 2) en cámaras fotográficas, impresoras, PDAs, lectores MP3, ordenadores portátiles y otros productos de gran consumo popular.

El UWB Forum utiliza DS-UWB (*Direct Sequence-UWB*) con solamente 2 canales, y códigos que permiten mayor variedad de alcances y velocidades hasta 1,4 Gbps. También entre sus objetivos está el mercado USB con la denominación *Cord-Free USB* para sus productos.

7.6.4. ZigBee (IEEE 802.15.4)

La ZigBee Alliance es un consorcio de empresas, en su mayoría fabricantes de semiconductores, como Philips, Mitsubishi, Samsung y Motorola, que han propuesto una red inalámbrica para terminales con las siguientes características: baja velocidad, bajo coste, ciclo de trabajo inferior al 1% (más del 99% del tiempo está inactivo), transmisión de paquetes de datos cortos, bajo consumo de potencia (autonomía de 2 años sin cambiar la pila), elevado número de dispositivos interconectados (más de 65.536 terminales), y comunicación segura y fiable.

Las orientaciones de mercado de esta red son, entre otras, las siguientes: automatización de edificios y hogares (seguridad, alarmas, climatización, lectura de contadores de agua, gas y electricidad, control de iluminación, control de accesos, riego, etc.); atención sanitaria (monitorización de pacientes, implantes clínicos, trazabilidad biométrica, etc.); control industrial (sensores y actuadores, control de procesos, instrumentación remota, etc.); interfaces hombre-máquina (teclados, ratones, palancas, botoneras, etc.) y mercado de consumo general (electrodomésticos de línea blanca, juguetes, juegos, material deportivo, etc.). A diferencia de otras normas para WPAN como Bluetooth y UWB, ZigBee no se plantea aplicaciones multimedia de gran ancho de banda con terminales en comunicación permanente (*always-on*), sino más bien satisfacer aplicaciones basadas en la actuación y captación de eventos muy simples y cortos como cerrar una puerta, encender una lámpara, conectar una máquina, medir una temperatura, etc.

El consorcio ZigBee propone una pila de protocolos de cuatro capas del modelo OSI. Para las dos primeras (PHY y MAC) adoptó la norma IEEE 802.15.4 de 2003, y definió sus propios protocolos para las capas de red (NWK)

y aplicación (APS, *Application Support*). Dado que la red ZigBee está destinada a satisfacer una gran variedad de aplicaciones, los usuarios y fabricantes de terminales pueden definir sus protocolos específicos en la capa de aplicación (ZDO, *ZigBee Device Object*) sobre la subcapa APS. En la capa física (PHY) el IEEE utiliza varias bandas libres con canales, velocidades y zonas geográficas diferentes tales como: 2,4 GHz (ISM), 16 canales, 250 kbps (internacional); 915 MHz, 10 canales, 40 kbps (USA); 868 MHz, 1 canal, 20 kbps (Europa). La multiplexación es CDMA con DSSS (apartado 7.5.1), y la modulación es del tipo BPSK (apartado 3.8.2.2) para las dos bandas bajas y O-QPSK (*Orthogonal-QPSK*) para la banda de 2,4 GHz. El nivel de potencia se limita a –3 dBm (apartado 7.3.3), aunque puede variar según el entorno geográfico, la cobertura, la velocidad y la tasa de error de paquetes permitida por la aplicación. La cobertura está entre 10 m y 100 m, siendo típica la de 30 m en aplicaciones de interior, aunque en aplicaciones de exterior puede haber alcances superiores a los 100m (por ejemplo en monitorización ambiental).

Debido a que una red ZigBee no tiene infraestructura (*Ad-hoc,* apartado 7.5.1), sus propios terminales asumen tareas como dispositivos de configuración, gestión y mantenimiento de la red global, y por ello presentan diferentes capacidades físicas y lógicas, que se describen a continuación. Desde el punto de vista de su realización física pueden ser de funcionalidad completa (FFD, *Full Function Device*) o de funcionalidad reducida (RFD, *Reduced Function Device*). Un dispositivo FFD puede asociarse con cualquier topología, puede asumir tareas de coordinador de red, y sus capacidades de computación y memoria le permiten actuar como encaminador (*router*) o como dispositivo con interfaces de usuario complejas, además de comunicarse con cualquier otro tipo de terminal. Un dispositivo RFD presenta las más bajas capacidades de procesamiento y memoria para reducir su coste (por ejemplo como sensor o actuador), y en consecuencia está limitado a las topologías en estrella, no puede ser coordinador de red y solamente puede comunicarse a través de un coordinador de red (más del 99% del tiempo está en estado “durmiente”).

Desde el punto de vista de las capacidades lógicas un dispositivo FFD puede actuar como coordinador y encaminador, además de como terminal de red. Un coordinador inicia la configuración de una red, administra los nodos y almacena información sobre ellos, necesitándose al menos uno por red. El encaminador participa en la red estableciendo trayectos de comunicación para hacer llegar los mensajes a los nodos emparejados. Todos los dispositivos deben tener direcciones de 64 bits, aunque se pueden asignar direcciones cortas de 16 bits (65.536 dispositivos) para reducir el tamaño de los paquetes.

Las distintas estructuras de trama se han diseñado para minimizar la complejidad y hacerla suficientemente robusta para su transmisión sobre un canal ruidoso. Cada capa del protocolo añade una cabecera a la estructura. La capa MAC de la norma IEEE 802.15.4 define cuatro tipos de estructuras de trama: señalización (el coordinador transmite señalizaciones), datos (para transmitir datos), reconocimiento ACK (confirma recepciones de tramas correctas) y órdenes MAC (transferencia de control entre entidades MAC). El tamaño total de la cabecera de un paquete sencillo es de 16 a 31 octetos (mínimo de 128 bits), dependiendo del tamaño de la dirección, y sin incluir cabeceras de seguridad.

El coordinador define una estructura opcional denominada supertrama, que transmite entre dos señalizaciones de red y se divide en 16 ranuras de tiempo equiespaciadas. La trama de señalización se transmite en la primera ranura de cada supertrama, y se usa para sincronizar dispositivos relacionados, para identificar la red y para describir la estructura de las supertramas. La supertrama permite dar acceso al canal a las aplicaciones que requieran un bajo retardo o un ancho de banda específico (datos de alta prioridad), asignando las correspondientes "ranuras de tiempo garantizado" (GTS, *Guaranteed Time Slots*), durante el llamado "período libre de contienda" (CFP, *Contention Free Period*). El método de acceso al medio es CSMA/CA ranurado durante el CFP, mientras que fuera de la supertrama, en el período de contienda, es el CSMA/CA normal (apartado 7.5.2).

En la capa de red se configuran diversas alternativas de conexión para adaptarse a la aplicación y al entorno. Existen cuatro topologías principales: "entre iguales" (*Ad-hoc* o también *peer-to-peer*) en la que dos terminales se conectan directamente entre sí; "estrella" (*star*), en la que todas las comunicaciones entre terminales pasan por el único coordinador; "racimo" (*Cluster-Tree*), que es una agrupación de estrellas conectadas mediante controladores centrales al único coordinador; "malla" (*mesh*), que presenta varios racimos enlazados por encaminadores conectados a un único coordinador, y permite proporcionar rutas diferentes entre dos terminales cualesquiera con objeto de aumentar la fiabilidad de la comunicación y disminuir el consumo de energía (auto-configuración y auto-regeneración). Estas topologías sugieren el nombre de ZigBee, en el sentido de que las redes mencionadas se asemejan a enjambres de tamaño variable en los que las abejas (como los terminales) se comunican entre ellas volando en zigzag (Fig. 7.5).

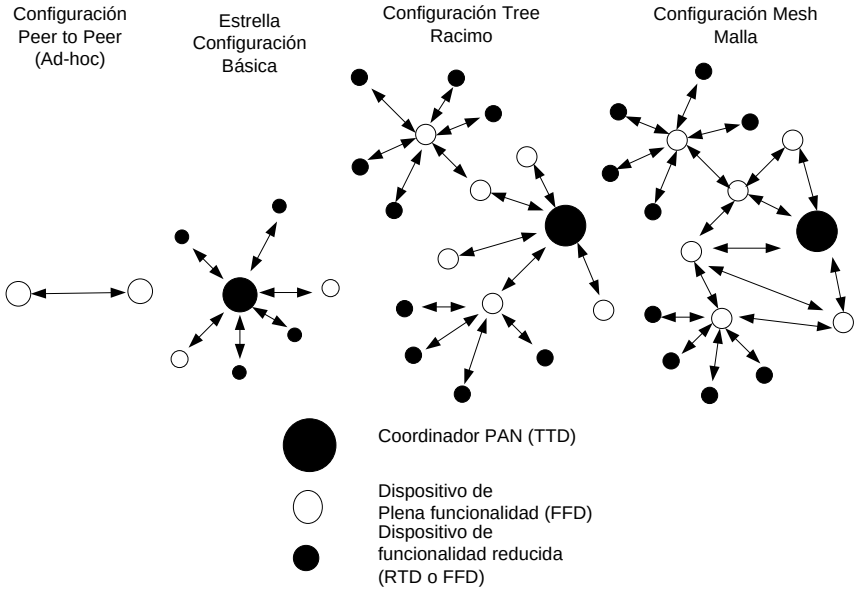


FIGURA 7.5. Topologías de la norma IEEE 802.15.4 (ZigBee).

Las redes ZigBee soportan muchos tipos de tráfico con características únicas, tales como datos periódicos, datos intermitentes y datos repetitivos con "requisitos de bajo retardo". Los "datos periódicos" están definidos normalmente por la aplicación, como por ejemplo un sensor o un medidor inalámbrico. Los datos periódicos se transmiten habitualmente usando un sistema de señalización por lo que el sensor se "despierta" a una hora determinada, comprueba la señalización, intercambia los datos y vuelve al estado "durmiente" (ejemplos típicos son los contadores de agua, gas o electricidad que transmiten las medidas a una central). Los "datos intermitentes" pueden ser manejados en un sistema sin señalizaciones o desconectado. En operación desconectada el dispositivo sólo se incorpora a la red cuando necesita comunicarse, ahorrando así una gran cantidad de energía (por ejemplo un interruptor de luz). Esto permite un ciclo de trabajo muy corto (menor del 1%) ya que un nodo que se enciende puede ser reconocido y asociado a la red en 30ms, despertarse y dormirse le lleva 15ms, lo mismo que acceder al canal y enviar un dato. Los datos repetitivos con "requisitos de bajo retardo" pueden usar ranuras de tiempo garantizado GTS, que aportan calidad de servicio (QoS) asignando a cada dispositivo una duración específica de tiempo definida por el coordinador en la supertrama, para realizar tareas de gran prioridad que no deben sufrir penalizaciones de tiempo de contienda ni retardos (por ejemplo, mensajes de: alertas médicas, control industrial, avisos de seguridad, etc.).

Tanto en la capa MAC como en la capa de red (NWK) ZigBee soporta varios niveles de seguridad (cifrado y autenticación) mediante el algoritmo AES (*Advanced Encryption Standard*) con una clave de 128 bits. Esto es fundamental en aplicaciones que requieran altos grados de confidencialidad y de autenticidad de la información, para evitar riesgos como consecuencia de fallos catastróficos o alteraciones en el contenido de la información, en sectores sensibles tales como: inmótica (automatización de edificios), control industrial, monitorización de pacientes, etc.

Dentro de la capa de aplicación de ZigBee, la subcapa APS se encarga de las tareas de "descubrimiento" (*Device Discovery*) y de "asociación" (*Binding*), es decir, de identificar a los dispositivos que están operando en el rango de cobertura de la red, y de enlazar a dos o más dispositivos basándose en sus requerimientos y reenviar mensajes entre dispositivos limítrofes respectivamente. La subcapa ZDO asume la definición del papel de los dispositivos dentro de la red (coordinador, encaminador, controlador de estrella y terminal), y el inicio o la respuesta a peticiones de asociación, además del establecimiento de una conexión segura entre dispositivos de la red. Los perfiles de servicio son conjuntos de capacidades que utilizan las capas de protocolos de un dispositivo ZigBee, así un ejemplo de perfil sería el control de iluminación de una casa.

7.6.5. IrDA

La norma IrDA tiene gran aplicación en equipos portátiles de gran consumo como: PCs, teléfonos, periféricos de ordenador, telemandos, etc. La comunicación se basa en la utilización de infrarrojos (IR entre 850 nm y 900 nm) para conexiones punto a punto de corto alcance (distancias inferiores a los 10 m) con visión directa (LOS), que emplea un haz estrecho en el emisor y un receptor con un pequeño ángulo de apertura. Esta norma se desarrolló en 1993 por el consorcio IrDA (*Infrared Data Association*) formado por fabricantes de USA y Japón líderes en ordenadores, equipos de comunicaciones y semiconductores.

La primera versión de 1994 (V 1.0) también llamada SIR (*Slow* IR), especifica una transmisión serie asíncrona en modo alternado (*half duplex*) con modulación OOK (apartado 7.4.1), y permite alcanzar una velocidad de 115,2 kbps. La segunda versión en 1995 (V 1.1) incorporó la transmisión síncrona y algunos cambios en la modulación para periféricos rápidos: en la especificación MIR (*Medium* IR) se cambian las duraciones de los bits en la modulación OOK para alcanzar 1,152Mbps, mientras que en la especificación FIR (*Fast* IR) una

modulación de impulsos en posición (PPM, *Pulse Phase Modulation*) permite obtener una velocidad de 4 Mbps (apartado 3.8.2.1). Nuevas realizaciones permiten alcanzar 16 Mbps a una distancia de 2 m.

7.7. REDES FIJAS DE ACCESO INALÁMBRICO (WMAN)

Se agrupan bajo este apartado las redes inalámbricas que ofrecen acceso a usuarios con nula ó limitada movilidad. Su interés deriva de la rapidez de la instalación con similares características técnicas que las redes de cables de área metropolitana (MAN), lo que reduce drásticamente los costes de infraestructura, y por ello representan una alternativa más económica que las redes cableadas, con las que entran en competencia. Para su estudio se clasifican en redes de banda ancha y de banda estrecha.

7.7.1. Banda ancha (MMDS y LMDS)

Son redes fijas de comunicación inalámbrica de uso público con una cobertura superior a la de las WLAN y LAN (3 km a 56 km), y están destinadas a competir en el mercado empresarial con las redes de los operadores de: televisión por cable CATV (apartado 3.9.1), y telefonía fija con servicios RDSI y xDSL (apartado 3.9.2). Su mercado fundamental es el acceso a Internet y otros servicios de datos, no obstante por su misma situación competitiva es dudosa su integración con redes de telefonía fija y móvil, y CATV, de otros operadores. Las soluciones inalámbricas fijas tienen la ventaja de proporcionar una mayor eficiencia espectral en términos de "bps/Hz" (apartado 7.3.3) utilizando antenas de haz estrecho en el entorno fijo, lo que les permite mayores velocidades cuando disponen de un elevado ancho de banda, como es el caso de LMDS y MMDS. Las redes fijas de acceso inalámbrico en banda ancha MMDS y LMDS, como su nombre indica no permiten la movilidad de los usuarios, aunque presentan áreas de cobertura celulares y macrocelulares, y sus velocidades sobrepasan los 20 Mbps, de ahí su denominación de "banda ancha".

A continuación se detalla cada una de ellas:

- Las redes fijas de acceso inalámbrico en banda ancha, para servicios de "distribución multicanal multipunto" (MMDS, *Multichannel Multipoint Distribution Service*) presentan las siguientes características: utilizan radiofrecuencias en el margen de 2,5 GHz a 2,69 GHz (UHF), y por ello se denominan de "bandas bajas" (inferior a 11 GHz), su cobertura es de

30 km a 70 km en comunicación punto-multipunto, presentan la modulación QAM64 (apartado 3.8.2.2), tienen menores costes de despliegue que LMDS pero con menor número de usuarios, y la velocidad en el sentido red-usuario (*downlink*) es de 27 Mbps a 34 Mbps y en el sentido usuario-red (*uplink*) es de 1 Mbps a 3 Mbps.

Su mercado es el de los usuarios de zonas rurales con baja densidad de población. El retorno se hace por estaciones de base, y por ello puede presentar dificultades en aplicaciones que requieran alta velocidad de retorno como el hospedaje de páginas Web en Internet (*Web hosting*). Inicialmente representaron una alternativa a las redes CATV en zonas inaccesibles a los cables, y por ello fueron históricamente denominadas en el ámbito anglosajón como "cable inalámbrico" (*wireless cable*) y "horizontal DBS".

- Las comunicaciones terrestres mediante redes fijas de acceso inalámbrico en banda ancha para servicios de "distribución local multipunto" (LMDS, *Local Multipoint Distribution Service*) presentan las siguientes características: utilizan las microondas en el margen de 27,5 GHz a 29,5 GHz, y por ello se denominan de "bandas altas" (superior a 11 GHz), presentan modulaciones QPSK, QAM16 y QAM64, necesitan visión directa de la antena (LOS), su cobertura es de 3 km a 10 km por célula según la pluviosidad, a diferencia de MMDS son bidireccionales, tienen mayores costes de despliegue que MMDS pero con mayor número de usuarios debido a su cobertura celular como la de telefonía móvil, y la velocidad en el sentido red-usuario es de 27 Mbps a 55 Mbps y en el sentido usuario-red es de 1 Mbps a 10 Mbps. Se destinan a entornos ciudadanos con alta densidad de población. No parece que tengan resuelto el problema de suministrar telefonía, debido a la competitividad con los operadores de otros tipos de redes (RTC, telefonía móvil y CATV).

7.7.2. Normas ETSI

Desde las instituciones de normalización internacionales han surgido alternativas a LMDS y MMDS que han recibido diversas denominaciones. Dentro del proyecto BRAN (apartado 7.4.2) del ETSI, las normas para redes fijas de acceso inalámbrico en banda ancha, son las denominadas HIPERACCESS para bandas altas e HIPERMAN para bandas bajas. HIPERACCESS presenta una cobertura de 5 km, con una velocidad de 25 Mbps.

HIPERMAN presenta una cobertura de 10 km con una velocidad de 6Mbps sin visión directa (NLOS), para lo cual utiliza la multiplexación OFDMA (apartado 3.7.2). La norma HIPERLINK del ETSI está orientada hacia la conexión entre HIPERACCESS y redes HIPERLAN, y sus características son: cobertura de 50 m a 500 m, banda de 17 GHz y velocidad de 155 Mbps.

El IEEE presenta la norma IEEE 802.16 para bandas altas (11 GHz a 66 GHz), mientras que la versión IEEE 802.16 "a" está destinada a las bandas bajas (2 GHz a 11 GHz) con características similares a HIPERMAN. En IEEE 802.16 existen muchas opciones, de tal modo que puede considerarse como un conjunto de herramientas en varias cajas (*toolbox*), y se describen a continuación.

7.7.3. WiMAX (IEEE 802.16)

Esta norma está orientada a proporcionar accesos de alta capacidad a distancias inferiores a 50 km con velocidades hasta 70M bps, permitiendo transmisión y recepción en los dos extremos por duplexación FDD en bandas licenciadas (tráfico de voz) y TDD en bandas libres (tráfico asimétrico como Internet). La norma original IEEE 802.16 especifica el rango espectral de 10-66 GHz, y posteriormente la 802.16a añadió el margen de 2-11 GHz que además de introducir bandas libres (ISM) presenta la ventaja de realizar comunicaciones sin visión directa (NLOS) (apartado 7.3.1). Una característica importante de la norma es que define una capa de acceso al medio (MAC) que soporta múltiples especificaciones físicas (PHY), lo que supone un intento de mezclar muchas tecnologías para cubrir necesidades muy variadas con claras rentabilidades prácticas. La revisión IEEE 802.16e (año 2005) es la que añade movilidad ,ya que las anteriores se refieren a sistemas estáticos, es decir, si el terminal cambia de célula se pierde la conexión.

El WiMAX Forum (año 2003) es la entidad encargada de proporcionar certificaciones de interoperabilidad de los equipos de diferentes fabricantes compatibles con las especificaciones IEEE 802.16. Por esta razón se suelen utilizar indistintamente ambos términos para referirse a estas normas. Una variante del estándar IEEE 802.16e de tecnología móvil para Internet banda ancha, desarrollada por la industria de las telecomunicaciones en Corea, recibe la denominación de WiBRO (*Wireless Broadband*) y puede encontrarse también como denominación alternativa de esta versión.

Los beneficios de las soluciones basadas en WiMAX (*Worldwide interoperability for Microwave Access*), tanto en bandas licenciadas como en

bandas libres, son la eficiencia en costes, escalabilidad y flexibilidad. En general las bandas licenciadas ofrecen una mayor calidad de servicio con un mayor coste de entrada por el pago del espectro (problemas de subastas, precios altos y fuertes retrasos), mientras que las bandas libres presentan una menor calidad por restricciones de potencia y la acción de señales interferentes de otros operadores, pero tienen un menor coste y una mayor interoperabilidad. Los proveedores de servicios en mercados emergentes, como países en desarrollo o países avanzados con áreas subdesarrolladas (por ejemplo en zonas rurales), pueden reducir el tiempo de acceso al mercado y los costes iniciales optando por bandas libres, e incluso los operadores con licencia pueden emplear las bandas libres como red de reserva de su red principal.

Emplea diversos métodos de multiplexación como TDM, FDM, OFDM, WOFDM (símbolos de entrenamiento y código FEC con baja tasa de interferencia) y Flash-OFDM (*Fast Low-Latency Seamless Handoff*) para obtener mayor ancho de banda, distancia y movilidad. La especificación IEEE 802.16e utiliza solamente el método OFDM con un número variable de portadoras en múltiplos desde 128 a 2048, y los perfiles de la banda libre de 2,4 GHz con TDD emplean desde 512 a 1024 subportadoras.

Utiliza una técnica de diversidad denominada MISO (*Multiple Input-Single Output*) debido a que en contraposición a MIMO (apartado 7.4.3.1), tiene dos antenas emisoras y solamente una receptora. Se efectúan dos transmisiones consecutivas en el tiempo, de tal forma que para una determinada subportadora envía el símbolo A por la antena 1 y el símbolo B por la antena 2, mientras que en el instante siguiente lo hace al revés. El receptor recibe en su única antena el mismo símbolo por canales y períodos diferentes, denominándose combinación espacio-tiempo (conocida también como transmisión Alamouti).

Si bien esta norma utiliza procedimientos de modulación multibit (apartado 3.8.2.2) dado que los mayores órdenes de modulación permiten enviar más bits por símbolo (estado de modulación), y por tanto alcanzar más altas velocidades binarias (*throughput*) y valores crecientes de eficiencia espectral, el uso de técnicas de modulación de orden alto como QAM64 implica necesariamente una elevada relación S/N para eliminar las interferencias y mantener una tasa de error (BER) moderada (apartado 3.7.1). Por ello WiMAX utiliza la técnica de modulación adaptativa (apartado 7.3.3), de tal forma que a mayor distancia de la estación base menor debe ser el orden de la modulación, pasando por: QAM64, QAM16, QPSK Y BPSK (Fig. 7.6).

El valor de la relación S/N en dB (apartados 3.3 y 3.5.1) en el receptor determina el orden seleccionado según: 22 dB (QAM64), 16 dB (QAM16) y 9 dB (QPSK).

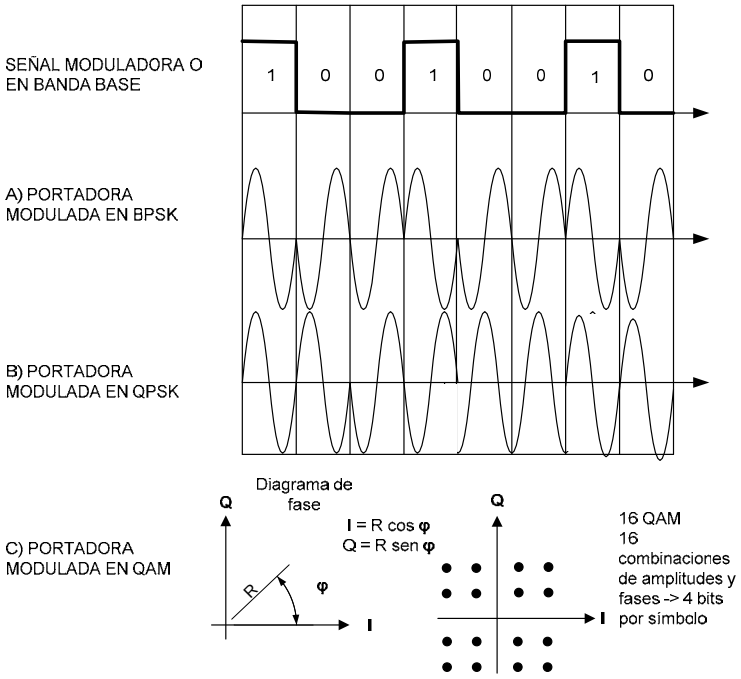


FIGURA 7.6. Tipos de modulación para WiMAX.

Para el control de errores WiMAX utiliza codificación de canal tipo FEC (apartado 7.3.2), con códigos tipo Reed-Solomon, Viterbi y Fano, así como métodos ARQ cuando fallan los FEC como la variante H-ARQ (*hybrid-ARQ*) mediante el reenvío de la información con errores. Existen diferentes formas de realización de H-ARQ, dado que en el reenvío se puede añadir información redundante para facilitar la decodificación, cada retransmisión puede ser autocodificable, se puede aplicar diversidad temporal ponderando múltiples copias de un paquete según la relación S/N de cada transmisión, etc.

En cuanto a la banda de emisión con licencia para operadores de banda ancha cabe destacar la comprendida entre 3,4-3,6 GHz, aunque hay posibilidades en el rango de 3,3-3,4 GHz (China e India) y en el de 3,6-3,8 GHz (Francia, UK, resto de Europa y USA). La banda de 3,4-3,6 GHz ha sido reservada en la mayoría de los países para servicios fijos, móviles y por satélite

(apartado 3.9.3), a excepción de USA. En ellas WiMAX opera tanto en TDD como FDD, existiendo canales de 3,5 MHz y 7 MHz.

La banda de 2,5-2,6 GHz (MMDS) fue reservada por USA, México, Brasil y países asiáticos (Singapur) para transmisión de TV, y el WIMAX Forum trata de hacerla disponible internacionalmente para aplicaciones de banda ancha tanto fijas como móviles, en FDD y TDD con canales de ancho de banda de 5 MHz y 6 MHz (31 canales en USA). En Asia, Australia, Corea del Sur y Nueva Zelanda se emplea la banda de 2,3 GHz, que se espera que pueda ser cubierta con los sistemas de 2,5 GHz. Esta banda de 2,3, llamada WCS (*Wireless Communication Service*), está formada por dos márgenes de 15 MHz (2305-2320 MHz y 2345-2360 MHz) con una separación de 25 MHz reservada para servicios de radio digital (DARS), y se considera que esta subbanda intermedia puede suponer una fuente de interferencias. Estas bandas están orientadas a los servicios móviles bajo la norma IEEE 802.16e ya que presentan un mejor comportamiento para este tipo de servicios.

Dado que las señales se propagan más lejos cuanto menor sea la banda de frecuencia utilizada, en las redes móviles existe una relación directa con el número de estaciones base utilizadas para cubrir un área de servicio dada, de tal forma que a menor frecuencia menor es el número de estaciones base necesarias, lo que disminuye el coste de las instalaciones, siendo por tanto fundamental el acceso a bandas de frecuencia menores para un despliegue rápido y económico. El WiMAX Forum trabaja con los organismos reguladores para la reserva de espectro tanto licenciado como libre, en bandas por debajo de 1 GHz, especialmente en aquéllas que ya no sean utilizadas por migración a la TV digital. En USA se estima que estarán disponibles las del rango de 700 MHz.

Las topologías utilizadas son: punto a punto, punto a multipunto, multipunto a multipunto (enrutamiento de tráfico con redes de varios saltos tipo Internet), metropolitanas (mezcla de las tres anteriores en entorno urbano), nodos mixtos (similar a las metropolitanas pero en entornos de interior) y tipo malla que basadas en IEEE 802.11s, los nodos son como encaminadores que mantienen configuraciones en caminos dinámicos (apartado 7.10.2).

7.7.4. IEEE 802.20 e IEEE 802.22

Otras normas relacionadas con WiMAX son la IEEE 802.20 y la IEEE 802.22. La norma IEEE 802.20 (MBWA, *Mobile Broadband Wireless Access*) está orientada al acceso móvil mundial de banda ancha basado en paquetes, y optimizado para el transporte de servicios IP (red WAN móvil).

Se especificarán las capas MAC y PHY con bandas de frecuencia inferiores a 3,5 GHz, tasas de bit superiores a 1 Mbps y desplazamientos de terminales hasta 250 km/h. Se estima que con canales de ancho de banda de 5 MHz la tasa máxima de usuario pueda llegar a 4 Mbps en sentido red-usuario (bajada) y 1,2 Mbps en sentido contrario (subida), con tasas agregadas de 16 Mbps en bajada y 3,2 Mbps en subida, con una eficiencia espectral superior a 1 bps/Hz (apartado 7.3.3). La especificación IEEE 802.20 se propone para su uso en bandas licenciadas y está llamado a competir con la IEEE 802.16e de WiMAX.

La norma IEEE 802.22 (WRAN, *Wireless Regional Area Network*) se propone el uso de bandas libres en el espectro reservado para el servicio de difusión de TV. Los márgenes previstos serán las bandas libres de 54-862 MHz, lejos del rango de WiMAX. Las señales a estas frecuencias se propagan hasta los 40 km o más en función de la estación base, y se considera rentable para zonas con poca población como alternativa a las líneas xDSL o módems de cable, en donde el despliegue de cables es económicamente inviable. Se considera que esta norma será interoperable con WiFi, pero no con WiMAX al no cubrir ésta las bandas de TV.

7.7.5. Banda estrecha (WLL)

Reciben la denominación de "lazo local inalámbrico" (WLL, *Wireless Local Loop*), las redes fijas de comunicación inalámbrica de uso público que con una cobertura microcelular y en banda estrecha (velocidad inferior a 2 Mbps), están destinadas a competir en el mercado de las redes de los operadores de telefonía fija (RTC) y servicios RDSI (apartado 3.9.2). Su mercado fundamental es proporcionar una red de telefonía y datos de bajo coste en zonas rurales o de población dispersa, que por su nivel económico o por la ausencia de infraestructuras de telecomunicaciones, no pueden acceder a las redes de los operadores de telefonía y datos convencionales. Aunque existen realizaciones propietarias de este tipo de redes, las normas internacionales que se han utilizado en estas aplicaciones provienen del ámbito de la telefonía inalámbrica de corto alcance, inicialmente diseñadas para entornos urbanos y suburbanos. Estas normas se denominan DECT, PACS y PHS.

La norma DECT (*Digital Enhanced Cordless Telecommunications*) está propuesta por el ETSI. En 1994 el JTC (*Join Technical Committee*) presentó en USA la norma PACS (*Personal Access Communications System*). A finales de 1993 el RCR (*Research and developement Center for Radio systems*) japonés, conocido como ARIB (*Association of Radio Industries and Bussinesses*) aprobó

la norma RCR STD-28, que completada por el TTC (*Telecommunications Technology Committee*), dio lugar a la aparición comercial de PHS (*Personal Handyphone System*) en 1995.

El interés de estas tres normas internacionales en las aplicaciones WLL se basa en las siguientes características comunes:

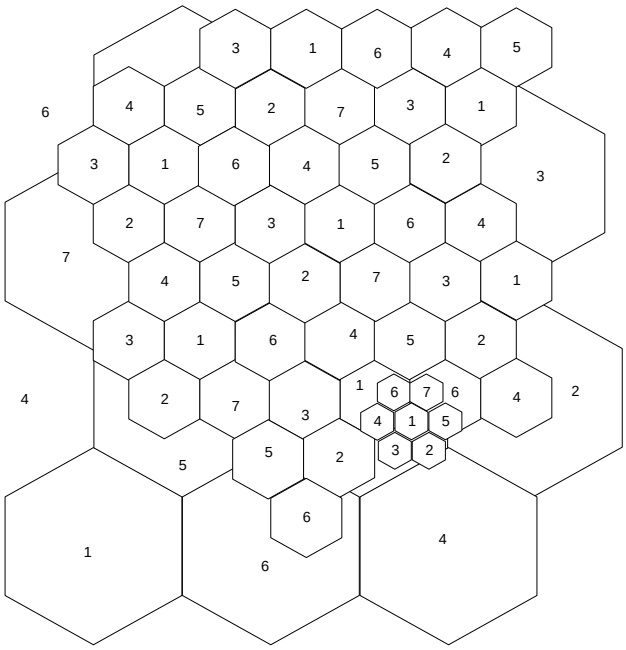
- Calidad de voz: basada en la codificación de voz denominada "modulación por impulsos codificados adaptativa y diferencial" (ADPCM, *Adaptive differential Pulse Code Modulation*), con una velocidad de 32 kbps. Este método de codificación ofrece una calidad de voz superior a otros utilizados en normas de redes celulares de baja velocidad, y es comparable a la calidad proporcionada por PCM (apartado 3.8.2.1) en redes cableadas.
- Conexión a red: capacidad de proporcionar conexión con la red de cable mediante conmutadores comunes, sin necesidad de una infraestructura separada, así como la posibilidad de conexión a la RDSI.
- Mayor velocidad: permiten velocidades de datos de 32 kbps, que son superiores a las proporcionadas por las redes móviles (máximo de 9,6 kbps). Existen posibilidades de aumentos de velocidad en múltiplos de 32 kbps.
- Capacidad: debido a su cobertura microcelular pueden satisfacer requerimientos en áreas urbanas densas, proporcionando mayores capacidades que los sistemas y normas macrocelulares convencionales.
- Coste: por su sencillez de instalación y bajo coste del equipamiento, son especialmente adecuadas para aplicaciones WLL en que la economía supone un factor decisivo.

7.8. REDES DE ACCESO CELULAR (WWAN)

7.8.1. Configuración de los sistemas celulares

En los sistemas celulares el despliegue de las estaciones de base se realiza estructurando el área de cobertura mediante polígonos o células que pueden formar grupos (siete, doce, etc.), y dentro de cada grupo se reparte toda la banda asignada al sistema. En general un sistema celular está formado por sectores en forma de polígonos hexagonales distribuidos en nueve grupos con siete células

cada uno (Fig. 7.7). La asignación de subbandas de frecuencia en cada grupo sigue la misma ordenación en sentido horario, y la ubicación de los grupos impide que existan dos células adyacentes con la misma subbanda de frecuencia, repitiéndose las frecuencias 1-6-4-5-2-7-3 en cada línea horizontal, y 1-4-2-3-6-5-7 en cada línea vertical. De esta configuración en el plano se deduce que cuanto menor sea el radio de las células (el lado del hexágono es igual al radio de la célula) mayor es la capacidad de reutilización de las frecuencias. No obstante, la disminución de radio está limitada por la potencia máxima utilizable por los terminales móviles que podrían interferir con células de grupos próximos (apartado 7.3.1). En el interior de edificios el radio de las células es inferior a 500 m (picoceldas y nanoceldas) con cobertura tridimensional, en zonas urbanas y en rutas con tráfico elevado (autovías, autopistas, etc.) el radio de las células está entre 500 m y 5 km (microcélulas), mientras que en las zonas rurales puede ser superior a los 5 km (macrocélulas).



N(1 a 7): Subbanda de frecuencia (N=B/7)
B: Ancho de banda del sistema celular (MHz)

FIGURA 7.7. Cobertura poligonal de un sistema celular con diferentes radios (pico, micro y macrocélulas), y asignación de subbandas de frecuencia.

En sistemas celulares que deben soportar elevadas cargas de tráfico existe la alternativa de asignación aleatoria de subbandas de frecuencia a las células en función del tráfico. Como ejemplo de dimensionamiento de un sistema celular (apartado 7.4.2), se realiza un cálculo aproximado del número de usuarios por célula en función de los tipos de multiplexación y modulación utilizados. Se considera que el sistema móvil celular dispone de un ancho de banda de 12,5 MHz (B) para cada sentido de comunicación (transmisión y recepción) entre las estaciones móviles y las estaciones base, y que cada canal de voz ocupa 30 kHz. De esta forma cada célula dispone de una subbanda de 1,8 MHz (12,5 MHz/7) lo que permite atender a 60 usuarios por célula (1,8 MHz/30 kHz) con multiplexación FDM (apartado 3.7.2). Si se utiliza la TDM con una modulación digital que permite canales de voz de 2,4 kHz, se consigue introducir a 12 usuarios en cada canal de 30 kHz (30 kHz/2,4 kHz) lo que permite atender a 720 usuarios por célula (12x60). Con métodos de multiplexación CDMA cada célula puede disponer en teoría de los 12,5 MHz, puesto que no se necesita la reutilización de frecuencias, lo que permite 1900 usuarios por célula con canales de 6,6 kHz (12,5 MHz/6,6 kHz), y como cada usuario habla la mitad del tiempo, puesto que estadísticamente el 50% del tiempo de conversación es silencio, en un instante dado pueden hablar simultáneamente el doble alcanzando 3800 usuarios por célula (21x1900).

Dado que normalmente en un sistema celular cada estación móvil soporta a un sólo usuario (el conductor de un automóvil, una persona en movimiento, etc.), ésta se identifica con el terminal del usuario que puede ser de tres tipos según su grado de autonomía (número de horas de funcionamiento sin recarga):

- Terminal móvil

  Cuando no puede retirarse de un vehículo y está permanentemente conectado a la batería del mismo.

- Terminal portátil

  Tiene un peso inferior a los 300 gramos y por sus dimensiones puede guardarse en un bolsillo, con una autonomía aproximada de 90 minutos de conversación o bien de 20h en espera (baterías Níquel-Cadmio). Está destinado a la telefonía personal.

- Terminal transportable

  Tiene mayor peso y dimensiones que el portátil (inferior a un maletín) y su autonomía es mayor. Se utiliza en aplicaciones que no disponen de

suministro continuo de energía como en la construcción, barcos de río o costa, emergencias (atascos de tráfico rodado, catástrofes, etc.).

En los sistemas celulares las estaciones de control realizan las funciones de conmutación de llamadas en curso (*handoff* o *handover*), así como de localización de terminales móviles para llamadas entrantes (*roaming*). Para facilitar la tarea de conmutación de células a la estación de control, un método típico consiste en la emisión continua por el canal de señalización de una señal piloto, no detectable por la estación base, al terminal móvil, que éste a su vez retransmite en sentido contrario. La estación base monitoriza el nivel de la señal piloto recibida, y cuando detecta un descenso por debajo de un cierto valor solicita a la estación de control el cambio de célula. La estación de control encarga a las estaciones base de las células adyacentes a la del móvil la medida del nivel de la señal piloto, y selecciona la célula que haya recibido el mayor nivel. A continuación ordena a la nueva estación base la asignación de un canal y a la anterior que indique la conmutación del canal al terminal móvil.

En circunstancias ideales este proceso es transparente al usuario salvo un silencio de varios milisegundos, pero se producen dificultades cuando existe igualdad en el nivel alcanzado por dos o más células (efecto de "peloteo" o "ping-pong" entre estaciones base), o bien cuando por fenómenos de propagación (multicamino) que refuerzan la señal piloto lejos del móvil, se asigna una célula muy distante (primero se oye una conversación en el canal anterior y luego se corta la comunicación por desvanecimiento del canal nuevo). La superación de estos efectos se consigue mediante un control preciso de la potencia emitida por el terminal móvil desde la estación base, lo que además de reducir la interferencia entre canales mejora la autonomía del terminal móvil (apartado 7.3.2).

La localización de terminales móviles con llamadas entrantes (*roaming*) se obtiene mediante el canal de control existente en cada célula, que es utilizado por la estación base para la supervisión de los terminales móviles de su entorno.

El sistema celular detecta el paso del terminal móvil de una célula a otra, lo que permite el seguimiento del mismo por el centro de conmutación. De esta forma, cuando se produce una llamada desde la red fija o el sistema móvil hacia un terminal móvil, el centro de conmutación puede encaminar la llamada a través de la estación de control correspondiente a la estación base de la célula en la que en ese instante se encuentra el móvil.

7.8.2. Normas internacionales

Son normas de radio móviles de uso público para telefonía digital cuyas redes presentan una cobertura internacional similar a las cableadas de área extensa (WAN) y, que además de prestar el servicio telefónico convencional, están destinadas a competir en el mercado del acceso a Internet y otros servicios de comunicación de datos, a través de los teléfonos móviles de los usuarios particulares. Desde la década de los años 70 del siglo XX estas redes han experimentado una gran evolución desde la primera generación completamente analógica (1G), a la segunda (2G) completamente digital, hasta la tercera (3G) con servicios de transmisión de datos añadidos. La evolución hacia la cuarta generación (4G) tiene por objetivo ofrecer los mismos servicios con idéntica calidad a los usuarios móviles de WWAN que los cableadas (WAN).

El "acceso al medio" es del tipo determinista por asignación fija, y sus métodos están basados en las técnicas de multiplexación TDM, FDM y CDM, recibiendo por ello los nombres específicos de TDMA, FDMA y CDMA (Fig. 7.8). En Europa la norma de telefonía móvil digital de segunda generación (2G) se denomina GSM, mientras que en USA coexisten las normas IS (*Intermediate Standard*) y GSM. Las alternativas de la norma IS son: IS-95 con método de acceso CDMA y las IS-54 e IS-136 con método de acceso TDMA. En Europa la norma GSM emplea el método de acceso TDMA y utiliza para transmitir las frecuencias de 890 MHz a 915 MHz y de 1,71 GHz a 1,785 GHz, mientras que para recepción emplea respectivamente las de 935 MHz a 960 MHz y de 1,805 GHz a 1,88 GHz. En USA la norma GSM utiliza las bandas PCS (*Personal Communication Services*) que en transmisión presentan las frecuencias de 1,85GHz a 1,91GHz y en recepción de 1,93 GHz a 1,99 GHz.

La norma de telefonía móvil digital norteamericana IS-136 es idéntica a la IS-54, salvo en las bandas de frecuencia utilizadas. Mientras IS-54 utiliza las frecuencias desde 824 MHz a 849 MHz para transmitir, y desde 869 MHz a 894 MHz para recibir, la norma IS-136 utiliza las mismas bandas PCS que GSM. La norma IS-95 utiliza las mismas frecuencias de transmisión y recepción que IS-54. El primer servicio de comunicación de datos ofrecido a través del teléfono móvil de segunda generación (2G) es el denominado de "mensajes cortos" (SMS, *Short Message Service*), cuya velocidad es de 9,6 kbps. La necesidad de aumentar la velocidad de los datos a través del teléfono móvil ha llevado a la aparición de normas internacionales, propuestas tanto por el Instituto Europeo de Normas para Telecomunicaciones (ETSI) como por la Asociación de la Industria de Telecomunicaciones en USA (TIA), para las redes celulares avanzadas de segunda generación (2,5G) y de tercera generación (3G).

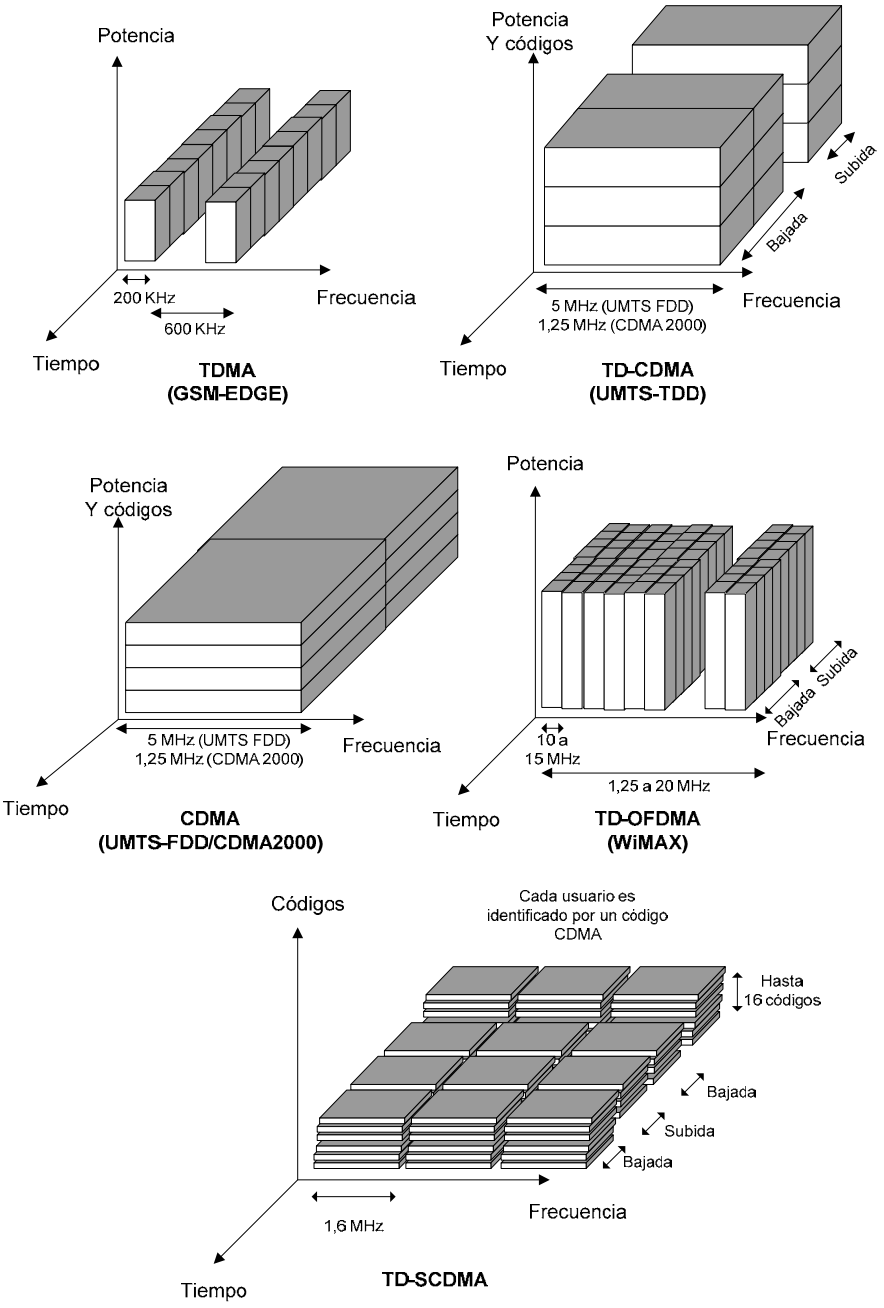


FIGURA 7.8. Tipos de multiplexación en redes celulares de 3ª Generación (3G).

La denominación de algunas de las normas son las siguientes:

- Datos por circuitos conmutados de alta velocidad (HSCSD, *High Speed Circuit Switched Data*)
- Norma intermedia-95b (IS-95b)
- Servicio general de paquetes de radio (GPRS, *General Packet Radio Service*)
- Acceso múltiple por división de código 2000 (CDMA2000)
- Velocidad de transmisión de datos mejorada para la evolución de GSM (EDGE, *Enhanced Data Rates for GSM Evolution*)
- Sistema universal de telecomunicaciones móviles (UMTS, *Universal Mobile Telecommunications System*)

Cada norma de las mencionadas se distingue por los siguientes conceptos: la institución que la propone, el método de acceso empleado, la velocidad máxima alcanzable, las velocidades usuario-red y red-usuario, el tipo de encaminamiento por conmutación y si admite conexión a WLAN. A continuación se resumen sus características por el orden indicado (Fig. 7.9):

- HSCSD: ETSI para GSM, TDMA, 38,4 kbps a 57,6 kbps, 28,8 kbps, Circuitos
- IS-95b: TIA para IS-95a, CDMA, 64 kbps, 32 kbps, 48 kbps, Circuitos
- GPRS: ETSI/TIA para GSM/IS-136, TDMA, 115 kbps, 22 kbps, 44 kbps, Paquetes
- CDMA2000: TIA para IS-95a, CDMA, 2 Mbps, 100 kbps, 384 kbps, Paquetes, WLAN
- EDGE: ETSI/TIA para GSM/IS-136, TDMA, 384 kbps, 40 kbps, 100 kbps, Paquetes
- UMTS: ETSI para GSM, WCDMA, 384 kbps a 2 Mbps, 100 kbps, 384 kbps, Paquetes, WLAN

Las normas de tercera generación (3G) para telefonía móvil celular, como UMTS, están orientadas a facilitar aplicaciones multimedia de tipo personal, tales como la oficina virtual (con servicios de fax, correo electrónico, vídeo, etc.) y la navegación en tiempo real para vehículos. Para ello ofrecen dos velocidades de transmisión: 2 Mbps en estático y 384 kbps en movimiento.

Los teléfonos móviles utilizados son unos terminales con capacidades mejoradas de escritura y visualización, como si fuesen "ordenadores de bolsillo". El acceso a Internet desde un terminal móvil, además de que la red permita velocidades adecuadas para la transferencia de información, debe incorporar los protocolos de la arquitectura de la red adecuados. El "protocolo de aplicaciones inalámbricas" (WAP), representa el esfuerzo de normalización realizado para soportar aplicaciones móviles de datos y sus características son las siguientes:

- Independencia del tipo de red móvil
- El terminal móvil debe incorporar un navegador y los protocolos de sesión y transporte
- El formato de página ha de ser escrito en WML para presentar las páginas en el terminal

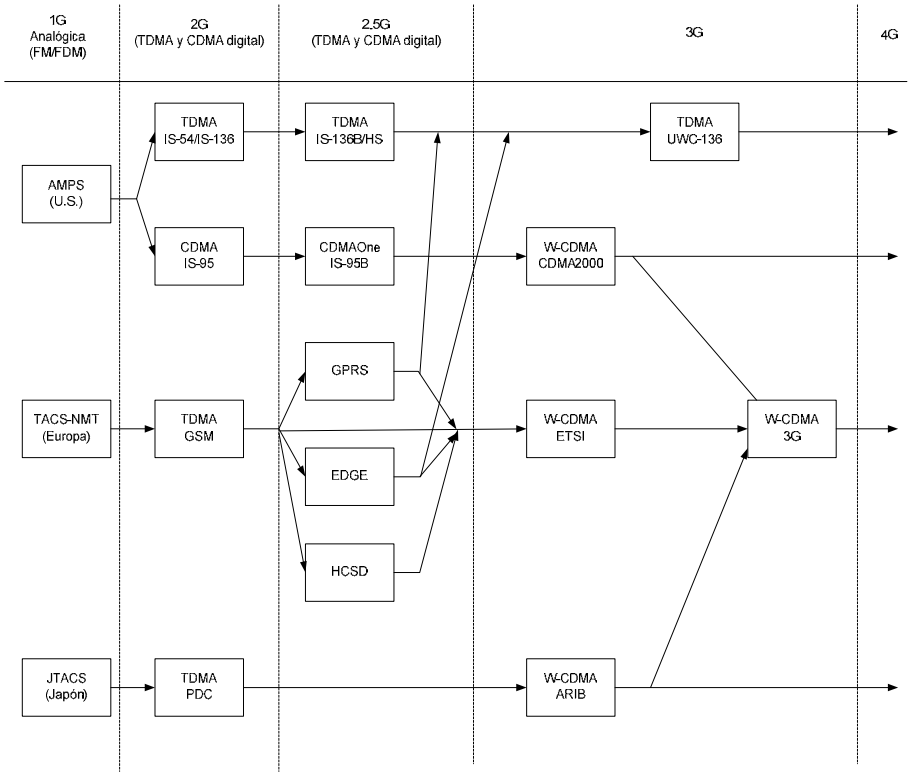


FIGURA 7.9. Las generaciones de la telefonía móvil celular.

El sistema japonés de telefonía móvil digital de segunda generación (PDC, *Personal Digital Communications*), emplea el método de acceso TDMA y utiliza para transmitir las frecuencias de 940 MHz a 956 MHz y de 1,477 GHz a 1,501 GHz, mientras que para recepción emplea respectivamente las de 810 MHz a 826 MHz y de 1,429 GHz a 1,453 GHz. Utiliza un protocolo para Internet móvil diferente de WAP, denominado I-Mode. Para la cuarta generación (4G) se estiman velocidades superiores a 20 Mbps con objeto de proporcionar servicios multimedia móviles en banda ancha, con interoperabilidad entre redes heterogéneas inalámbricas (itinerancia en redes celulares, satélite, WLAN, etc.) y fijas, uso eficiente del espectro, esquemas adaptativos de calidad de servicio (QoS), terminales reconfigurables (por ejemplo teléfonos celulares duales para redes terrestres y de satélites), y servicios IP accesibles a todos los usuarios.

7.9. REDES DE SATÉLITES ESPACIALES

7.9.1. Sistemas celulares basados en satélites

Los sistemas celulares de satélites presentan la particularidad de que debido a su configuración orbital tipo LEO y MEO (apartado 3.6), cada constelación crea una cobertura de todo el planeta cuyas estaciones base son los satélites, por esta razón la función de traspaso (*handoff*) está determinada más por el desplazamiento de la célula que el del terminal móvil. Para dar soporte tanto al sistema como al servicio, existen estaciones terrestres de comunicación y control. Las estaciones terrestres asumen las funciones de centros de conmutación del sistema (CC) con las redes fijas, y de estaciones de control (EC) para la localización de terminales móviles realizando la función de itinerancia (*roaming*) y la función de traspaso de células, además de las funciones de gestión de tráfico y tarificación (apartado 7.4.1).

Los sistemas celulares de satélites existentes se indican a continuación, mencionando el tipo de órbita, número de satélites operativos, método de acceso, banda, años previstos de operación, tipo de encaminamiento, y servicios:

- Inmarsat-P: MEO, fax, datos, radiomensajería y teléfonos digitales personales
- Iridium: LEO (780 km), 66, TDMA, L (1,62 GHz), 5 a 8, ISL, voz (GSM), datos (2,4 kbps), fax, radiomensajería y GNSS

- Globalstar: LEO (1410 km), 48, CDMA, L (2,49 GHz/1,62 GHz), 7,5, Terrestre, voz (GSM), datos (9,6 kbps), fax, radiomensajería y GNSS
- New ICO: MEO (10.390 km), 10, TDMA, S (2,19 GHz/2 GHz), 12, ISL, voz, datos (64 kbps), fax, radiomensajería y GNSS
- Ellipso: LEO (8.040 km), 14, CDMA, 5 a 7, voz y otros
- Odissey: MEO (10.354 km), 12, CDMA, 15, voz (GSM), datos, fax y servicio de mensajes cortos (SMS)
- SkyBridge: LEO (1.469 km), 80, Ku (13,6 GHz/11,7 GHz), Terrestre, lazo local de banda ancha (16 kbps a 2 Mbps/16 kbps a 20 Mbps)
- Teledesic: LEO (1.350 km), 288, TDMA/FDMA, Ka (28,9 GHz/19,1 GHz), 10, ISL, servicios de banda ancha (16 kbps a 2 Mbps/16 kbps a 64 Mbps)

7.9.2. Servicios de navegación global

Los servicios de navegación global (GNSS, *Global Navigation Satellite Services*) se basan en redes de satélites que han sido diseñadas para determinar la posición de un terminal móvil terrestre, marítimo, o aeronáutico, en Longitud (E, W), Latitud (N, S) y Altitud (m), así como el Tiempo horario (s) y su Velocidad de desplazamiento (m/s). Las redes típicas de satélites para servicios GNSS se configuran como una constelación de satélites MEO a 20.000 km de altitud, que entre 20 y 30 se sitúan en 3 o 6 órbitas circulares inclinadas cuyos planos están equidistantes. Para realizar estas medidas el terminal móvil GNSS recibe simultáneamente en su posición las señales de cuatro satélites situados en órbitas inclinadas MEO, y calcula las diferencias de tiempo entre las señales procedentes de los satélites y su propio reloj interno. Las distancias a tres satélites se determinan multiplicando las respectivas diferencias de tiempo por la velocidad de la luz (300.000 km/s), con lo que se obtiene por triangulación la posición del terminal móvil (Fig. 7.10).

Las señales emitidas por los satélites son cadenas de impulsos pseudoaleatorios cuyo código es distinto para cada satélite, con objeto de impedir que cualquier señal radioeléctrica que llegue al terminal móvil pueda ser confundida con una señal GNSS. Otra ventaja del código pseudoaletorio es permitir la amplificación de esas señales en el terminal móvil GNSS sin necesidad de utilizar grandes antenas. Los errores cometidos en la sincronización de los códigos pseudoaleatorios de las señales de los tres satélites involucrados en la medición, y en el terminal mismo, se subsanan con

la señal pseudoaleatoria procedente de un cuarto satélite, que mediante su procesamiento en el terminal permite calcular un factor de ajuste válido para una determinación más precisa de la posición. Hay terminales móviles GNSS capaces de sintonizar a más de cuatro satélites simultáneamente para mejorar la precisión. La medida de velocidad se determina calculando el tiempo que se tarda en pasar de una posición a otra del terminal móvil.

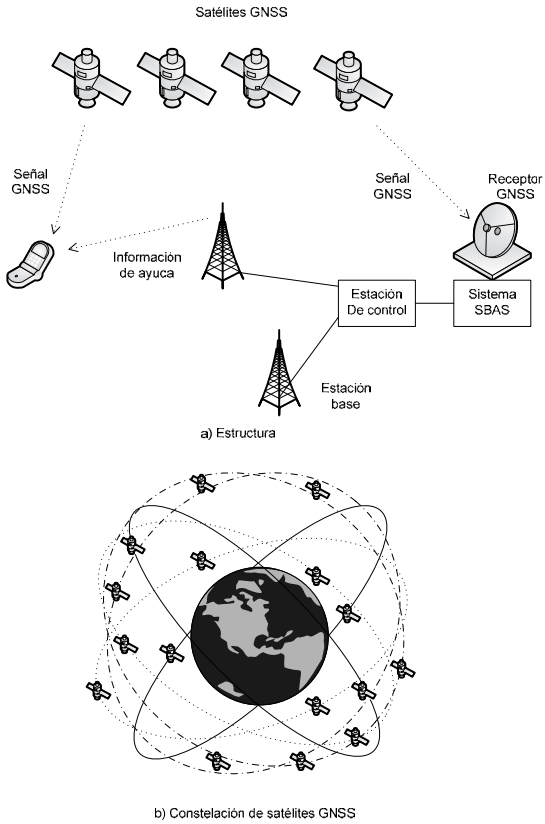


FIGURA 7.10. Representación de un sistema GNSS básico.

Si bien el funcionamiento de un sistema GNSS está basado en las medidas realizadas por el terminal móvil, a partir de las señales pseudoaleatorias de cuatro satélites MEO, existen fuentes de error que limitan la precisión de las medidas.

Por ejemplo, en la navegación aérea se pueden tener precisiones en altitud de 20 m o de 5 m, en función de los errores que puedan ser incorporados a las señales pseudoaleatorias. Las fuentes de error principales en servicios GNSS son:

- Efemérides: predicción de órbita
- Posición relativa (GDOP): entre los satélites y el terminal móvil
- Ionosfera: retardo
- Troposfera: retardo
- Reloj: variaciones de tiempo
- Multitrayecto: rebotes de las señales pseudoaleatorias (apartado 7.3.1)
- Disponibilidad selectiva (SA): errores introducidos a propósito por el suministrador del servicio

El error de "efemérides" procede de las variaciones de la órbita que sufre el satélite como consecuencia de las acciones gravitatorias del sol y de la luna, así como de la presión de la radiación solar. El terminal móvil posee un almanaque programado que contiene la posición de cada satélite del servicio GNSS en el espacio en cada instante. Asimismo el proveedor del servicio controla de forma permanente la altitud, posición y velocidad de cada satélite, y esa información llega al terminal móvil con la señal pseudoaleatoria.

La "posición relativa" (GDOP, *Geometric Dillution of Precision*) entre el terminal y los satélites influye en el error de la posición, por ejemplo si los satélites que participan en la determinación están muy juntos. Esto es debido a que la triangulación se produce con ángulos de muy escasa diferencia entre sí lo que incrementa el margen de error. El terminal móvil debe determinar cuales son los satélites de entre toda la constelación GNSS, que en un instante dado y para una posición determinada son elegibles para recibir las señales pseudoaleatorias que introducen menos error de posición relativa (GDOP).

Los errores introducidos por la ionosfera y la troposfera añaden retardos a las señales procedentes de los satélites, como consecuencia de los cambios de densidad que afectan a su velocidad de propagación. Esto debe corregirse en el terminal móvil GNSS para mejorar la precisión.

Las variaciones de tiempo introducidas por los relojes de los satélites y del terminal presentan errores en la determinación de la posición.

El multitrayecto representa el efecto de las señales reflejadas en las proximidades del terminal móvil, por posibles obstáculos en el entorno del terminal, sobre todo en la navegación terrestre debido a la orografía, que ven alterada su trayectoria y por ello añaden retardo, lo que empeora la precisión.

La "disponibilidad selectiva" (SA, *Service Availability*) es una degradación intencionada en los datos de corrección del reloj de cada satélite, para que en la triangulación de un terminal móvil GNSS se acumulen errores que impidan posicionamientos de elevada precisión. Esta degradación intencionada se justifica desde el punto de vista de que los sistemas GNSS están operados por instituciones militares de USA (sistema GPS) y Rusia (sistema GLONASS), que ante situaciones de emergencia limitan las aplicaciones civiles del servicio GNSS según los requerimientos de la seguridad militar.

Precisamente la acción negativa de la "disponibilidad selectiva" (SA) sobre los usuarios civiles de los servicios GNSS, ha llevado a la aparición de los servicios complementarios de los sistemas GNSS, denominados SBAS (*Satellite-Based Augmentation Services*). Estos servicios se basan en que todos los terminales móviles cerca de un punto geográfico presentan una correlación entre sus errores, y si otro terminal además de recibir las señales GNSS que le llegan a los demás conoce muy bien su posición, puede calcular las correcciones oportunas y enviárselas a los terminales móviles GNSS para aumentar su precisión. Los servicios SBAS utilizan estaciones terrestres y satélites GEO para corregir las fuentes de error, y aumentar la precisión de las medidas de los GNSS, que puede ser de 15 m a 1 m de precisión horizontal y vertical, y de 50 ns de precisión temporal. Como servicios GNSS y SBAS pueden mencionarse:

- GNSS: todos utilizan la banda L (1 GHz a 2 GHz en el segmento espacial), y 20.000 km (MEO)
  - Militares: 24 satélites, 3/6 órbitas inclinadas equidistantes a 60° y 15° del ecuador (GPS)
    - GPS (*Global Positioning System*): iniciado en 1983 por el Departamento de Defensa (DOD) en USA (CDMA)
    - GLONASS (*Global Orbitography Navigation Satellite System*): iniciado en 1993 por el ejército ruso (3 órbitas con 8 satélites, FDMA)
  - Civiles
    - Proyecto GALILEO: 30 satélites en 3 órbitas inclinadas (UE y la ESA, en 2008)

- SBAS
  - EGNOS (*European Geostationary Navigation Overlay System*): satélites de Inmarsat promovido por la UE, Eurocontrol y ESA (*European Space Agency*)
  - WAAS (*Wide Area Augmentation System*): norteamericano
  - MSAS: japonés

Los servicios GNSS, además de su uso en todo tipo de transportes por tierra, mar y aire, también tienen interés en comunicaciones móviles para la sincronización de los sistemas CDMA, y en el posicionamiento de los teléfonos móviles. Otras aplicaciones están en los sectores de la energía, agrícola, obras civiles, emergencias, actividades deportivas, etc.

7.10. LAS REDES INALÁMBRICAS Y LOS BUSES DE CAMPO

7.10.1. Tecnologías inalámbricas en redes industriales

En el apartado 3.9.4 se resumieron las características más importantes de los buses de campo industriales, todos ellos de tipo cableado. Si bien las normas de buses de campo cableados se han diversificado en diferentes sectores de aplicación como domótica e inmótica, control industrial, automoción, aviónica y espacio, con diferentes requerimientos en cuanto a fiabilidad, robustez, coherencia de datos, respuesta en tiempo real, tamaño de paquetes, velocidad, etc., no existe, por el contrario, ninguna norma específica con tecnología inalámbrica en el mercado de las redes industriales. Las normas inalámbricas y móviles analizadas en este capítulo no tienen un objetivo específico en dicho mercado, salvo el caso de ZigBee (apartado 7.6.4) que sólo satisface parcialmente algunas de las características de las aplicaciones en campo, con coberturas muy limitadas.

El uso de enlaces inalámbricos por dispositivos de campo como sensores y actuadores permite una instalación y mantenimiento más flexibles en planta, además de satisfacer directamente las operaciones con robots móviles, reduciendo el problema del incremento de cableado. Sin embargo un sistema de comunicación inalámbrico para operar eficientemente en el entorno industrial debe garantizar entre otras prestaciones, como sus equivalentes cableados: alta

fiabilidad, tiempos de respuesta menores de 10 ms, elevado número de sensores y actuadores y muy bajo consumo de energía. En el entorno industrial además el canal inalámbrico está afectado por el ruido procedente de motores eléctricos, transformadores, y una variedad de equipamiento generador de descargas eléctricas (por ejemplo soldadura eléctrica), cuyo resultado es el aumento de la tasa de error (BER) y la pérdida de paquetes.

Un enfoque conjunto que incorporase todos los requerimientos ya satisfechos por las capas de enlace y aplicación de los buses de campo cableados existentes, y que añadiese a la capa física las contramedidas inalámbricas adecuadas (apartado 7.3.2), llevaría a una nueva investigación y desarrollo de protocolos específicos del modelo OSI, para buses de campo inalámbricos (*wireless fieldbus systems*) especializados, en los que habría que estudiar nuevas métricas y factores de mérito (apartado 7.3.3) que validaran las topologías de referencia (apartado 7.10.2) y el rendimiento de los protocolos frente a las aplicaciones concretas (apartado 7.10.3). En las aplicaciones industriales esta tendencia "radical" es propia de instalaciones experimentales y laboratorios de investigación, y por ello no cabe esperar una aparición inmediata de este tipo de soluciones en el mercado.

La opción "integradora" en el diseño de redes inalámbricas industriales se basa en la coexistencia de redes cableadas específicas que se despliegan en la planta (LAN) o en un ámbito geográfico más amplio (MAN y WAN) bajo normas genéricas, junto con redes inalámbricas y móviles también del mercado de las comunicaciones, que se interconectan para unificar los flujos de información entre los diferentes terminales fijos, móviles e inalámbricos, en un contexto de compartición de datos y aplicaciones al nivel global de una empresa o una corporación. Dentro del modelo OSI pueden ser utilizados los equipamientos de interconexión requeridos como repetidores, puentes, encaminadores y pasarelas, entre otros, cuyo diseño relaciona protocolos fijos e inalámbricos capa por capa. Esta solución presenta el inconveniente de que los equipos pueden resultar caros y complejos, además de que los retardos de conversión de pilas de protocolos hacen inviables las aplicaciones con tiempos de respuesta muy cortos, pero puede ser de gran utilidad en casos concretos en los que sea posible limitar las conversiones de los equipos de red y simplificar los flujos de datos entre los terminales.

Otra opción más simple es la "híbrida", en la que la red inalámbrica actúa como mero soporte de comunicación por radio en el entorno industrial, es decir sustituyendo cables en su área de cobertura, como una típica WPAN, WLAN, WMAN ó WWAN, que se conecta mediante interfaces de radio a nodos de un

bus de campo actuando como puntos de acceso a la red cableada, destinados a ser meros receptores y emisores de información. Esta solución elimina los requerimientos específicos de los buses de campo del canal inalámbrico, y por ello puede ser fácil de incorporar a las instalaciones industriales, en aquellos casos en que esa separación sea viable por el tipo de aplicaciones a satisfacer (por ejemplo en aplicaciones distribuidas de monitorización ambiental).

En cualquiera de las tres opciones de diseño e implantación de las tecnologías inalámbricas en las redes industriales, radical, integradora e híbrida, se necesitan herramientas de desarrollo para la planificación, configuración y mantenimiento de dichas redes. En el apartado 7.4 se presenta esta problemática desde un punto de vista general y se analiza un caso concreto.

7.10.2. Redes en malla (*Mesh*)

Son redes también denominadas: "cooperativas", "autorreparables" (*selfhealing*), "multisalto" (*multihop*), y MANET (*Mobile Ad-hoc NETworks*). Este tipo de redes están constituidas por colecciones de nodos fijos y móviles conectados por radio, los cuales libremente y de forma dinámica se autoconfiguran en topologías arbitrarias temporales, permitiendo una conexión permanente en áreas sin infraestructura (*Ad-hoc*). Ejemplos de este tipo de redes se encuentran raramente en aplicaciones comerciales (apartados 7.5.1 y 7.6.4), pero son habituales en misiones especializadas como recuperación de desastres en áreas sometidas a catástrofes de todo tipo, en sistemas tácticos militares (por ejemplo: operaciones conjuntas en tierra, mar y aire), y en aplicaciones experimentales diversas (astronomía, biosfera, climatología, ecología, instrumentación distribuida, etc.).

No obstante existe un enorme interés por satisfacer aplicaciones comerciales mediante este tipo de redes tanto en entornos de exterior para comunidades de tipo rural con infraestructuras escasas, o de interior en servicios de bajo coste, y en general en aquellas aplicaciones especializadas cuya red está creada por los propios terminales inalámbricos y móviles para servicios tales como: sistemas de transporte inteligentes (ITS, *Intelligent Transportation Systems*); acceso público a Internet por entidades independientes (no ISPs), y proveedores de servicio (ISP, *Internet Service Providers*) competitivos, en los mercados residencial, de negocios e itinerancia, en áreas escasamente pobladas; seguridad en entidades públicas (policía, bomberos, protección civil, emergencias médicas, monitorización ambiental, etc.); telemetría con gran densidad de captadores de medidas en campo para la industria y la investigación, etc.

El desarrollo de los protocolos adecuados para estas redes que son muy sensibles a la aplicación, dista mucho de estar unificado, sobre todo en la capa de acceso al medio (MAC, *Medium Access Control*), debido a que cada terminal debe actuar como un encaminador (*router*) de comunicaciones, lo que limita la escalabilidad de la red por degradación del rendimiento y aumento del retardo a medida que se extiende el diámetro de la red (cobertura). En cuanto a la capa física (PHY) se suele recurrir a las normas para WLAN y sobre todo WPAN ya existentes (apartado 7.6.4).

Dada la necesidad de estándares abiertos en tecnologías de radio para facilitar las economías de escala, reduciendo los costes de los equipos y asegurando la interoperabilidad, se han desarrollado normas específicas de redes en malla (WPAN, WLAN y WMAN) para las capas MAC y PHY tales como: IEEE 802.15.5, IEEE 802.11s (apartado 7.5.2), IEEE 802.16a (apartado 7.7.3), e IEEE 802.20 (apartado 7.7.4). Siguen estando abiertos a la investigación protocolos capaces de proporcionar: interfaces de radio fiables y de alta capacidad entre terminales, funciones de red escalables (independientes del número de usuarios y rutas de comunicación) y, gestión eficaz de los recursos del sistema de tal forma que todos los usuarios mantengan la calidad de sus prestaciones con independencia del tráfico, las condiciones de los canales, la contienda por el acceso y su ubicación geográfica.

7.10.3. Ingeniería de protocolos de capa cruzada (*Cross-Layer*)

El diseño de protocolos en las redes inalámbricas y móviles ha sido paralelo al desarrollo de sus equivalentes cableadas, siguiendo el paradigma del modelo OSI de siete capas apiladas, en el cual cada capa de la pila es diseñada y operada independientemente con interfaces estáticas entre su capa superior y su capa inferior, sin considerar las restricciones impuestas a la red concreta ni el tipo de aplicaciones. La razón de esta metodología de diseño se basaba en la explotación de la modularidad del sistema de comunicación global, lo que simplificaba el diseño de equipos y su interoperabilidad. Sin embargo las características especiales de las redes de radio tales como la movilidad de los usuarios, el fallo frecuente de los enlaces debidos a la propagación, la limitada capacidad de los canales, y las restricciones en el suministro de energía (baterías) y los recursos de procesamiento de los terminales (número y tamaño de los chips), junto con las diferentes demandas de servicio (QoS) en las aplicaciones, imponen nuevos desafíos en el diseño de los protocolos de las capas de enlace, encaminamiento y transporte en redes de radio de alta velocidad.

La metodología de diseño de protocolos de "capa cruzada" supone la consideración dinámica de las interacciones entre los protocolos de las diferentes capas, lo que conlleva una gran complejidad debida a la aparición de numerosos parámetros y a la naturaleza no lineal de las máquinas de estado de los protocolos. No obstante se considera que una cuidadosa explotación de algunas interacciones entre protocolos de diferentes capas (*cross-layer*), puede aportar prestaciones eficientes en los protocolos de la capa física que ayuden a mejorar los protocolos de la capa de aplicación en diferentes escenarios de red (por ejemplo: celular, malla, etc.). Un caso típico es el de las redes en malla (apartado 7.10.2) donde las variaciones en movilidad de terminales, densidad de usuarios y dimensiones de red hacen que los requerimientos de calidad de servicio para las aplicaciones sean difíciles de satisfacer.

La ingeniería de protocolos de capa cruzada es un área de investigación emergente, de naturaleza interdisciplinar que implica la colaboración entre ámbitos de trabajo como: procesado de señal, modulación y codificación adaptativas, modelado de canales de radio, modelado de tráfico, teoría de colas, diseño de protocolos y técnicas de optimización.

7.11. CONCLUSIONES

La ausencia de normas específicas para redes inalámbricas en las aplicaciones industriales, exige el conocimiento de las normas existentes en las aplicaciones inalámbricas y móviles genéricas, con el objetivo de poder adaptarlas convenientemente a las demandas el entorno industrial, o de evaluar convenientemente sus prestaciones para satisfacer los requerimientos en campo. Este conocimiento se aporta mediante la descripción de las normas para redes de radio con diferentes áreas de cobertura y servicios, justificando las diferentes capas de protocolos con una previa profundización en los fundamentos de la propagación electromagnética que caracterizan a los canales de comunicación.

7.12. BIBLIOGRAFÍA

P. Mariño. Las nuevas tecnologías de la información y las comunicaciones. Ed. FSG, 2001.

P. Mariño. Las comunicaciones en la empresa: Normas, redes y servicios. Ed. RA-MA, 2003.

M.G. Moreno y J.L.G. Moreno. Comunicaciones móviles. Ediciones UPC, 2002.

Quobis Networks. WiMAX: la revolución inalámbrica. Ed. Quobis, 2006.

7.13. EVALUACIÓN

7.13.1. Ejercicios resueltos

Ejercicio 7.1

Una operadora estima que la métrica U (bps/Hz/km²) en redes móviles de cuarta generación (4G) alcanzará un valor 1000U (U1) para coberturas de nanoceldas de 10 m de radio, mientras que para microceldas de 100 m de radio en redes móviles de tercera generación (3G) ese mismo parámetro no subió de 8U (U2). Determinar la ganancia experimentada en la métrica de "eficiencia espectral" desde la 3G a la 4G.

Solución

La eficiencia espectral se expresa en "bps/Hz", y por tanto hay que calcularla a partir de la métrica U de cada generación y cobertura.

a) Cálculo de las coberturas de cada tipo de celda en km²:

La superficie de la nanocelda es de: $S1 = \pi . R1^2 = \pi.km^2/10.000$, para 4G

La superficie de la microcelda es de: $S2 = \pi . R2^2 = \pi.km^2/100$, para 3G

b) Cálculo de la "eficiencia espectral" de cada generación y cobertura es, multiplicando por los km² de la cobertura en cada expresión de la métrica U, y haciendo el cociente de ambas:

Ganancia de "eficiencia espectral" (GEF); GEF = U1 x S1 / U2 x S2

GEF = 1000 (bps/Hz).(π/10.000)/8 (bps/Hz).(π/100) = (1/10)/(8/100) = 10/8

Ejercicio 7.2

En la norma IEEE 802.20 (MBWA) para acceso móvil mundial en banda ancha el máximo valor de las frecuencias portadoras (Fp) será de 3,5 GHz, con tasas de bits superiores a 1 Mbps, y velocidades de desplazamiento (V) de los terminales de usuario hasta 250 km/h.

Calcular el máximo desplazamiento espectral (Fd) que experimentarán las portadoras por efecto Doppler, y la máxima frecuencia alcanzable por una portadora en un terminal móvil (Fpmáx).

Solución

El desplazamiento espectral debido al efecto Doppler depende de la velocidad del móvil, de la frecuencia de la portadora y de la velocidad de la luz (C) según la ecuación 7.1:

Fd = V.Fp/C = 250 km.3500 MHz/(3600 s.300.000 km/s) = 810,19Hz

Por lo tanto la máxima frecuencia alcanzable será de:

Fpmáx = 3,5 GHz + Fd

e igualando resulta:

Fpmáx = 3.500.000.810,19 Hz

Ejercicio 7.3

El límite que impone la Comisión Federal de Comunicaciones a la métrica de "densidad espectral de potencia" para la norma UWB en la banda de 3,1 GHz a 10,6 GHz es de 75 nW/MHz. Calcular el valor de dicha métrica en dBm/MHz.

Solución

Para ello hay que tomar el valor de la potencia en nanowatios y pasarlo a dBm, mediante la ecuación 7.2:

dBm =10.log (75 nW/1 mW) = 10.log (75/1.000.000) = 10.(log 75 – 6)

El resultado es

dBm = 10.(1,8751 – 6) = -41,25

y por tanto la "densidad espectral de potencia" pedida equivale a;

-41,25 dBm/MHz

Ejercicio 7.4

En una red celular con un ancho de banda de 20 MHz y grupos de 7 células, cada canal de voz ocupa 30 kHz. Calcular el número de usuarios por célula en los siguientes casos:

a) Se utiliza multiplexación FDM.

b) Se cambia a multiplexación TDM con una modulación digital que permite canales de voz de 2,4 kHz.

c) Se aplica una utilización estadística de los silencios en un 50%.

Solución

a) Todo el ancho de banda se reparte entre cada grupo de 7 células, disponiendo cada célula de una subbanda de:

20 MHz/7 = 2,85 MHz

lo que permite atender a 95 usuarios (2850 kHz/30 kHz) en cada célula.

b) Al reducirse el canal a 2,4 kHz el número de usuarios se multiplica por 12 (30 kHz/2,4 kHz) y resulta un total de 1140 usuarios por célula (95 x 12).

c) Si se aprovechan los silencios al 50% pueden hablar simultáneamente el doble, resultando por tanto un total de 2280 usuarios por célula (2 x 1140).

7.13.2. Evaluación objetiva

1. El GPS es:
   A) Un método de modulación
   B) Un sistema GNSS
   C) Un tipo de telefonía celular digital
   D) Una red de área local

2. Las redes de acceso celular son del tipo:
   A) LAN
   B) RDSI
   C) xDSL
   D) WWAN

3. La red de servicios de distribución multicanal multipunto (MMDS):
   A) Es de tipo WWAN
   B) Sus frecuencias están en el margen de las microondas
   C) Su cobertura es de 30 km a 70 km
   D) Cumple la norma WiFi

4. La norma WiMAX está basada en:
   A) IEEE 802.15.4
   B) IEEE 802.16
   C) FDDI II
   D) Bluetooth

5. En un servicio GNSS el error debido a variaciones de la órbita se denomina:
   A) Efemérides
   B) Posición relativa
   C) Multitrayecto
   D) Disponibilidad selectiva

6. Un servicio SBAS de navegación global es:
   A) WAAS
   B) GLONASS
   C) GSM
   D) GALILEO

7. La función de traspaso (*handoff* ó *handover*) en una red móvil significa que el teléfono móvil:
   A) Cambia de célula en el área de cobertura durante una llamada
   B) Ha de ser localizado para recibir una llamada entrante
   C) Debe conectarse a un teléfono de la RTC
   D) Recibe interferencias cocanal

8. La función de itinerancia (*roaming*) en una red móvil significa que el teléfono móvil:
   A) Recibe interferencias cocanal
   B) Cambia de célula en el área de cobertura durante una llamada
   C) Ha de ser localizado para recibir una llamada entrante
   D) Debe conectarse a un teléfono de la RTC

9. Un sistema celular de satélites es:
   A) GLONASS
   B) ZigBee
   C) Globalstar
   D) VSAT

10. En una WPAN de topología provisional (*Ad hoc*) es cierto que:
   A) Puede ser de tipo temporal
   B) Cumple la norma GSM
   C) Tiene puntos de acceso a otras LAN
   D) Su cobertura es superior a los 300m

TEMA 8

SEGURIDAD EN LAS COMUNICACIONES. CALIDAD DEL SERVICIO

8.1. Introducción y objetivos
8.2. Factores de inseguridad en sistemas y en dispositivos
8.3. Soluciones actuales en sistemas y dispositivos de seguridad
8.4. Política de seguridad para redes
8.5. Defensas no criptográficas
8.6. Defensas criptográficas
8.7. Seguridad en protocolos industriales
8.8. Introducción al concepto de calidad del servicio en redes
8.9. Aproximaciones estándar a la calidad del servicio en redes
8.10. Los encaminadores y la calidad del servicio
8.11. Conclusiones
8.12. Bibliografía
8.13. Evaluación

8.1. INTRODUCCIÓN Y OBJETIVOS

En este capítulo se expondrán los conocimientos básicos sobre dos aspectos de las redes que, de ser poco importantes durante años, han pasado a ser fundamentales para cualquier buen profesional de redes: la seguridad y la calidad de servicio.

En el momento actual no se puede hacer un análisis completo y realista de ninguna organización en la que las redes sean una parte esencial sin tener en cuenta los peligros desde el punto de vista de seguridad a los que están expuestas tales redes. Va más allá de esta publicación analizar cuáles pueden ser las posibles motivaciones de los atacantes, pero, sean las que sean, se debe estar dispuesto a afrontar la posible pérdida de datos, revelación de información sensible, caídas de la red interrumpiendo el servicio, en general todos aquellos problemas de seguridad que no tienen que ver con desastres naturales y que tienen un origen humano. El análisis de todos estos problemas de seguridad será el objetivo de los primeros apartados de este capítulo.

La importancia relativa dentro de la organización de los datos gestionados tanto en sistemas como en su envío a través de redes ha crecido hasta tal nivel que no permite obviar desafíos a la seguridad de tales datos sin tener que prepararse para afrontar grandes pérdidas. En este sentido se deben conocer cuáles son los principales peligros reales a los que se está expuesto, físicos, lógicos, humanos y organizativos. De igual manera un profesional de las redes debe conocer también cuáles son las principales estrategias de defensa que puede utilizar. Tales estrategias y sus soluciones serán el objetivo del siguiente apartado. Hay que conocer en qué casos concretos es más inteligente utilizar cada una de las posibles defensas, tanto las organizativas como las técnicas.

Como pieza clave dentro de todas las estrategias, se debe estar dispuesto a utilizar una política de seguridad que articule todas las defensas utilizadas y que se convierta en el “motor” de todas las medidas de seguridad. Éste será el contenido del siguiente apartado del capítulo, con el que se dará una visión de conjunto de las medidas de defensa informática. Como parte final del análisis dedicado a los problemas de inseguridad informática en redes se expondrán, en dos apartados, las principales características no criptográficas que exhiben los dispositivos de seguridad de redes y se describirán los fundamentos y protocolos de los sistemas criptográficos en la informática de las comunicaciones en redes.

De igual manera un profesional cualificado debe conocer cuáles son las limitaciones de las redes con las que trabaja, debe saber cómo obtener el máximo de prestaciones para los recursos que tiene en sus redes. Hasta fechas recientes había poco que decir sobre cómo mejorar el servicio ofrecido en redes. Era suficiente con que la red estuviera funcionando correctamente 24 horas al día. Hoy, y cada vez en más organizaciones, esto ya no es suficiente. Es necesario que la red se adapte a las necesidades cambiantes de cada una de las aplicaciones que la usan como medio de transmisión. Esto quiere decir que habrá aplicaciones que necesitarán más ancho de banda en momentos determinados u otras para las que el retardo en la señal deba ser menor de una cierta cantidad para poder darán servicio como el ofrecido a los clientes de tales aplicaciones. Tales ideas forman parte del apartado de este capitulo en el que se trata del concepto actual de Calidad de Servicio en redes.

Afortunadamente es éste un problema que ya ha encontrado soluciones técnicas reales y se dispone de estándares suficientemente probados y aceptados que se deben conocer desde un punto de vista general, en sus objetivos. Éste será el contenido del siguiente apartado del capítulo en el que se describirán estos estándares de manera comparada.

Por último, como contenido del último apartado, se debe conocer cuáles son las armas de las que se dispone cuando se quiere implantar una cierta política de calidad del servicio en redes. Son los dispositivos de comunicaciones básicos de una red de redes (los conmutadores y los encaminadotes) los utilizados para implantar tales políticas, mediante configuraciones de sus sistemas que lo permiten

8.2. FACTORES DE INSEGURIDAD EN SISTEMAS Y DISPOSITIVOS

Una red está compuesta de sistemas y dispositivos de comunicaciones. Desde el punto de vista de los objetivos de la organización son los primeros los que más cuentan. Ellos gestionan el almacenamiento de los datos, el control de acceso a tales datos y el envió y recepción de los datos según sea necesario.

Por otro lado, los usuarios utilizan diferentes aplicaciones que se ejecutan gracias al sistema operativo del ordenador correspondiente. Tanto los sistemas como las aplicaciones exhiben problemas de seguridad, con lo que hay que tratarlos ambos.

Se puede hacer una división no completamente estricta entre los sistemas que se pueden encontrar en redes atendiendo a su importancia relativa y dividirlos en sistemas servidores y sistemas clientes. Es importante darse cuenta de que, en general, es mucho más relevante cuidar la seguridad en los servidores que en los clientes, aunque sin descuidar la importancia de los clientes que pueden, cada vez más, contener datos sensibles de la organización.

Asimismo sería iluso suponer que en nivel de detalle los sistemas operativos son todos iguales. En este apartado, no obstante, se analizan aquellos problemas graves que lo son para cualquier sistema típico en una red y, siendo más explícitos, esto cubriría cualquier sistema de la familia Windows, cualquier sistema UNIX (HP-UX o Tru64-UNIX de HP, AIX de IBM, Solaris de Sun, Red-Hat Linux, SUSE Linux, etc.) y sistemas algo menos frecuentes como Novell Netware o el sistema OpenVMS de HP.

De igual manera se analizarán los problemas de seguridad típicos asociados con las aplicaciones más generalizadas como son todas las que dependen de servicios HTTP o web, todas las que tienen que ver con el uso del correo electrónico, las que permiten el acceso remoto a sistemas o aplicaciones y las que tienen que ver con sistemas multimedia.

Con el objeto de hacer más sencilla la exposición se van a catalogar los problemas de la siguiente manera:

- Problemas de seguridad física.
- Problemas generales de seguridad en sistemas operativos y aplicaciones.
- Problemas comunes de seguridad en sistemas operativos.
- Problemas comunes de seguridad en aplicaciones.

8.2.1. Problemas de seguridad física

Los problemas de seguridad física más comunes para sistemas son diferentes en el caso de servidores de aplicaciones o de datos y en el caso de equipos clientes.

Es muy importante darse cuenta de los problemas de seguridad que pueden presentarse cuando los **servidores de la red** no están correctamente ubicados desde el punto de vista físico, pues el resultado de un ataque en estas circunstancias puede ser desastroso.

Los principales problemas que pueden encontrarse son:

- La **destrucción total o parcial del equipo**, lo que puede ser muy costoso desde el punto de vista puramente económico de la máquina o, peor aún, si se piensa en los datos de cualquier tipo que podía albergar y a los que, muy probablemente, estaba proporcionando acceso remoto.

- El **acceso no permitido al sistema operativo** y. como consecuencia, a los datos y aplicaciones del sistema. Prácticamente todos los sistemas disponen de un mecanismo de arranque "mínimo" diseñado para solucionar problemas de gestión poco habituales pero probables, como el olvido de la contraseña del administrador del sistema o el cambio de un parámetro del sistema especialmente delicado. Este mecanismo pasa por introducir una secuencia especial en el arranque físico de la máquina desde su consola y esto precisa acceso físico al equipo. Una vez arrancado este mini sistema operativo se obtiene acceso al sistema sin petición de contraseña ninguna. Si la máquina no está en un entorno físico controlado y restringido, cualquiera con los conocimientos mínimos puede acceder a este entorno y acceder al sistema completo como administrador sin conocer su contraseña. A partir de ese momento es dueño (ilícitamente) del sistema, con consecuencias muy comprometedoras para el propio sistema.

  En el caso de los **sistemas clientes** los problemas comunes pueden parecer, en principio, menos devastadores pero es necesario considerarlos igualmente. Los principales problemas son:

- **Todo tipo de ataques** a la seguridad de servidores, clientes y dispositivos de comunicación, originados en el cliente y puestos en marcha a conciencia, debido a la mala voluntad del propietario del equipo, lo cual es posible debido a la funcionalidad completa de los sistemas operativos típicos de cliente.

- La **destrucción del equipo** y de la información que contenía por falta de seguridad física del mismo y por razones similares a las ya citadas.

- El **acceso no permitido** a datos sensibles que pueden residir en el equipo cliente, conseguido mediante técnicas como la citada anteriormente o por otras más sencillas, como el robo de contraseñas de cuentas del equipo cliente.

- El **acceso no autorizado a** aplicaciones y datos en el equipo **servidor**, desde el equipo cliente, obtenido con las mismas técnicas ya citadas.

- La **inyección en la red de diferentes tipos de virus** o troyanos a partir del equipo cliente, debido a la mala configuración de seguridad del equipo al que le falta, por ejemplo, un buen antivirus actualizado, lo que permite la inserción primero en el equipo y, después, en los otros equipos de la red, de distintos tipos de programas que se aprovechan de esta situación.

Muchas de estas circunstancias se agravan en el caso de portátiles que pueden moverse dentro de la oficina o fuera de ella, resultando más sencillo inyectarles software mal intencionado y resultando más sencillo su uso malintencionado al crecer la probabilidad de ser robados.

8.2.2. Problemas generales de seguridad en sistemas operativos y aplicaciones

Sin entrar en el detalle de la estructura interna de los sistemas operativos baste decir que los actuales consiguen unas funcionalidades cada vez más interesantes a base de implementar una serie de características complejas:

- Son cada vez más extensos, ocupando millones de líneas de código.
- Son cada vez más modulares, permitiendo realizar distintos tipos de operaciones a base de interacciones entre diferentes módulos.
- Permiten cada vez más interconexiones en red para facilitar las tareas de la informática distribuida.

Estas características los hacen especialmente difíciles de analizar desde el punto de vista de seguridad, provocando diferentes tipos de problemas.

Por otro lado las aplicaciones actuales han seguido el mismo camino y exhiben características similares salvo la de la extensión relativa: no son tan extensas como un sistema operativo pero son mucho mayores que sus versiones anteriores. Además cada vez más son el objetivo de la mayor parte de las herramientas de ataque desarrolladas. Hay que citar, especialmente para el caso de las aplicaciones, la implementación de los protocolos de comunicación citados en diferentes capítulos anteriores, que provoca en muchos casos la creación de posibles avenidas de ataques en la red.

En realidad, en todos estos casos se está analizando el mismo tipo de **objetos software** que, con ligeras diferencias, exhibe el mismo tipo de estructuras y tipos de problemas de seguridad comunes semejantes. Con el objeto de analizarlos más claramente se pueden clasificar como:

- Problemas debidos a la codificación incorrecta del software.
- Problemas debidos a ataques de ingeniería social

8.2.2.1. Mala codificación

Según diferentes estudios recientes[1] se estima que existen entre 6 y 16 errores de codificación por cada 1000 líneas de código software en una aplicación o sistema general. Esto es así para sistemas y aplicaciones y lleva a miles de posibles problemas de funcionamiento. Algunos de ellos (el número dependerá del énfasis puesto en la seguridad en el diseño del software) son *bugs* de seguridad o vulnerabilidades, que pueden ser aprovechadas para lanzar diferentes tipos de ataques contra la pieza de software que los contiene.

En el caso de una cualquiera de ellas se pueden separar diferentes etapas de su historia:

1. **Descubrimiento de la vulnerabilidad**. Esta fase ha mejorado con los años; hoy en día puede ser descubierta por los creadores del software o por cualquier investigador. Antes estos últimos podían ser acusados, aún usando medios discretos para indicar el error a los fabricantes, de distintos tipos de delitos. La industria ha evolucionado y admite los avisos, aunque no le guste que se hagan públicos.

2. **Anuncio de la vulnerabilidad**. Esta fase es problemática. Si el descubrimiento se ha hecho por cauces "lícitos" es lógico que se haga el anuncio por parte del fabricante pero, en muchos casos, no es así.

3. **Creación y distribución del corrector, o "parche" de la vulnerabilidad**. Lo primero a señalar es que este corrector es otro programa software que hay que instalar. Si no se instala, el programa original está expuesto a los ataques desarrollados para aprovechar la

[1] Ver "RFID malware, truth vs. Myth", M. Rieback, B. Crispo y A. Tannenbaum, IEEE Security & Privacy, Julio de 2006.

vulnerabilidad ya conocida y difundida por el propio fabricante. Si se instala hay que tener en cuenta los posibles "efectos laterales" que puede provocar, como, por ejemplo, el que deje de funcionar una aplicación diferente o que el citado "parche" haya de ser "parcheado" más tarde otra vez por tener también problemas de seguridad. Además hay que dedicar un esfuerzo (y un coste) a estar informado y a instalar correctamente la actualización.

No siempre se instalan las actualizaciones aún cuando la vulnerabilidad ha sido popularizada ya, lo que abre un riesgo de exposición considerable al ataque.

Si se tiene en cuenta que todos los sistemas operativos y aplicaciones están expuestos a este tipo de problemas todos los días y que aparecen cada año cientos de vulnerabilidades de cada fabricante, se puede concluir que este problema es quizás el más grave con diferencia, por lo poco que un administrador de sistemas o de redes puede hacer para mitigarlo: estar al tanto de todas las vulnerabilidades y decidir, con mayor o menor precisión, si se instalan las actualizaciones y cuándo y cómo.

8.2.2.2. Ingeniería social

Bajo este nombre se hace referencia a distintos tipos de técnicas que coinciden todas en aprovecharse de la buena voluntad o de la inconsciencia de la gente. Se trata de aprovechar el eslabón más débil de la cadena de seguridad: el factor humano.

Si no se ha dado una formación básica de seguridad adecuada a los empleados de una organización, se dan fácilmente situaciones en las que el atacante, haciéndose pasar por un compañero, un administrador de sistemas, un colaborador externo o un cliente, puede obtener cualquier tipo de información confidencial (números de teléfono, contraseñas de acceso, datos fiscales, etc.) sin más que preguntar. Hay una parte de "teatro" que debe ser ejecutada fríamente, pero que da buenos resultados en general si el receptor de las preguntas no está correctamente aleccionado para identificar claramente a quien hace las preguntas.

En los últimos años se está asistiendo a una derivación peligrosa de este tipo de ataques, conocida como "***phishing***" y sus sucedáneos. En este caso se trata de mensajes de correo electrónico recibidos en la cuenta de cualquiera, en los que, alegando un problema de seguridad de una organización (típicamente un banco), se piden datos personales y privados del receptor del mensaje, como

números de cuenta, contraseñas de acceso remoto, etc. La técnica suele consistir en ir desde la ventana del correo a un enlace proporcionado en el mismo y este enlace nos lleva a una ventana idéntica a la que se espera de nuestro banco, sólo que no lo es, y al introducir los datos requeridos en realidad se proporcionan a los atacantes, que pueden así con facilidad hacerse pasar por el atacante en el sitio web real de la entidad. Se calcula que alrededor del 4% de los ataques tiene éxito, lo que podría parecer un número ridículo hasta que se sabe que son millones los mensajes de este cariz que circulan por Internet. Hasta ahora casi todos estos ataques han sido dirigidos contra la seguridad de un individuo pero su naturaleza permite pensar en la obtención de información sensible de la organización mediante métodos semejantes.

8.2.3. Problemas comunes de seguridad en sistemas operativos

Hasta aquí se ha analizado problemas comunes generales a cualquier pieza de software, pero también los hay particulares. En el caso de los sistemas operativos los más comunes son los debidos a mala administración del sistema. Entre ellos los más significativos son:

- **Configuración insegura de los ficheros de autorización**. Hay típicamente dos formas de acceder a un sistema como usuario: localmente y de manera remota. En el primer caso hace falta acceso físico, en el segundo un cliente accede mediante una aplicación como telnet o ssh al servidor. En cualquiera de los dos se contrasta su información de acceso (típicamente un nombre de cuenta y una contraseña) contra un fichero de autorización de usuarios. Este fichero puede, a su vez, ser local, residir en el sistema al que se pretende acceder, o remoto, residiendo en un servidor de la red. En cualquiera de los casos si este fichero no está correctamente protegido (sólo legible y actualizable por el administrador) o si los nombres y, especialmente, las contraseñas no son seguras, el acceso no autorizado al sistema resulta especialmente sencillo. Por ejemplo no debería estar permitido el tener contraseñas idénticas al nombre de la cuenta de usuario ni cuentas "automáticas" de acceso para usuarios no privilegiados. Es, además, muy importante tener contraseñas seguras para las cuentas de administración. Si el fichero no está correctamente protegido se pueden sufrir ataques de "***cracking***" que, mediante programas que hoy son de libre distribución, permiten obtener las contraseñas de las cuentas. Todo esto es inútil, de todas formas, si no se contempla, asimismo, el posible uso no autorizado en la red de analizadores de protocolos.

- **Configuración insegura de los permisos de sistemas de ficheros**. Hoy en día los sistemas operativos permiten que cada entidad de almacenamiento tenga la protección que dicte nuestra política de seguridad. Esto significa que cada volumen lógico, disco, partición, directorio y fichero se pueda proteger suficientemente. Tanto en el caso de ficheros de gestión (de red, de registro de eventos, de parámetros, de arranque de aplicaciones, etc.) como de ficheros de datos (bases de datos, documentos sensibles de la organización) se pueden aplicar los criterios restrictivos necesarios. Por ejemplo puede haber ficheros que deban estar cifrados y con acceso permitido sólo a un proceso, otros pueden permitir el acceso de lectura de cualquier proceso pero nada más, puede haber discos enteros, de uso de una base de datos, con permisos especiales sólo para un grupo de usuarios. Si esto no se protege adecuadamente, conforme a una política de seguridad, se estará poniendo en peligro los activos de la organización residentes en el equipo.

- **Inexistencia de copias de seguridad**. Debe existir una (o varias) copias de seguridad de los datos si se quiere responder rápidamente a su desaparición. Además, hay que asegurarse de que contienen lo que se cree que contienen probándolas con una cierta periodicidad y hay que asegurarse que residen en un lugar seguro desde el punto de vista de su posible destrucción.

- **Configuración incorrecta de la compartición de ficheros**. Esto es en realidad un caso particular de mala configuración de los sistemas de ficheros pero en el que una configuración incorrecta abre el paso a muchos (y muy peligrosos) tipos de ataques. Se trata de sistemas que usan herramientas que en parte son aplicaciones y en parte son del sistema como el "*Network File System*", NFS, o la compartición de ficheros de Windows de Microsoft. Si su configuración es incorrecta desde el punto de vista de seguridad se "abre" el sistema a distintas avenidas de ataque.

8.2.4. Problemas comunes de seguridad en aplicaciones

Se puede observar una tendencia clara en los últimos años a un mayor número de ataques que aproveche todo tipo de vulnerabilidades o debilidades de

las aplicaciones que se usan en nuestros sistemas[2], ya sean estas locales o remotas. Aunque no se va a hacer una taxonomía completa (disponible en referencias de la bibliografía) entre las más significativas se pueden citar:

- Las debidas a mala implementación de protocolos de comunicaciones.
- Las debidas a debilidades propias de los protocolos.
- Las debidas a mala selección de protocolos o a un uso incorrecto de ellos.
- Las debidas a la configuración incorrecta de las aplicaciones.

8.2.4.1. Implementación incorrecta de protocolos

Los problemas de una incorrecta codificación de los protocolos de comunicación dan lugar a situaciones en las que los ataques pueden ser realmente fáciles de ejecutar. Es simplemente un recordatorio del apartado en el que se ha analizado el problema de la codificación incorrecta del software.

Entre los más significativos aparecen:

- Los ataques de denegación de servicio como el “ping de la muerte” en el que una mala codificación del protocolo ICMP da lugar a un desbordamiento de memoria en el equipo atacado (al que se la hace ping) pudiendo llegar a provocar una caída del equipo.

- Todo tipo de problemas de seguridad relacionados con una incorrecta implementación del protocolo SMTP, responsable del tráfico de mensajes de correo electrónico en las redes. Los programas que lo implementan son muy extensos y, como consecuencia, propensos a tener abundantes vulnerabilidades.

- Los ataques basados en “puertas falsas” (*trap doors*), que son código únicamente conocido por el fabricante del software y que permite el acceso a cualquier sistema o aplicación.

[2] Ver, por ejemplo, el informe del instituto SANS, http://www.sans.org/top20/

8.2.4.2. Diseño peligroso de protocolos y aplicaciones

Cuando se diseñó la versión actual de los protocolos IP no se tuvo en cuenta demasiado la seguridad, lo cual es normal si se piensa que los objetivos del diseño no eran las tareas para las que ahora se utilizan. Esto trae como consecuencia la existencia de diversas "debilidades" que hacen que sean posibles cierto tipo de ataques aún cuando no existan problemas en la implementación. Entre los más comunes se pueden citar:

- Aplicaciones sin control de autenticación como el TFTP (*Trivial File Transfer Protocol*) o el propio SMTP, que hacen relativamente fácil la suplantación de la identidad en la red. En la misma categoría de falta de autenticación, o debilidad en la misma, entran el DNS (*Domain Name System*), responsable de la traducción de nombres de equipo a direcciones IP y pieza básica de gestión en cualquier red, o el protocolo SNMP (*Simple Network Management Protocol*), responsable de los mensajes de gestión de red de las principales aplicaciones que se encargan de la citada gestión. Este problema lo tienen incluso algunos protocolos de encaminamiento como el popular RIP (*Routing Internet Protocol*).

- La naturaleza de los propios mecanismos de conexión en TCP/IP, que permiten ataques como el "*smurf*" en el que se envían pings a la dirección de *broadcast* de la red donde reside la victima, habiendo colocado en el mensaje como dirección IP origen la del equipo victima.

- Los ataques de tipo SYN FLOOD, de denegación de servicio en los que el atacante, usando una dirección IP falsa, se aprovecha el hecho (Fig. 8.1) de que un servidor de una aplicación TCP tiene que seguir el protocolo "*3-step handshake*", citado en el capítulo 4, y no dar por perdida una petición de conexión hasta pasado un cierto tiempo, lo que puede llenar su tabla de "conexiones por establecer" o "embriónicas" rápidamente y hacer imposible una conexión TCP lícita al estar atendiendo únicamente a las ilícitas.

Se deben señalar también todas aquellas aplicaciones que sufren de lo que se podría llamar "efecto facilidad", diseñadas con el objetivo de que su uso sea especialmente sencillo, poniendo en peligro en muchos casos la integridad de los sistemas donde residen. Entre ellas cabe señalar:

- Las macros de muchos programas (como el WORD de Microsoft) que están habilitadas por defecto, haciendo que, al abrir un documento, se ejecuten posibles programas ilícitos de tipo *script*.

- Las facilidades de uso, como los sistemas preparados para poner en marcha un ejecutable con un nombre concreto al colocar un CD en su unidad correspondiente, sin discriminar qué va a realizar el programa.

- Los "*applets*" de *ActiveX* de Microsoft, que se pueden ejecutar automáticamente al ser descargadas desde un navegador de Internet.

- Los "*cookies*", pequeñas estructuras de datos que los servidores web les dan a los navegadores, para que al volver al conectarse al servidor sea más fácil la navegación por el, pero que pueden estar llenos de sorpresas, como que sirvan para conseguir la lista de los sitios web visitados en Internet por el usuario del equipo.

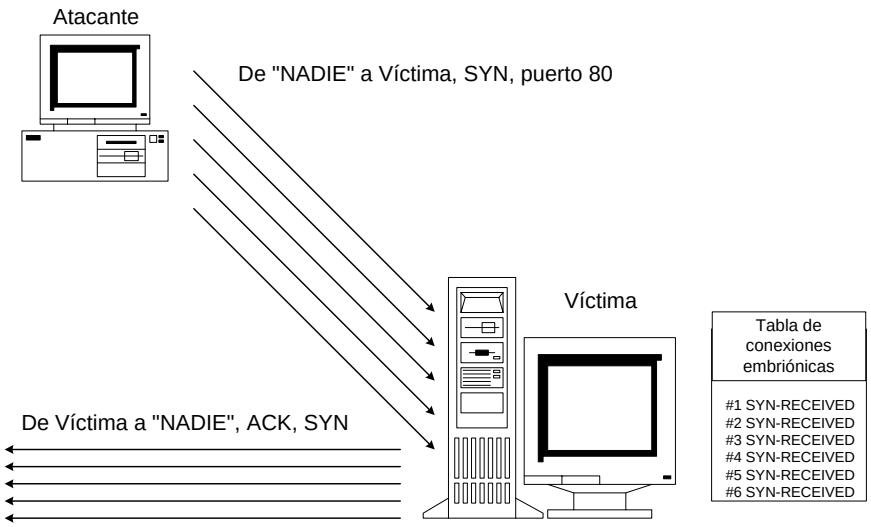


FIGURA 8.1. Ataque de tipo SYN FLOOD TCP.

8.2.4.3. Incorrecta selección de protocolos de aplicación

Como consecuencia muchas veces de la falta de conocimientos y otras de la falta de presupuesto, se dan también muchos casos de selección de protocolos o de aplicaciones inseguras, habiendo otros que, cumpliendo las mismas misiones, son más fiables. Entre ellos los más significativos son:

- Para aplicaciones de acceso remoto a ordenadores aún se utiliza mucho la aplicación *Telnet*, que envía toda la información a la red (incluyendo nombre de usuario y contraseña) en texto claro, haciendo muy sencilla su obtención para alguien trabajando con un analizador de protocolos o "*sniffer*", poniendo en peligro la integridad y la información asociada con el ordenador al que se hizo la conexión. En su lugar puede usarse la aplicación SSH (*Secure SHell*), que cifra los mensajes, teniendo características similares, desde el punto de vista del usuario, a *Telnet*.
- Para aplicaciones de transferencia de ficheros, y de forma idéntica a la anterior, se usa mucho aún la aplicación FTP (*File Transfer Protocol*), con las mismas características de *Telnet* citadas. Igualmente se puede usar SFTP en su lugar.
- Para sitios web que necesitan de una seguridad especial no se debe utilizar el protocolo http, estando disponible el protocolo https, versión "segura" del anterior, que, como se verá más adelante, al hablar de SSL, permite cifrar y autenticar mensajes.
- Para cuestiones mucho más específicas, como la creación y gestión de tablas de encaminamiento dinámico para los encaminadores sigue utilizándose aún bastante el protocolo RIP que, como ya se ha analizado, no autentica los mensajes, estando disponibles hace mucho tiempo ya otros protocolos de encaminamiento que si que permiten gestionar autenticación de mensajes entre encaminadores, como RIPv2, OSPF o EIGRP.

8.2.4.4. Otros

Esta rápida exposición de algunas de los problemas de seguridad más importantes asociados con las aplicaciones no estaría completa sin citar otro tipo problemas generales:

- Los **virus y "caballos de Troya"** que, inyectados en la aplicación o el sistema desde un correo electrónico o un programa de macros o un control de *Activex*, pueden aparecer en el sistema y tener distintos objetivos, todos ellos dañinos, como, por ejemplo, borrar o alterar información, enviar esta información a otros o estudiar el comportamiento del usuario del equipo. Un tipo de ellos especialmente reciente es el conocido como "***spyware***", programas especialmente diseñados para espiar el comportamiento de las aplicaciones del sistema, los sitios web visitados, las aplicaciones puestas en marcha o, en el peor de los casos, obtener información sensible del equipo, como firmas digitales, números de tarjetas electrónicas, etc.

- Los ataques especialmente "creativos", basados en la combinación de diferentes tipos de ataques de los ya citados. Entre ellos, ataques como "*loki*", en el que, utilizando como medio de transporte el campo de texto de los mensajes ICMP de un ping, se intercambia información de cualquier tipo, pasando por encima de las restricciones impuestas por cortafuegos o los ataques de denegación de servicio distribuidos (DDOS), como el "stacheldrath" o el "*tribe* flood", que pueden convertir en "agentes" para un ataque concreto contra un solo equipo X a cientos o miles de equipos, insertando determinado tipo de troyanos que responderán a una orden remota para atacar al equipo victima todos a la vez. En este caso se crea una especie de "red de esclavos" (Fig. 8.2), que responden todos a la vez provocando una situación muy difícil de parar.

8.2.5. Factores de inseguridad en dispositivos de comunicaciones

Además de clientes y servidores en una red existen otros dispositivos de comunicaciones. Dejando para más adelante los específicamente dedicados a la seguridad, como los cortafuegos, esencialmente se tienen conmutadores y encaminadores. Tanto los unos como los otros crean los nexos de unión entre diferentes segmentos de red. Los conmutadores distribuyen el acceso para los equipos de una misma red IP y los encaminadores tienen como misión la cohesión de una sola red de redes (o *Internetwork*), formada por la unión de todas las redes con diferente dirección de red IP. Es fácil entender la importancia de que estos equipos estén funcionando permanentemente y que su configuración sea la correcta y los problemas que podría traer para la red como un todo cualquier fallo de sus características de seguridad.

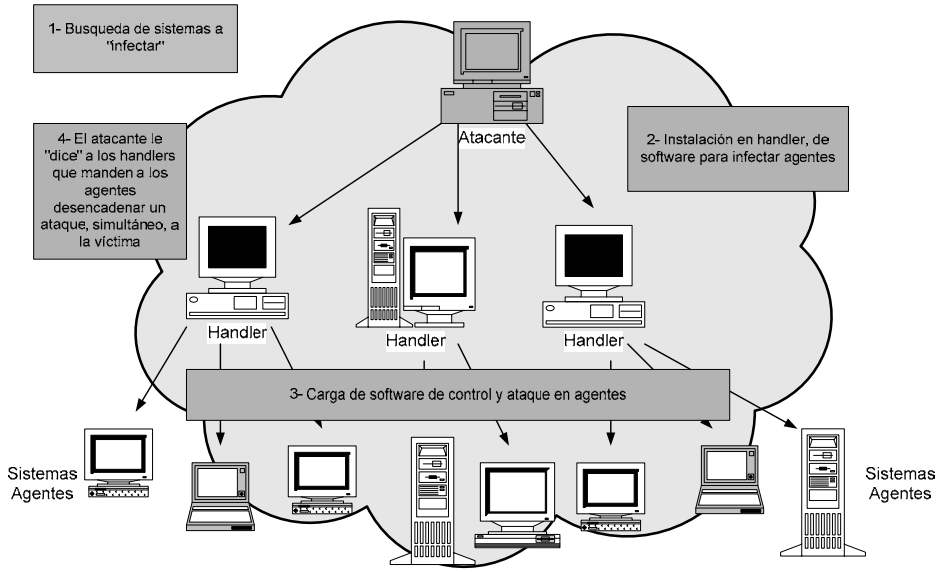


FIGURA 8.2. Esquema de un ataque DOS distribuido típico.

Lo primero que hay que destacar de estos dispositivos es que no son nada diferentes en lo esencial a los citados en el apartado anterior. Están compuestos por una plataforma concreta (hardware), un sistema operativo (software) y un conjunto de ficheros y aplicaciones, al igual, por ejemplo, que un servidor. Lo que les hace realmente diferentes es que cada una de estas partes está diseñada con el objetivo concreto de servir de nexo de unión en la red, no son sistemas de propósito general.

Dicho esto, se les puede aplicar todo lo comentado hasta ahora, remarcando únicamente cuáles podrían ser las consecuencias en la red de un ataque con éxito:

- Si no se está atento a las actualizaciones del sistema, se pueden sufrir ataques, basados en vulnerabilidades del software ya descubiertas, como ataques de denegación de servicio que, al dejar al encaminador inactivo, paralicen la actividad de la red o permitan un ancho de banda excesivo para ciertas aplicaciones y muy poco para otras.
- Si no existe seguridad física y resultan físicamente atacados se puede perder la cohesión de la red entera o, al menos, de una parte de ella. Además, si el acceso físico es fácil se puede reconfigurar el sistema con fines ilícitos, arrancándolo en su modo de arranque mínimo.

- Si su gestión remota se hace mediante protocolos poco seguros (*Telnet*, HTTP) se está expuesto al mismo tipo de problemas de reconfiguración.
- Si la configuración es incorrecta (contraseñas débiles, filtros de usuario poco coherentes, listas de acceso incompletas, mala parametrización) se puede abrir caminos a ataques específicos.
- Si los protocolos de encaminamiento que se usan son poco fiables (caso de RIP o de IGRP), se pueden dar casos de ataques al encaminamiento incorrecto de todas las rutas en la red.

Desde el punto de vista de la seguridad todos estos dispositivos se comportan como los citados anteriormente y lo único que se debe poder calibrar es la incidencia en la red de una configuración incorrecta de su seguridad.

8.3. SOLUCIONES ACTUALES EN SISTEMAS Y DISPOSITIVOS DE SEGURIDAD

Desde el punto de vista técnico se dispone de un arsenal de herramientas para poder hacer frente a todos estos problemas potenciales, con la excepción de la mala codificación, cuya única solución es escribir los sistemas y las aplicaciones correctamente desde el punto de vista de seguridad, algo claramente fuera de las expectativas reales de influencia de los administradores de redes.

Tal como se ve en la Tabla 8.1 se puede hacer fácilmente una enumeración de estas defensas:

- Contra la inseguridad física establecer un perímetro de seguridad alrededor de las máquinas y dispositivos, especialmente de los más sensibles.
- Establecer un método de actualización de software con problemas de seguridad, que permita estar al tanto rápidamente de las vulnerabilidades exhibidas por cada una de las piezas de software de la red.
- Establecer una correcta política de copias de seguridad, seleccionando bien cómo se hacen, cuándo y dónde se guardan y probándolas aún en ausencia de problemas.

- Contra los ataques de acceso a sistemas, elegir correctamente contraseñas, cambiarlas periódicamente, proteger correctamente los ficheros, etc.
- En el caso de necesidad de una autenticación más fuerte se pueden usar además ya sistemas más avanzados que las contraseñas como tarjetas de tipo token o sistemas biométricos basados en huella dactilar o el iris, aunque estos mismos tienen también sus problemas y están lejos de ser 100% aceptables.
- Si se necesita, además, autenticación de acceso remoto se puede usar sistemas AAA (*Autenticación Authorization Accounting)*, basados en protocolos como el RADIUS o el TACACS/+, que son más complejos de gestionar pero que permiten ofrecer una mayor (y extendida a más usuarios) seguridad en la autenticación, además de proporcionar herramientas de autorización de usuarios y de gestión del uso de recursos por conexión de usuario.
- Como se verá más adelante se puede usar también métodos de autenticación basados en la **firma digital** que, con una implementación más compleja por su fuerte componente criptográfico y su complejidad organizativa, ofrecen una seguridad mucho mayor.
- Hacer una administración de los sistemas correcta, lo cual significará conocer en profundidad el sistema del que se trate, aunque, en todos los casos, esto pasará por asegurar ficheros correctamente y elegir un sistema de ficheros que cumpla las propiedades de privacidad, integridad y autenticación deseadas en cada sitio.
- Utilizar sistemas antivirus y antispyware actualizados que se ejecuten periódicamente, manteniendo cada sistema libre de virus, troyanos, etc.
- Formar suficientemente a los usuarios para que entiendan los principios básicos de la seguridad informática y sean menos susceptibles a ataques de ingeniería social, así como para que comprendan la necesidad de usar los sistemas antivirus citados y entiendan los peligros asociados con correos electrónicos no identificados.
- Desde el punto de vista de los administradores de seguridad se dispone, además, de toda una batería de herramientas de gestión de la seguridad que se detalla más adelante como cortafuegos, sistemas de detección de intrusiones, detectores de vulnerabilidades, cuyo uso o no dependerá de diferentes factores. Algunas de estas herramientas se han vuelto casi obligatorias en las redes actuales.

TABLA 8.1. Herramientas típicas de defensa

PROBLEMAS DE SEGURIDAD	DEFENSAS POSIBLES
Inseguridad Física	Establecer perímetro de seguridad
Vulnerabilidades del software	Método seguro de actualización del software
Pérdida de datos sensibles	Política correcta de copias de seguridad
Ataques de acceso no permitido a sistemas	• Selección de contraseñas • Tarjetas token • Autenticación AAA • Firma digital • Sistemas biométricos
Virus, troyanos, *spyware*, etc.	Sistemas antivirus y antispyware actualizados
Ingeniería social	Formación básica para usuarios
Problemas de seguridad en redes en general	• Cortafuegos • Sistemas de detección de intrusiones • Detectores de vulnerabilidades

8.4. POLÍTICA DE SEGURIDAD PARA REDES

Siguiendo la definición ampliamente aceptada del *Site Security Handbook*, RFC 2196, una política de seguridad informática es:

> Una serie de sentencias formales, o normas, que deben ser cumplidas por todas las personas de una organización que dispongan de acceso a cualquier información, datos o tecnología que sean propiedad de la organización.

El propósito principal de una política de seguridad es informar a los usuarios, trabajadores y personal de la organización en general sobre los requisitos obligatorios a cumplir por cada uno de ellos para proteger los valores tecnológicos e información de la organización. La propia política debe especificar qué mecanismos utilizar para cumplir las normas y la organización y metodología a seguir en cada caso particular. Si es suficientemente completa deberá incluir una base suficiente para seleccionar con éxito las propias tecnologías, incluyendo sistemas y software.

Entre sus características principales se puede señalar:

- Debe poderse implantar, es decir debe permitir que siga realizándose el trabajo de manera normal y no alterarlo significativamente.
- Debe entenderse fácilmente, de manera que todos los miembros de la organización conozcan su existencia y las normas básicas generales, y que cada responsable particular esté al tanto de sus propias tareas concretas.
- Debe hacerse cumplir. Muchas de las normas exigirán algo más de trabajo, con lo que si no hay una cierta coerción posiblemente se convertirán en papel mojado. Deben definirse claramente los distintos grados de responsabilidad.
- Debe cumplir la legislación, en particular en el caso español, la Ley Orgánica de Protección de Datos (LOPD) y la Ley de Servicios de Sistemas de Información y Comercio Electrónico (LSSICE).
- Debe incluir mecanismos de respuestas a incidentes y de actualización de la propia política.

En la creación de una política de seguridad deben participar distintos grupos representativos, no solamente técnicos sino también de negocio, legales e incluso a veces puede ser interesante tener consultores externos.

En el trabajo de creación de la política hay que tener en cuenta cómo contestar a una serie de preguntas clave:

- **¿Qué se quiere proteger?** Esto implica realizar un inventario de los activos de la organización, tanto tangibles (máquinas, sistemas, cableado, etc.) como intangibles (aplicaciones, sistemas operativos, datos, etc.) y conocer distintos estándares organizativos, como el ISO 27001, y de selección, como el ISO 15408 o *Common Criteria.*

- **¿Quiénes pueden ser los atacantes?** Hay que decidir en quién se puede confiar y qué motivaciones podrían tener distintos tipos de atacantes.

- **¿Cómo se van a utilizar los sistemas y herramientas elegidos para implementar la política de seguridad?** Para ello hay que conocer lo mejor posible los distintos tipos de ataque y las distintas posibles defensas.

- **¿De qué dinero se dispone para ponerla en marcha?** Hay que tener en cuenta gastos de adquisiciones técnicas, pero también de formación, de configuración de sistemas y herramientas, gastos asociados al tiempo empleado en la administración correcta y mantenimiento de las herramientas, etc.

Una vez implementada se convierte en el motor del proceso de seguridad, un proceso que no puede pararse pues esto implicaría confiar en que ya todo está hecho y que no van a aparecer nuevos problemas.

Este **proceso de seguridad** tiene varias fases (Fig. 8.3).

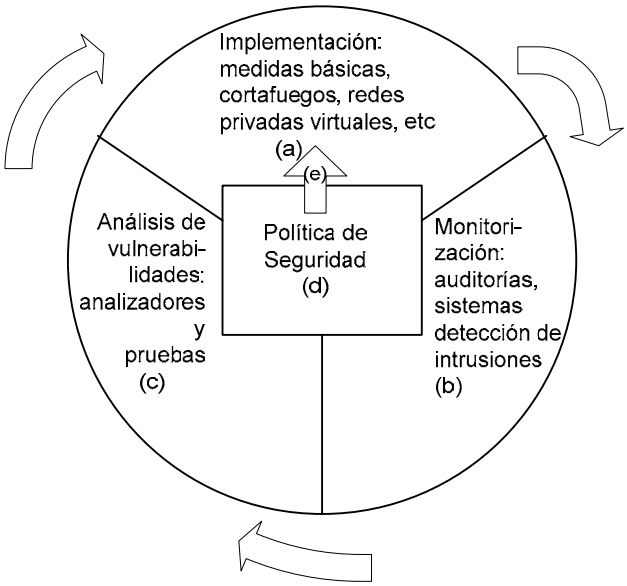


FIGURA 8.3. El proceso de seguridad informático, basado en la política de seguridad.

En la **fase de implementación** (Fig. 8.3 a), se activan todas las medidas organizativas, se configuran todas las medidas básicas de seguridad de sistemas y dispositivos de comunicación y se configuran todos los dispositivos de seguridad, como cortafuegos o sistemas de detección de intrusiones, IDS. En la **fase de monitorización** (Fig. 8.3 b), se auditan los sistemas y las redes mediante los registros de sistemas, aplicaciones o IDS, en busca de incumplimiento de alguna política o de aparición de amenazas o debilidades no tenidas en cuenta en la creación de la política seguridad. En la **fase de análisis de vulnerabilidades** (Fig. 8.3 c), se prueban los sistemas y aplicaciones

mediante detectores de las mismas o se hacen ataques simulados para comprobar su estado, en busca de vulnerabilidades no detectadas, o nuevas, y no tenidas en cuenta en la política de seguridad. Con todos los datos nuevos de inseguridades encontrados se debe actualizar la política (Fig. 8.3 d), para implantar las nuevas medidas (Fig. 8.3 e), y que todo vuelva a empezar.

No se debe olvidar, para finalizar, que, en el día a día, el cumplimiento consistirá en seguir una serie de normas, lo más sencillas posibles, sobre uso aceptable de equipos y servicios, sobre accesos remotos permitidos, protección de la información, seguridad física, respuestas a incidentes, cifrado aceptable, seguridad en las adquisiciones, seguridad de portátiles, políticas de antivirus, auditorias, etc.

8.5. DEFENSAS NO CRIPTOGRÁFICAS

Además de las defensas comunes ya citadas en apartados anteriores existen una serie de herramientas de seguridad susceptibles de ser utilizadas dependiendo del tipo de redes. Aún siendo una separación puramente académica (pues algunas de ellas pueden jugar también papeles de defensas criptográficas) se van a describir las más comunes, que son:

- Los cortafuegos.
- Los Sistemas de detección de intrusiones.
- Los detectores de vulnerabilidades.

8.5.1. Cortafuegos

Un cortafuegos implementa una aproximación de red al problema de mantener la seguridad de los sistemas de las redes. Desde un punto de vista ideal, un cortafuegos (Fig. 8.4), debe cumplir las siguientes características:

- Todo tráfico entre las dos redes debe pasar a través del cortafuegos.
- Solo el tráfico autorizado por el cortafuegos debe dejarse pasar.
- El cortafuegos debe ser completamente inatacable.

Entre los puntos fuertes de los cortafuegos se puede señalar especialmente que son un sitio habitual para implantar buena parte de la política de seguridad,

que pueden soportar técnicas avanzadas de autenticación, son un buen sitio para centralizar alarmas y registros de auditoria de tráfico y, al no ser sistemas de propósito general, necesitan poca administración de usuarios.

Entre sus puntos débiles hay que señalar que su uso puede hacer bajar la guardia en la implantación de mecanismos de seguridad de los sistemas y redes que se desea proteger, que si son muy sofisticados pueden necesitar una configuración muy compleja, que pueden representar un cuello de botella para el tráfico de la red y uno muy importante: **ningún cortafuegos puede evitar ataques originados en la parte protegida de la red y cuyo objetivo reside en la misma parte.**

Las tecnologías de cortafuegos más comunes son:

- La basada en **filtros de paquetes**. Operan al nivel IP y de transporte. Filtran mensajes IP, dejándolos pasar o no, a través de cada interfaz, basándose para ello en los valores de algunos de los campos más importantes de las cabeceras IP, TCP y UDP de cada mensaje. Entre los ejemplos reales más típicos se encuentran los encaminadores de prácticamente cualquier fabricante. Entre los campos de cabeceras más comunes para hacer de criterios para los filtros están las direcciones IP origen y destino, los números de puerto origen y destino TCP o UDP y el tipo de protocolo que aparece en la cabecera IP de cada mensaje.

- La basada en **servidores *proxy***, que exigen un proceso *proxy* en el cortafuegos por cada uno de los protocolos que se quiere filtrar, interceptando peticiones de clientes a servidores, las evalúa en función de una serie de criterios del nivel de red y de aplicación típicamente y, si permite la comunicación, crea una conexión entre el mismo y el servidor “a favor” del cliente y otra entre el mismo y el cliente, creando una asociación por cada pareja de conexiones de estas características. Quizás el ejemplo más popular sea el *Proxy Server* de *Microsoft*.

- La basada en inspección dinámica de tráfico por sesión y completa o “***stateful inspection***”. En estos cortafuegos se obtiene, almacena y manipula coherentemente información de todos los niveles de comunicación para cada comunicación TCP o UDP. Esta información, completa para cada mensaje y gestionada dinámicamente, se utiliza para decidir si se permite el paso del tráfico o no. Entre loe ejemplos más comunes están el *Cisco* PIX *Firewall* o el *Firewall-1* de la empresa israelí *Checkpoint*.

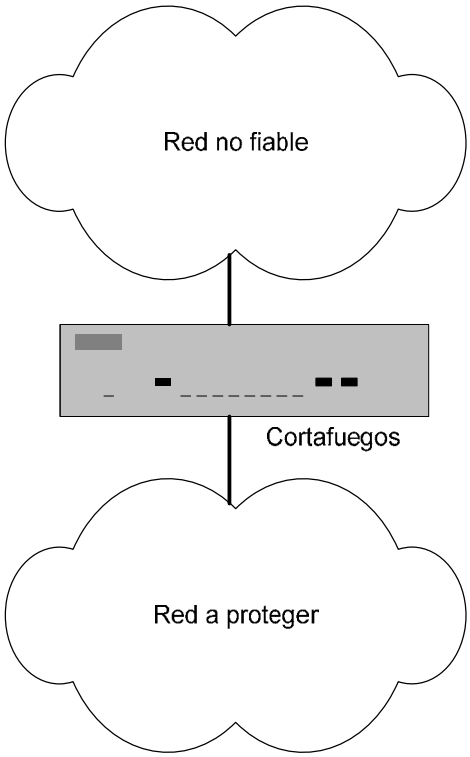


FIGURA 8.4. Un cortafuegos genérico.

Habría que remarcar que actualmente los productos comerciales ya citados y muchos otros combinan estas tecnologías siendo posible encontrarse un cortafuegos basado en tecnología de tipo stateful que, además, puede hacer filtrado de paquetes tradicional y configurar diferentes niveles de servidores proxy, tal como el citado Cisco PIX Firewall. De hecho estos productos suelen poder configurarse como encaminadores, puntos de entrada de redes privadas virtuales, sistemas de detección de intrusiones, etc. Se está asistiendo a la transformación de muchos de estos dispositivos que están dejando de ser simples cortafuegos y pasando a convertirse en súper-dispositivos de seguridad.

Para acabar esta sección hay que hacer referencia también a los cortafuegos personales, cada vez más extendidos (aunque polémicos en cierta manera). Estos cortafuegos consisten en una aplicación que se instala en el ordenador que se quiere proteger y se configura de manera que protege a un único sistema (aquel en el que está instalado) frente a los posibles ataques provenientes del resto de la red.

Las técnicas que utilizan suelen ser basadas en tecnología de tipo "*stateful inspection*" y su uso suele ser añadido al de los cortafuegos más tradicionales ya citados.

8.5.2. Sistemas de detección de intrusiones

Una intrusión puede definirse como un mensaje, o serie de mensajes, que implementan, de muy diferentes formas posibles, ataques a sistemas o dispositivos en la red. Un sistema de detección de intrusiones o IDS (*Intrusión Detection* System) suele ser un sistema en la red especializado en detectar y parar ese tipo de intrusiones.

Los primeros modelos se basaban en la **detección de anomalías**. Se crean perfiles de uso típico por tipo de usuario, por tipo de trabajo, etc. Son realmente muy difíciles de usar pues son muy propensos a producir muchos falsos positivos, mensajes identificados como de ataque (en este caso por no coincidir con perfiles correctos) cuando no lo son. En esta tecnología es muy difícil acertar pues el tráfico de una red compleja es todo menos predecible y, como consecuencia, es muy complicado crear perfiles con sentido.

Casi todos los modelos reales actuales se basan la **detección de firmas de ataque**. Una firma de ataque es un mensaje (por ejemplo un mensaje de tipo ping con una dirección IP destino de tipo *broadcast*) o grupo de mensajes (por ejemplo muchos mensajes seguidos con iguales direcciones IP destino y de tipo SYN TCP, es decir, identificativos de un ataque SYN FLOOD) que indican con una cierta fiabilidad que hay un ataque en curso.

Estos sistemas (por ejemplo los *Cisco* IDS o el *Real Secure* de ISS) se comercializan con un conjunto extenso de firmas de ataque, correspondientes a los ataques conocidos en el momento de la venta, pero tal conjunto hay que actualizarlo frecuentemente ya que, como ya se ha analizado, aparecen nuevos problemas con mucha frecuencia. Además, en algunos casos, los productos permiten crear firmas de ataque propias.

Una clasificación importante a tener en cuenta en los IDS es la que los separa entre de red y locales. Los sistemas IDS de red monitorizan el tráfico de uno o varios segmentos de la red (Fig. 8.5) independientemente de la dirección IP origen del tráfico y de su dirección IP destino.

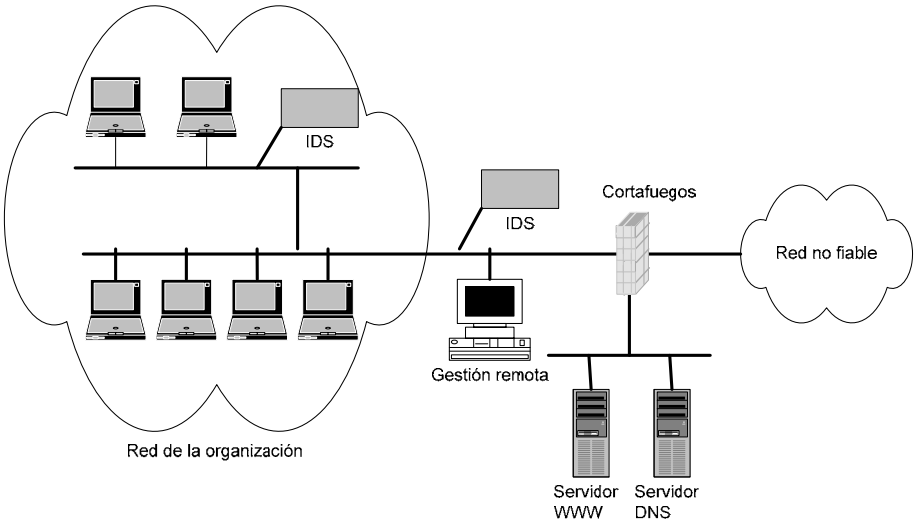


FIGURA 8.5. Detección de intrusiones basada en la red.

Los sistemas locales IDS se instalan como una única aplicación que salvaguarda un único sistema. Es una solución más barata pero sólo protege al citado sistema.

El procedimiento habitual de su funcionamiento es el siguiente:

1. Una vez configurados monitorizan todo el tráfico que pasa por la tarjeta de red que usan.

2. Comparan ese tráfico con cada una de las firmas para las que están configurados.

3. Si encuentran una firma que se ajusta a un tráfico determinado desencadenan la acción que tienen configurada, que puede ser una combinación de registrar el ataque, enviar una alarma o parar el ataque.

8.5.3. Detectores de vulnerabilidades

Se puede definir un detector de vulnerabilidades como un programa que busca, de manera automatizada, vulnerabilidades y debilidades de entre una gran lista que conoce y que se actualiza como en el caso anterior, en un sistema concreto o en muchos sistemas diferentes en una red. Su resultado final es un informe sobre los problemas encontrados y qué se debe hacer para subsanarlas.

Son sistemas preactivos, permiten afrontar un posible problema antes de que suceda. Además, algunas de sus características más comunes son:

- Deben poderse **actualizar frecuentemente**. La instalación de estas actualizaciones debe ser sencilla.

- Deben indicar claramente **cómo solucionar cada uno de los problemas** encontrados, teniendo algunos de ellos enlaces directos con las actualizaciones que hay que instalar y que detallan los pasos a dar.

- Deben informar de todo aquello que no son capaces de analizar, indicando que no fueron capaces de hacerlo.

8.6. DEFENSAS CRIPTOGRÁFICAS

Se empezará por definir las principales características de seguridad deseadas posiblemente para distintos tipos de mensajes en una red. Pueden desearse como propiedades únicas o en combinación, es decir se puede requerir, por ejemplo, únicamente privacidad para un grupo de mensajes y privacidad y autenticación para otro grupo de mensajes.

Se entiende por **autenticación** la propiedad que permite demostrar que uno es quien dice ser. En el contexto de los mensajes de red se habla de mensajes para los que se puede demostrar que han sido enviados desde la dirección IP origen que aparece en el propio mensaje, que se puede demostrar que han sido enviados por un usuario concreto de la red o una combinación de ambas cosas. Esta propiedad se busca típicamente para sistemas de firma digital o, en general, para sistemas de control de identidad y suele estar implementado mediante algoritmos matemáticos de criptografía de clave pública, a veces combinados con sistemas "*hash*".

Se entiende por **privacidad** la propiedad que permite que un mensaje con tal característica enviado de un punto a otro de la red solo sea legible, entendible, por los dos extremos origen y destino; si es interceptado en un punto intermedio no es legible. Es lo que se conoce vulgarmente como "mensaje cifrado". Esta propiedad se busca típicamente en tráfico confidencial entre diferentes sitios de una empresa, conectados entre si a través de una red pública como la red Internet y está implementado mediante algoritmos matemáticos de criptografía de clave secreta o simétricos.

Se entiende por **integridad** la propiedad que permite garantizar que un mensaje enviado no ha sido modificado en su transito al sitio de destino o detectar si ha sido modificado. Esta propiedad se necesita en cualquier sistema que tenga como misión avisar de posibles cambios no deseados, como por ejemplo en sistemas de firma digital y se implementa mediante algoritmos criptográficos conocidos como sistemas de una sola vía o funciones "*one-way hash*".

Estas características son la base de la arquitectura de cualquier sistema criptográfico (Fig. 8.6). Una vez elegidas las propiedades deseadas se cuenta con una serie de algoritmos utilizados por protocolos criptográficos cuya selección determinará el tipo de sistema que se deba utilizar. Es importante señalar que un profesional de la seguridad informática debe conocer estos fundamentos pero no tiene ninguna necesidad de ser un experto matemático sino entender la trascendencia de las diferentes configuraciones de los protocolos y sistemas criptográficos con los que trabaje.

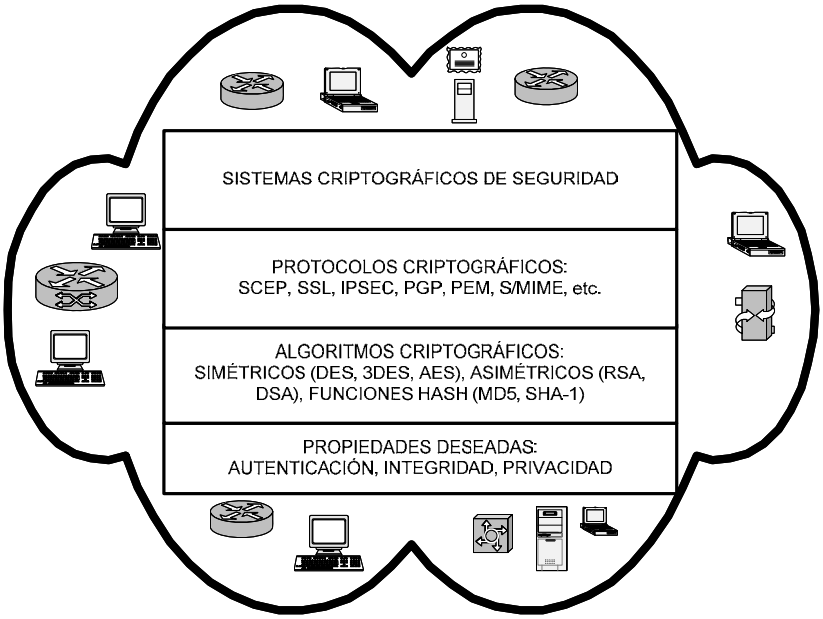


FIGURA 8.6. Niveles de la arquitectura de sistemas criptográficos.

Se van a describir seguidamente las características más típicas de cada uno de los tipos de algoritmos matemáticos implicados en la consecución de estas propiedades, pasando después a considerar como son utilizados dentro de protocolos criptográficos que los utilizan y, como ejemplo de un sistema final, se analizarán las características más importantes de una red privada virtual.

8.6.1. Algoritmos criptográficos: Clave privada, funciones *Hash* y clave pública

Antes de empezar es muy importante señalar que **los algoritmos criptográficos deben ser públicos** para poder ser considerados seguros. Antes de estandarizarse son suficientemente probados por la comunidad científica y profesional, se hace un control de calidad exhaustivo. Todo algoritmo criptográfico secreto, no público termina por provocar fallos graves de seguridad. La seguridad reside sobre todo en la clave o en las claves que se utilizan y en la metodología mediante la cual se utilizan los algoritmos, tal como se verá a continuación.

Un algoritmo de clave privada (o de **criptografía simétrica**) utiliza una única clave, que sirve tanto para cifrar un texto como para descifrarlo (Fig. 8.7), clave compartida únicamente entre los participantes del sistema.

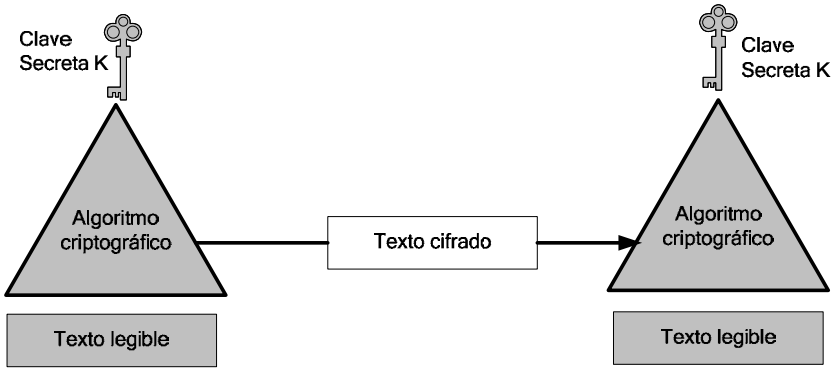


FIGURA 8.7. Algoritmo de criptografía simétrica.

Entre sus puntos fuertes se pueden destacar que son mucho más rápidos que los algoritmos de clave pública y que son los usados tradicionalmente en sistemas hardware de cifrado.

Entre sus puntos débiles está la necesidad de un sistema de distribución de la clave muy seguro. La clave hay que modificarla con una cierta periodicidad y existe el peligro de que, al caer en manos no autorizadas, la información cifrada esté disponible para quien no debería de estarlo. También es un problema cuando se desea utilizar por parejas de usuario, una clave diferente por cada pareja de usuarios, pues una red de n usuarios necesitaría gestionar n(n-1)/2 claves diferentes.

Entre los más ampliamente usados merece destacarse el DES (*Data Encryption Standard*), el 3DES, una versión mucho más sofisticada del DES y el AES (*Advanced Encryption Standard*), cada vez más utilizado en diferentes ámbitos.

Las **funciones de una sola vía** reciben su nombre debido a su naturaleza matemática: dado un mensaje (o un texto) x, es muy fácil mediante el algoritmo calcular el resultado f(x), al que se le denomina el *hash* de x. Lo significativo es que resulte prácticamente imposible (Fig. 8.8), dado el *hash* f(x) obtener x.

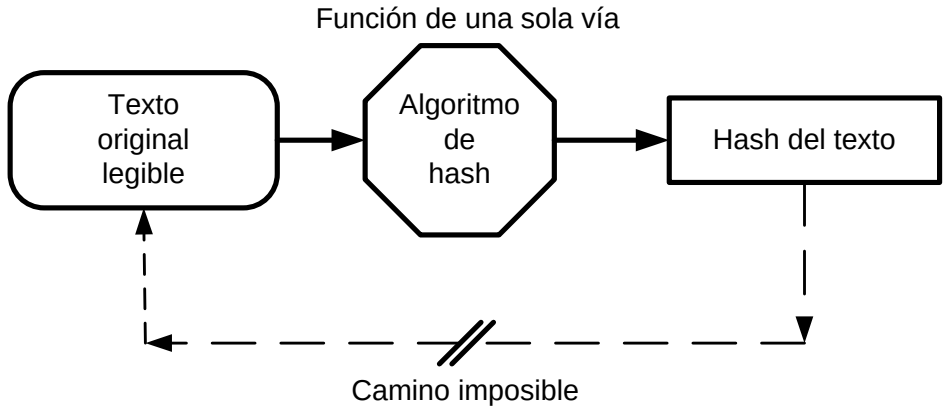


FIGURA 8.8. Mecanismo típico de una función de una sola vía.

Su uso más habitual es el de garantizar la integridad del texto, ya sea éste un fichero o un mensaje. Las más utilizadas dependen, para mayor seguridad, de una clave privada compartida y entre ellas se debe citar el MD5 (*Message Digest* 5) o el SHA (*Secure Hash Algorithm*). El procedimiento es simple: se

envía el mensaje junto con su *hash* y el receptor, al recibir el mensaje, separa éste del *hash*, aplica la misma función empleada en el origen del mensaje y compara el *hash* resultante. Si no son iguales, el mensaje ha sido modificado en el camino.

Un algoritmo de criptografía de clave pública (o de **criptografía asimétrica**) se basa en las siguientes características:

- Cada participante del sistema genera de manera simultánea, mediante el algoritmo, una pareja de claves íntimamente relacionadas entre sí, la clave pública del participante y la clave privada del participante.

- La clave pública puede ser conocida por todos los participante sin problema alguno de seguridad.

- La clave privada sólo es conocida por el propio participante.

- Cualquiera que conozca la clave pública del participante A puede cifrar mediante ella un mensaje y enviarlo a la red, pero únicamente el participante A podrá descifrarlo pues esta operación sólo se puede realizar mediante la clave privada de A (Fig. 8.9).

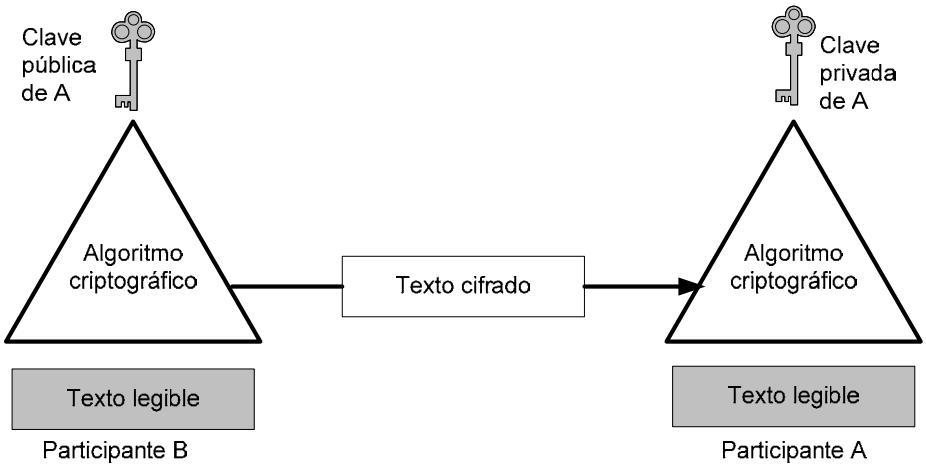


FIGURA 8.9. Criptografía de clave pública.

Entre los puntos fuertes de esta clase de algoritmos merece destacarse la facilidad para conseguir la autenticación mediante ellos y la posibilidad de combinarlos con otros métodos criptográficos con facilidad.

Entre sus puntos débiles están la lentitud en comparación con los de clave simétrica, con lo que no resultan aún aptos para operaciones en las que se busca la privacidad de los mensajes y, sobre todo, la necesidad de un buen sistema añadido de gestión de claves.

De entre los algoritmos de esta clase el más utilizado con diferencia es el RSA, acrónimo de los apellidos de sus creadores Rivest, Shamir y Adleman. Ha estado protegido por una patente de los laboratorios RSA hasta Septiembre del año 2000 lo que le hizo restringido durante años al entorno comercial.

El problema citado de la gestión de claves resulta especialmente importante debido al uso continuado y creciente de los **sistemas de firma digital** y consiste en la imposibilidad, con lo analizado hasta ahora, de asegurar que la clave pública de un participante del sistema es realmente la clave pública, es decir, en la autenticación de la clave pública. En una red con muchos participantes este problema, como puede uno darse cuenta fácilmente, es realmente complicado de resolver.

En el modelo de firma digital más extendido, el basado en el algoritmo RSA de criptografía pública (Fig. 8.10), el procedimiento de firma de un mensaje es el siguiente:

- El emisor genera un *hash* del mensaje, H1, mediante una función de una sola vía previamente pactada con el receptor.

- Este H1 se cifra mediante RSA usando la clave privada del emisor y el resultado es lo que se conoce como firma digital, FD, del mensaje, que se adjunta al mensaje. Nótese que la firma digital cambia cada vez que se envía un mensaje diferente.

- Cuando el mensaje llega a su destino, el receptor separa el mensaje de la firma digital.

- Calcula el *hash* del mensaje mediante la función pactada y obtiene un *hash* H2 y descifra la firma digital mediante RSA y la clave pública del emisor obteniendo H1, el *hash* original.

- Si H1 y H2 son idénticos, puede afirmarse que el mensaje fue enviado por el propietario de la clave pública usada (autenticación) y que no fue modificado en transito (integridad).

Se puede comprobar fácilmente que el punto débil más grave es la imposibilidad de asegurar que la clave pública del emisor esté autenticada. Aunque no se vaya a entrar en detalle en la complejidad de la creación y mantenimiento de un sistema de gestión de firma digital se ha de conocer al menos los rudimentos de tales sistemas.

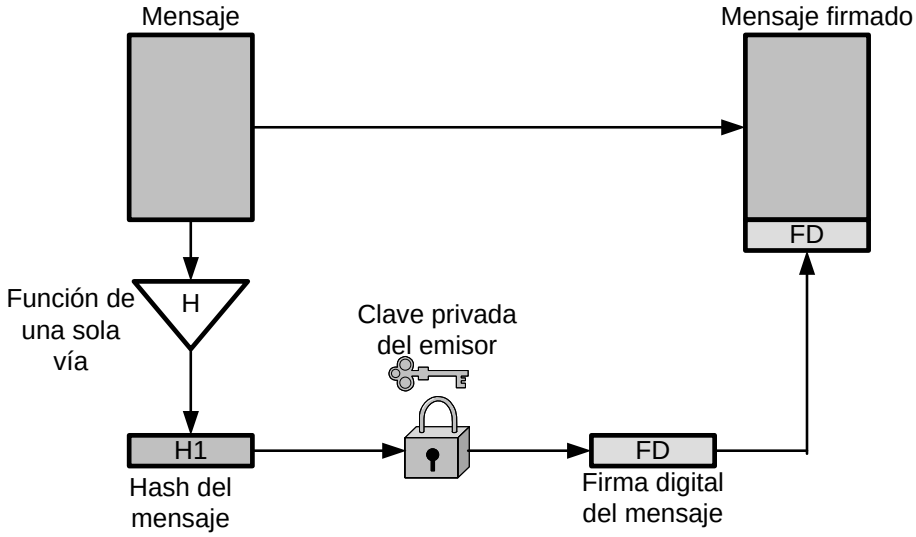


FIGURA 8.10. Procedimiento de firma digital antes del envío de un mensaje.

Debe existir en la red al menos un sistema de Autoridad de Certificación (AC) que emite un certificado digital, en formato estándar X.509 (conteniendo varios campos en los que se identifica la clave pública del participante, la AC, el algoritmo asimétrico utilizado, la función de una sola vía pactada, etc.), para la clave pública de cada participante legítimo y se lo hace llegar junto con un certificado propio de la AC. Estos certificados digitales tienen un formato concreto, dado por el estándar, y están firmados digitalmente por la propia AC.

Cuando un participante A quiere enviar un mensaje firmado digitalmente por el a otro, B, envía previamente su certificado digital a B y recibe de B el certificado digital de B.

En B, al disponer de un certificado digital de la AC se puede comprobar la validez del certificado de A, comprobando la firma digital mediante la clave pública de la AC y el algoritmo de *hash* pactado. En A el procedimiento es el mismo y esto lleva a la autenticación correcta en ambos sentidos, A autentica a B y B autentica a A.

Es importante darse cuenta de que cuando se dice A y B estos serán una cosa u otra dependiendo del sistema criptográfico en el que se trabaje. Pueden ser certificados de usuario, de servidor en la red, de encaminador, etc.

La complejidad del formato de los certificados, de los algoritmos utilizados, de conceptos como la lista de revocación de certificados y, quizás la más importante, la dificultad de gestionar la identificación correcta de los participantes para asociarles un certificado digital están fuera de la extensión del capítulo y se pueden consultar en los libros de la bibliografía.

8.6.2. Protocolos criptográficos: SSL, IPSec y otros

Un protocolo criptográfico es simplemente un protocolo de comunicaciones que, como parte de sus funciones, usa métodos criptográficos, independientemente de que su objetivo final vaya más allá de lo que estos permiten conseguir. Se crean usando métodos tradicionales de programación de protocolos y algoritmos criptográficos, pero sirven (Fig. 8.6) a la vez de piezas básicas con las que construir sistemas criptográficos más complejos.

El protocolo SSL (*Secure Sockets Layer*), utilizado en aproximadamente el 90% de los sitios de comercio electrónico de la red Internet, aparece como una herramienta más dentro de los navegadores principales (*Netscape Navigator*, *Internet Explorer*, *Mozilla*, etc.) como herramienta básica de seguridad. Se puede saber qué sitios web lo utilizan pues para conectarse a ellos hay que identificarlos como “**https**” es decir “http seguro”. No obstante su uso actual va mucho más allá del comercio electrónico, estando implicado, por ejemplo, en muchas de las herramientas de gestión remota de cualquier tipo de plataforma.

Las propiedades de seguridad que proporciona SSL son:

- **Autenticación**, permitiendo identificar los participantes y sus mensajes dentro de un grupo de participantes seleccionados. Es importante remarcar que el estándar del protocolo la marca como opcional, pero se debe utilizar siempre que se pueda. Se utiliza normalmente RSA.

- **Privacidad**, para poder cifrar y descifrar mensajes sólo dentro del grupo citado. Se cifra la cabecera de datos de cada mensaje SSL. Se utilizan normalmente DES o AES.

- **Intercambio seguro de claves**, para poder intercambiar el material necesario para asegurar las propiedades anteriores.

Al iniciarse una sesión SSL el cliente y el servidor se ponen de acuerdo en la versión del protocolo, los algoritmos de cifrado y, opcionalmente, si debe haber autenticación o no. Una vez llegados a un acuerdo, se autenticarán, normalmente mediante firma digital, se pondrán de acuerdo en la clave para el cifrado simétrico y empezarán el intercambio seguro de mensajes.

Otros protocolos de parecido objetivo son el **TLS** (*Transport Layer Security*), originado como desarrollo del SSL por parte del IETF o el **SET** (*Secure Electronic Transactions*) desarrollado con la intención de subsanar los defectos de SSL de cara al comercio electrónico y a la implicación de las entidades bancarias y aún no demasiado implantado.

Los protocolos **IPSec** (*IP Security*) son un grupo de estándares desarrollados por el IETF, dentro del proyecto general de renovación que ha supuesto la nueva versión IPv6, con el objetivo de permitir comunicaciones seguras a distintos niveles entre dos puntos cualesquiera de una red IP. Su uso en IPv4, versión utilizada todavía por más del 95% de los sistemas que trabajan en redes IP, ha resultado un éxito, convirtiéndose, por ejemplo, en los protocolos de seguridad más extendidos y aceptados par ala implementación de redes privadas virtuales.

IPSec ofrece integridad de los datos, autenticación del origen de los datos enviados, privacidad de los datos y protección contra repetición de mensajes, pudiendo combinarse las propiedades y no siendo necesario que todos los mensajes IPSec exhiban las mismas propiedades de seguridad.

Los algoritmos utilizados son todos completamente estándar y los protocolos están diseñados para permitir con facilidad la incorporación de nuevos algoritmos según esto vaya siendo necesario.

Cuando se habla de IPSec en realidad se habla de tres protocolos independientes, responsables cada uno de ellos de diversas funciones:

- El protocolo **AH** (*Authentication Header*), que proporciona integridad y autenticación de los datos mediante funciones hash MD5 o SHA aplicadas a la cabecera de datos de los mensajes.

- El protocolo **ESP** (*Encapsulation Security Payload*) proporciona privacidad de los datos mediante algoritmos simétricos como DES, 3DES o AES, aunque, opcionalmente, puede proporcionar también integridad y autenticación mediante los algoritmos de una sola vía citados.

- El protocolo **ISAKMP** (*Internet Security Association Key Management Protocol*) que se encarga de la administración y el intercambio seguro de todas las claves necesarias para ello. Puede usar, a su vez, otros protocolos para conseguir un procedimiento seguro, entre ellos el protocolo de *Diffie-Hellman.*

Mediante el ESP se puede, además, operar en dos modos diferentes. En modo transporte se protege todo el mensaje menos la cabecera IP y el es modo común para cuando entre origen y destino no hay ningún encaminador. En modo túnel se protege todo el mensaje incluyendo la cabecera IP lo que obliga a modificar el propio mensaje añadiéndole una nueva cabecera de IP en texto claro y con las direcciones origen y destino de los puntos de la red donde se cifra en ESP y en ese modo. Es típico de las redes privadas virtuales.

Cuando se va a utilizar IPSec entre dos puntos de una red previamente hay un intercambio de mensajes para crear una Asociación de Seguridad, que será el resultado de pactar entre los dos puntos los algoritmos que se van a utilizar en cada paso de la comunicación. Aunque los detalles están más allá del contenido de este capítulo se puede ilustrar con una de sus partes en la cual los dos puntos deben decidir cómo se van a autenticar el uno al otro. Para ello se dispone de:

- Autenticación mediante clave secreta compartida. En este caso deben compartir una contraseña que servirá para cifrar ciertos mensajes de presentación entre ellos. Es el método típico para pocos participantes en la comunicación pero es muy poco seguro según el número crece.

- Autenticación mediante firmas digitales. En este caso se debe aplicar los procedimientos y estrategias ya citadas en la sección anterior. Es mucho más complejo y difícil de gestionar pero más seguro para muchos participantes y el habitual a la hora de mantener grandes redes privadas virtuales a través de Internet.

Una vez han seleccionado cómo se van a autenticar deben hacerlo para poder seguir adelante en la creación de la Asociación de Seguridad y, una vez completada, ya están listos para el intercambio seguro de tráfico que tenga las propiedades de seguridad (integridad, privacidad, autenticación) pactadas.

Otros protocolos criptográficos bien probados hoy en día son, por ejemplo, los que tienen como objetivo proporcionar un correo electrónico más seguro, como PGP (*Pretty Good Privacy*) o S/MIME *Secure Multipurpose Internet Mail Exhange*). Aunque es cierto que cumplen bastante bien sus objetivos su uso no está muy extendido pues, para que así fuera, debería de cambiarse el esquema general de uso del protocolo de correo SMTP y su actual implantación mundial.

8.6.3. Redes privadas virtuales

Una red privada virtual (RPV) es, simplemente, una conexión segura entre dos o más partes de una red privada, creada a través de una red pública y considerada insegura. Tienen tres usos principales:

- Como una **Intranet** de la organización, sirve para conectar piezas disjuntas de la misma red privada. En este caso las comunicaciones seguras son, habitualmente, de red a red y no llega hasta los equipos individuales de cada una de las redes disjuntas.

- Como una **Extranet**, una modificación simple del caso anterior en la parte técnica pero compleja en la parte de seguridad. Parte de las redes disjuntas unidas mediante la RPV no pertenecen a la misma organización sino a otra o a otras que colaboran con la primera.

- Como una **red de usuarios móviles** de una organización que trabajan en su casa o que cambian constantemente de ubicación a causa de su trabajo pero que consiguen una conexión segura con la red de su oficina sin más que usar una conexión RPV. Este tipo de RPV mantiene la comunicación segura entre el equipo del usuario y el punto de entrada a la red de la oficina.

La forma en la que se da este tipo de comunicación segura es con protocolos criptográficos y algunos de los más utilizados hoy en día son SSL, IPSec o MPLS. Un caso muy común es aquel en el que la RPV, utilizando IPSec, entre los puntos de entrada a Internet (Fig. 8.11), gestiona el tráfico seleccionando las propiedades de seguridad a aplicar a cada tipo de mensajes. Esto implica el

poder de gestionar qué mensajes deben ir cifrados para mantener la privacidad, cuáles deben ser recibidos con gestión de la integridad, cuáles sin ningún tipo de cifrado, etc.

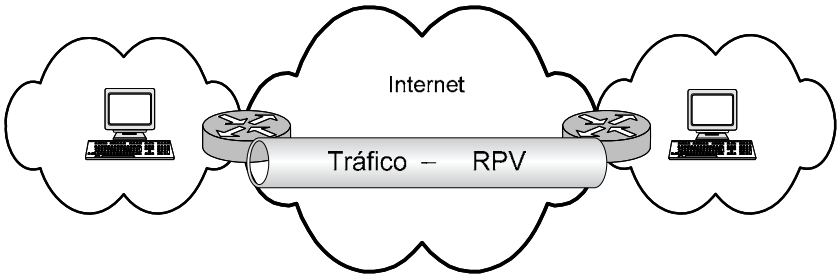


FIGURA 8.11. Red Privada Virtual típica a través de Internet.

Además de las citadas características de seguridad las redes privadas virtuales suelen tener también otros componentes:

- Componente de autenticación por usuario o de control de acceso. Sirve para proporcionar autenticación a nivel de usuario mediante sistemas como los AAA añadida a la conseguida para los puntos intermedios mediante IPSec.
- Componente de auditoría y registro de actividades, para asegurar el funcionamiento correcto y capacidad de recuperación.
- Componente de calidad del servicio, para asegurar un buen rendimiento, que no haya degradación de los recursos, especialmente para ciertos tipos de tráfico.

8.7. SEGURIDAD EN PROTOCOLOS INDUSTRIALES

A continuación se describen algunas de las funciones de seguridad definidas en una serie de normas estándar de protocolos de comunicaciones industriales

8.7.1. OPC

La fundación OPC (*Open Process Control*) define, dentro de las normas OPC de comunicaciones industriales, una serie de interfases opcionales de

seguridad para los objetos OPC. Estas especificaciones tratan de controles de acceso y se basan en el modelo de seguridad de Windows de Microsoft. Un servidor OPC puede implementar uno de entre tres posibles niveles de seguridad:

- Seguridad deshabilitada: no se obliga a cumplir ningún criterio de seguridad.

- Seguridad DCOM: Los permisos de ejecución y de acceso al servidor OPC se dan sólo para un cierto conjunto de clientes. Éste es el nivel de seguridad proporcionado por el modelo de COM distribuido de Microsoft y se configura siguiendo el modelo de credenciales de seguridad del entorno de Windows.

- Seguridad OPC: En este caso el servidor OPC sirve como monitor de referencia para controlar el acceso a objetos de seguridad específicos de cada fabricante, que están gestionados por el servidor OPC. La norma, no obstante, no define ni sugiere qué objetos deben asegurarse. Si se implementa esta opción hay que configurar antes la seguridad DCOM para decidir quién accede a las interfases del servidor.

En esta especificación sólo se detallan controles de acceso y no se dice nada acerca de la confidencialidad o de la integridad de las transmisiones.

8.7.2. MMS

Otro grupo de especificaciones es la del MMS, *Manufacturing Messaging Specification*. Se trata de una norma del nivel de aplicación para la comunicación mediante mensajes a y desde dispositivos o PLCs en un entorno de fabricación gestionada informáticamente. Es ampliamente utilizado en el entorno de fabricación de automóviles y es la base del estándar IEC 61850. Es posterior al OPC y define no sólo las comunicaciones cliente/servidor sino las comunicaciones entre iguales (*peer-to-peer*).

MMS define una serie de características de control de acceso basadas en una autenticación simple mediante contraseñas. En la especificación no se dice nada sobre la confidencialidad o la integridad. El valor de la contraseña (que es opcional) se usa para controlar el acceso a los objetos MMS. Cada objeto contiene listas de control de acceso que concretan las condiciones en las que se permiten los servicios pedidos a los objetos. Las condiciones son identidad de

usuario (o de aplicación) y/o la contraseña mientras que los servicios suelen ser crear, leer, escribir, ejecutar, borrar o modificar.

En MMS se asume la red de comunicación es Ethernet y las mayor parte de las implementaciones actuales asumen la pila de comunicaciones TCP/IP. Se puede incluir, asimismo, seguridad de cifrado mediante protocolos SSL o IPSec.

8.7.3. IEC 61850

La IEC (*Internacional Electrotechnical Commision*) especifica en su norma 61850 las entidades software, modelos de datos, servicios, protocolos y formatos de datos para la automatización de subestaciones de las redes de potencia eléctrica. Desde el punto de vista que nos ocupa son especialmente relevantes las opciones de control de acceso y la asociación de protocolos de comunicaciones, que determinan las características de seguridad de la red.

El IEC 61850 estipula que los nodos deben proporcionar control de acceso basado en la identificación de nodo (para la comunicación de máquina a máquina) y en la autenticación de usuario y de control de acceso al sistema para usuarios..

Las asociaciones del IEC 61850 están directamente relacionadas con las del MMS. El IEC 61850 no se refiere especialmente a seguridad de comunicaciones sino que se refiere al trabajo sobre seguridad genérica realizado por el IEC. Siempre que IP sea parte de la pila de protocolos de comunicaciones, se puede emplear los servicios de seguridad de IPSec, por ejemplo.

8.7.4. IICP

La IEC ha desarrollado también la norma ICCP (*Inter-Center Control Protocol*), que es un estándar para la comunicación en redes de área amplia entre centros de la red de transmisión eléctrica. Es similar a la norma de OPC pero no está ligada a ningún sistema operativo en particular.

La mayor parte de las implementaciones del ICCP funcionan sobre MMS y TCP y, como consecuencia, debe implementarse las funciones de control de acceso de MMS. Para proteger la comunicación entre centros se recomienda el uso del protocolo SSL. Con la intención de hacer uso de SSL incluso en implementaciones sin TCP de ICCP, basadas en la pila ISO, se define también la encapsulación de mensajes SSL en estos paquetes ISO, no TCP.

8.7.5. Recomendaciones de políticas de seguridad

Se muestran a continuación una serie de normas y organismos involucrados en el desarrollo de políticas de seguridad para entornos de comunicaciones industriales:

- IEEE 1402: Es la norma de seguridad de subestaciones, especialmente dedicada la seguridad física.

- *Process Control Security Requirements Forum* (PCSRF): Puesta en marcha por el NIST, tiene como objetivo crear conjuntos de normas de seguridad en la creación de nuevos sistemas de control de procesos. Ha desarrollado ya normas para sistemas DCS o SCADA como la SCP-ICS (*Industrial Control System Security Capabilities Profile*) y la SPP-ICS (*Industrial Control System System Security Profile*).

- ISA SP99: El comité SP99 del ISA (*Instrumentation Systems and Automation*), conocido como "*MAnufacturing and control systems security*", intenta crear documentos de guía para introducir la seguridad informática en los sistemas de control automatizado. Tiene ya dos guías publicadas sobre seguridad en sistemas PLCs, SCADA, etc.

- IEC TC65: El subcomité técnico 65C del IEC lleva desde 2004 trabajando sobre normas de seguridad para buses de campo y otras redes de comunicación.

8.8. INTRODUCCIÓN AL CONCEPTO DE CALIDAD DE SERVICIO EN REDES

El término Calidad de Servicio hace referencia a una serie de técnicas y métodos cuyos objetivos son asignar distintas prioridades a distintos tipos de tráfico en la red e implantar políticas que gestionen la latencia y las necesidades de ancho de banda. Todas estas técnicas simplemente tratan de que el tráfico correspondiente a las aplicaciones más críticas de la red tenga un trato preferencial en todos los puntos de la red por los que viaje y se han convertido en un mecanismo esencial para asegurar el correcto rendimiento de las aplicaciones en la red.

Para conseguirlo es necesario realizar **una monitorización constante** en tiempo real, una **ingeniería cuidadosa del tráfico** de la red y, aprovechando todo lo anterior, se puede hacer una buena **planificación** para obtener los rendimientos requeridos por aplicación. Cada una de ellas requerirá un cierto nivel de servicio y el los objetivos se pueden resumir en:

- Que todo el tráfico de la red alcance sus niveles de servicio. Esto se traduce, por ejemplo, en mantener el tráfico de voz sobre IP con una latencia menor de un cierto valor para obtener una buena calidad.

- Que, durante los periodos de congestión en la red, el tráfico más importante disfrute de los recursos que necesita siempre.

Como un beneficio añadido estas técnicas permiten optimizar el uso de los recursos de la red, permitiendo dilatar en el tiempo lo más posible la necesidad de gastar más dinero en añadir recursos.

Hoy en día todas las técnicas de Calidad de Servicio en redes se usan de manera más completa en redes grandes, con multitud de servicios diferentes y suele ser la “última milla” la parte de la red en la que se hace más énfasis.

8.9. APROXIMACIONES ESTÁNDAR A LA CALIDAD DE SERVICIO EN REDES

Para entender mejor qué aproximaciones estándar se realizan se va a detallar la secuencia lógica de los pasos a dar para implantar unas políticas de Calidad de Servicio en redes.

En primer lugar se ha de realizar la **clasificación del tráfico**. Hay que identificar cada servicio por su tráfico correspondiente: esto se hace mediante el reconocimiento de los flujos de tráfico del servicio correspondiente. Los dispositivos de la red deben inspeccionar cada paquete que reciben y buscar en él marcas que indiquen la clase de servicio del paquete. Una vez reconocida una marca se aplican los mecanismos apropiados al paquete.

Existen diversas aproximaciones para hacer esta clasificación. Entre ellas las más extendidas son:

- La impulsada por el DMTF (*Distributed Management Task Force*) que consiste en un servicio de directorio de aplicación de políticas para el acceso a recursos de la red. Crea una lista de prioridades para el tráfico de la red que van desde la 0 (menos prioritaria) que se aplica al tráfico por defecto (como un fichero de usuario o el correo electrónico) hasta la 7 (máxima prioridad) que se aplica al tráfico de gestión de red, pasando por distintos tipos de tráfico sensible, como el tráfico de voz con unas necesidades de latencia menor de 10 ms, prioridad 6, tráfico de video con necesidad de latencia menor de 100 ms, prioridad 5, etc.

- La clasificación de Servicios Integrados (*Intserv*), avalada por el IETF, que usa el protocolo RSVP para reservar recursos de red (como ancho de banda o latencia) para flujos de tráfico concretos. Los problemas asociados con una correcta gestión del RSVP y de la sobrecarga de recursos usados en mantener el estado de cada flujo de la red hacen que esta aproximación sea aún poco utilizada.

- Otra aproximación propuesta por el IETF y mucho más extendida es la conocida como de Servicios Diferenciados (*Diffserv*). En ella cada paquete IP se clasifica en su entrada a la red mediante el campo de "tipo de servicio" (TOS, "*Type Of Service*") de la cabecera IP del paquete como se analiza más adelante. Basándose en este campo se aplicará paquete a paquete el tratamiento de calidad de servicio. Una vez establecido el TOS del paquete, el paquete se encola y se procesa usando un mecanismo conocido como "*Weighted Random Early Detection*" o WRED, que se analiza más adelante.

Tras la clasificación del tráfico hay que "marcarlo". El **marcado de paquetes** consiste en codificar la clasificación del servicio dentro del propio paquete con el objetivo de que cada dispositivo de la red pueda identificar su clase. Normalmente este paso y el anterior se realizan en el borde de la red, en su entrada, aunque, además, un paquete podría volver a ser marcado en su viaje por la misma red.

En redes conmutadas (nivel 2 de OSI) las tramas se marcan utilizando el estándar IEEE 802.1p, que define un valor de 3 bits para asignar hasta 8 clases de prioridades diferentes a cada trama, que se inserta en la parte de la trama IEEE 802.1Q, la norma Ethernet para LAN virtuales.

En redes IP (nivel 3 de OSI) se codifica el citado campo TOS con uno de los valores siguientes:

- El DSCP (*Differentiated Services Code Point*) que puebla los primeros 6 bits del TOS y especifica el comportamiento por salto a aplicar al paquete. Es importante señalar que, en la actualidad, no todos los fabricantes de encaminadores tienen soporte de DSCP.

- La precedencia IP, campo de 3 bits en el TOS, con valores de 0 a 7 dependiendo del tráfico. Esta aproximación cada vez se usa menos siendo sustituida por la del DSCP.

- El propio campo TOS, codificado con valores de 0 a 15, también en fase de extinción y sustituido por el DSCP.

Finalmente se llega a la fase de aplicación de la política de calidad de servicio, que consiste en forzar el tratamiento de los paquetes basándose en su clasificación y en las condiciones de la red. Tanto los paquetes entrantes a la red como los salientes sufren la aplicación mediante diferentes mecanismos. En los periodos de congestión de la red el tráfico de baja prioridad sufre más retrasos en beneficio del correspondiente a servicios de prioridad mayor. Entre los mecanismos más utilizados se deben señalar:

- *Traffic Shaping*. Si se detecta congestión, se reduce la cantidad y el ritmo de paquetes entrantes y salientes correspondientes a ciertos flujos concretos. Tales paquetes pueden ser encolados o descartados. Suele aplicarse a nivel TCP. Para que sea una técnica eficaz debe usarse uniformemente a través de toda la red.

- *Queuing* o encolamiento, que consiste en asignar paquetes a diferentes colas para su proceso, dependiendo de su clase. A su vez permite utilizar diferentes aproximaciones. En una de ellas, "*packet dropping*", al llenarse una cola no permiten la entrada de nuevos paquetes resultando estos descartados. Otra es "*Weighted Fair Queuing*" que permite reservar distintos anchos de banda a cada una según reglas definidas por el administrador. Otra más es "*priority queuing*" que asigna prioridades a las colas y las procesa en eses orden, desde la de mayor prioridad a la de menor prioridad.

- *Random Early Detection*, que se desarrolló originalmente para los encaminadores del núcleo de Internet.

  Es una forma de control de congestión que monitoriza las colas de paquetes en el encaminador y descarta paquetes cuando una cola se

llena. Puede provocar una pérdida excesiva de paquetes, provocando un número muy grande de retransmisiones de los mismos y degradando el rendimiento, por lo que normalmente se utiliza en la versión ya citada del WRED.

8.10. LOS ENCAMINADORES Y LA CALIDAD DE SERVICIO

La forma en la que se especifican los parámetros y reglas que controlan la calidad de servicio es mediante herramientas software y dispositivos que aplicarán las diferentes políticas.

En muchos casos la calidad de servicio se gestiona en cada dispositivo, caso a caso. En otros, más ambiciosos y más difíciles de gestionar, se hace una administración a nivel de red utilizando habitualmente herramientas de gestión de red, como el HP *OpenView* o el *Cisco Works*, que, además de funciones de gestión general y de monitorización, permiten aplicar las diferentes políticas y metodologías mencionadas,

Aunque no son los únicos dispositivos en los que se implementan las distintas fases de puesta en marcha de políticas de calidad de servicio, los encaminadores de la red, especialmente los que están en las “fronteras” de la red, son los dispositivos más frecuentemente utilizados. En ellos se calculan las estadísticas de calidad de servicio usando los datos en tiempo real de los paquetes que procesan. Hay que tener en cuenta que la implementación de las características analizadas previamente obliga a usar una serie de recursos en los encaminadores, que, si no se ha previsto adecuadamente, podrían provocar un rendimiento incorrecto del propio encaminador, tanto en su función como implementador de las funciones citadas como en sus funciones básicas de encaminamiento de paquetes en la red.

Además los encaminadores son, en general, inflexibles en la redistribución de recursos a servicios cuando hay cambio de condiciones. De hecho cualquier buena puesta en marcha de una estrategia de calidad de servicios pasa por que la configuración de los encaminadores de la frontera de red esté coordinada con la de los encaminadores del proveedor de servicio WAN o del proveedor de servicio de Internet.

Además de las funciones citadas en los apartados anteriores algunos encaminadores pueden implementar otro tipo de características que ayudan en la

mejor puesta en marcha de una política de calidad de servicio global para la red. Entre ellas se ha de citar:

- El balanceo de carga, que puede servir, correctamente implementado, para aliviar cuellos de botella, especialmente en los encaminadores de los extremos de la red.

- La posibilidad de mantener memorias de páginas web o documentos especialmente frecuentados, lo que se traduce en un incremento del rendimiento del tráfico de la red.

- La capacidad de detectar congestión en la WAN a la que están conectados, como el caso de *Frame Relay* y los encaminadores que detectan los bits FECN (*Forward Explicit Congestion* Notification) y BECN (*Backward Explicit Congestion* Notification) en la cabecera de nivel 2 de los mensajes que vienen de la red *Frame Relay*.

8.11. CONCLUSIONES

Se ha hecho una exposición de los problemas de seguridad más habituales que suelen aparecer en las redes de comunicaciones industriales, que son, esencialmente, los mismos de cualquier red actual de ordenadores, analizando por separado tanto problemas comunes relacionados con los sistemas físicos como problemas derivados de la naturaleza software de los sistemas operativos y aplicaciones y problemas de organización relacionados con la “ingeniería social”.

Se ha expuesto también el arsenal de herramientas y metodologías de defensa actuales, haciendo especial énfasis en la importancia de usarlos dentro de un proceso de seguridad concreto, dirigido mediante una política de seguridad cuidadosamente elaborada, algo que debe hacer ver a los alumnos cómo deben utilizarse todas las herramientas citadas.

Es especialmente importante señalar el análisis de las propiedades de seguridad fundamentales y de los algoritmos matemáticos utilizados para poder implementar las propiedades, así como los protocolos criptográficos que los usan.

Resulta para el alumno igual de importante hoy en día conocer los rudimentos de los métodos usados para poder conseguir una calidad de servicio

más que obligatoria para las redes de comunicaciones industriales y, en este sentido, se han analizado las tareas más significativas de su puesta en marcha, así como los estándares y normas más utilizados para ello.

8.12. BIBLIOGRAFÍA

W. Cheswick y W. Bellovin. Firewalls and Internet security. Ed. Addison-Wesley, 2003.

G. Díaz, F. Mur, E. Sancristóbal, M. Castro y J. Peire. Seguridad en las comunicaciones y en la información. Ed. UNED, 2004.

S. Garfinkel y G. Spafford. Practical UNIX and Internet security. Ed. O'Reilly, 1999.

M. Liotine. Mission-critical network planning. Ed. Artech House, 2003.

J. Marcelo. Riesgo y seguridad de los sistemas informáticos. Ed. Univ. Polit. Valencia, 2003.

B. Schneier. Applied cryptography. Ed. Wiley, 1996.

B. Schneier. Secrets and lies. Ed. Wiley, 2000.

W. Stallings. Cryptography and network security: Principles and practice, Ed. Prentice Hall, 2003.

D. Dzung, M. Naedele, T.P. Von Hoff y M. Crevatin, Security for Industrial Communication Systems, Proceedings of the IEEE, vol. 93, nº 6, Junio de 2005

8.13. EVALUACIÓN

8.13.1. Evaluación objetiva

1. ¿Cuál de los siguientes sería un algoritmo correcto para un sistema criptográfico en el que se buscase la privacidad de los mensajes entre diferentes equipos de una red? La distancia típica de trabajo de los buses de campo es:

   A) El algoritmo MD5

   B) El algoritmo RSA modificado para gestión de recursos

   C) El algoritmo AES

   D) EL algoritmo SSH

2. En un cortafuegos de tipo filtro de paquetes en el que se desea filtrar mensajes de correo electrónico, ¿cuál de las siguientes parejas de criterios serían válidos?

   A) La dirección IP destino del mensaje y el tipo de mensaje SMTP

   B) El número de puerto destino y el tipo de protocolo

   C) Las direcciones IP origen y destino y el campo de HELO de SMTP

   D) Los campos HELO y DATA de la cabecera de SMTP del mensaje

3. En una comunicación segura entre dos equipos mediante IPSEC, que utilice ESP en modo túnel y cifrando mediante DES, ¿qué campos de cabeceras del mensaje viajan cifrados?

   A) Todos los campos de las cabeceras del mensaje original

   B) Todos los campos de las cabeceras menos los de la cabecera de IP del mensaje original

   C) Sólo la cabecera de datos del mensaje

   D) Ninguno. Esto solo sucede en mensajes ESP en modo transporte

4. Dentro del diseño de una política de calidad de servicio para redes se llega a la conclusión de que se necesita separar y clasificar el tráfico en 12 categorías diferentes. ¿Cuál de las siguientes aproximaciones permitiría tal diseño?

   A) La basada en marcas IEEE 802.1p en paquetes Ethernet

   B) La basada en DSCP dentro de la cabecera IP de los mensajes

   C) La basada en el campo TOS dentro de la cabecera IP de los mensajes

   D) La basada en el campo TOS dentro de la cabecera Ethernet de los mensajes

5. Si en una red se utiliza como protocolo de encaminamiento OSPF en su configuración en lugar de RIP, ¿Se puede afirmar que sólo por este cambio de protocolo la red es más segura?

   A) No, dependerá de si se configuran o no las características de autenticación de OSPF o no

   B) Si, ya que la red responderá automáticamente a ataques a su privacidad

   C) No, RIP es más seguro siempre que OSPF

   D) Si, siempre que el número de encaminadores no supere el máximo de OSPF

6. Dentro de las normas de acceso físico de una política de seguridad de redes, ¿cuál de las siguientes debería ser obligatoria?

   A) La que gestiona el acceso de un cliente a su ordenador personal

   B) La que gestiona el acceso de un empleado móvil al ordenador de su cliente

   C) La que gestiona el acceso de la policía al entorno físico

   D) La que gestiona el acceso de cualquier persona a los servidores más importantes de la organización

7. ¿Qué diferencias esenciales hay entre las cuestiones a tener en cuenta para la seguridad del software de los dispositivos de comunicaciones, como los encaminadores, con respecto al caso de los servidores?

   A) La mayor complejidad de tal software

   B) Ninguna, en ambos casos el planteamiento general es el mismo

   C) La menor complejidad de tal software

   D) La necesidad de un monitor de referencia de seguridad más complejo

8. Se ha descubierto una vulnerabilidad en una aplicación importante que funciona sobre el sistema operativo de un servidor y se dispone del "parche" adecuado. ¿Cuál de los siguientes sería un posible problema de seguridad que podría derivarse de la instalación de tal parche?

   A) Ninguno, la instalación soluciona todos los problemas

   B) La incompatibilidad entre el "parche" y la aplicación

   C) En muchos casos hay que "parchear" el "parche", que introduce otros problemas no esperados

   D) Necesidad de hacerlo en un entorno físico abierto

9. Se pretende rastrear el origen de una serie de ataques DDOS que están sufriendo una serie de servidores de la organización. ¿Cuál de las siguientes herramientas sería más útil para la tarea?

   A) Un sistema de detección de intrusiones en la red

   B) Un detector de vulnerabilidades de sistemas operativos

   C) Un sistema anti-*spyware* en cada ordenador de la red

   D) Una red privada virtual configurada dentro de la LAN de la organización

10. ¿En qué tipo de redes se suele utilizar el estándar IEEE 802.1p como método de clasificación de paquetes?

    A) En redes ATM, dentro de la cabecera básica de la celda

    B) En redes IP, en las cabeceras de IP, en cualquier tipo de red

    C) En redes inalámbricas WLAN

    D) En redes conmutadas en las que los conmutadores investigan la cabecera Ethernet

TEMA 9

INTRODUCCIÓN A LAS COMUNICACIONES INDUSTRIALES Y A LOS SISTEMAS DE TIEMPO REAL

9.1. Introducción y orientaciones para el estudio
9.2. Objetivos
9.3. Características de las redes de comunicación industrial
9.4. Modelo de integración de redes de comunicación
9.5. Comunicación y sistemas de tiempo real
9.6. Sistemas disparados por tiempo y disparados por evento
9.7. Planificación de mensajes de tiempo real
9.8. Prioridades estáticas y dinámicas
9.9. Análisis de tiempo real en redes de comunicación
9.10. Conclusiones
9.11. Bibliografía
9.12. Evaluación

9.1. INTRODUCCIÓN Y ORIENTACIÓN PARA EL ESTUDIO

En este capítulo se expondrán los conocimientos básicos sobre dos componentes fundamentales en los sistemas industriales, como son los sistemas de comunicación y los sistemas de tiempo real.

La evolución en la automatización industrial ha ido a la par con el desarrollo tecnológico. Específicamente la introducción de tecnologías de la comunicación permitió pasar de los clásicos sistemas centralizados, típicos en la década de los setenta basados en equipos de altas prestaciones y de coste elevado, a los actuales sistemas distribuidos basados en equipos más sencillos conectados a través de redes de comunicación.

Los sistemas distribuidos pueden adquirir diferentes arquitecturas atendiendo a las necesidades de automatización de una planta. Así pues, se pueden encontrar instalaciones conectadas a través de una única red de comunicación dedicada, hasta sistemas más complejos formados por la combinación de dos o más redes de comunicación. Esto último ha dado lugar al establecimiento de una jerarquía de comunicaciones en el ámbito industrial, que obliga a analizar cuidadosamente los distintos requerimientos y prestaciones de cada nivel atendiendo a la función que realiza. Esta jerarquía de comunicaciones es abordada en detalle en la primera parte de este capítulo.

Por otra parte, los sistemas distribuidos y en especial los sistemas distribuidos de control han pasado a ser el corazón de un amplio grupo de importantes sectores de nuestra sociedad, tales como, sector industrial, sector del transporte, defensa, etc. En estos sistemas, un conjunto de acciones son llevadas a cabo por un número de ordenadores y dispositivos interconectados a través de una infraestructura de comunicación formando un lazo de control sobre un proceso dado. Básicamente, el sistema hace una medición de ciertas entradas y genera las salidas necesarias las cuales son aplicadas al proceso a través de dispositivos de actuación. Todas estas acciones deben ser ejecutadas en intervalos específicos de tiempo, es decir, existen restricciones temporales de funcionamiento. En los sistemas distribuidos de control un retardo en una salida puede traer más o menos graves consecuencias dependiendo de su aplicación. Esto significa que las salidas generadas deben ser correctas tanto lógicamente como en el tiempo, lo que lo convierte en un sistema de tiempo real.

Así, los sistemas distribuidos son un tipo de sistema de tiempo real. En ellos, los principales problemas a dar solución son los derivados a la planificación de las diferentes acciones a realizar sobre los diferentes nodos o equipos que se comunican a través de una red, de tal forma que las restricciones temporales sean cumplidas en todo momento.

Por lo tanto este capítulo se centra también en el estudio y el análisis de los conceptos y la teoría básica de los sistemas de tiempo real y su aplicación a los sistemas de comunicación industrial.

9.2. OBJETIVOS

Este capítulo persigue varios objetivos, en primer lugar se pretende estudiar y analizar los sistemas de comunicación industrial y el modelo de integración de las redes de comunicación en el entorno industrial.

En segundo lugar, se proporcionan los conocimientos básicos de sistemas de tiempo real (tipos de sistemas de tiempo real, restricciones temporales, algoritmos y políticas de planificación) y su aplicación a los sistemas distribuidos de tiempo real.

Al finalizar este capítulo se debe ser capaz de poder dar respuesta a los siguientes puntos:

- Componentes y características de los sistemas distribuidos de tiempo real.

- Diferentes niveles del modelo de integración de redes de comunicación industrial y sus características.

- Conceptos básicos de tiempo real, tales como: sistemas disparados por eventos o por tiempo, sistemas críticos y no críticos, planificación de tares, planificación con desalojo y sin desalojo, planificación estática y dinámica.

- Características y funcionamiento de los principales algoritmos de planificación.

9.3. CARACTERÍSTICAS DE LAS REDES DE COMUNICACIÓN INDUSTRIAL

Los últimos años han sido testigos y partícipes del vertiginoso desarrollo tecnológico en los campos de los sistemas informáticos y de la electrónica. La automatización en la industria[1] ha seguido un proceso gradual, aplicando la tecnología disponible en cada momento. Esto ha dado lugar a las denominadas 'islas automatizadas', término empleado para designar a una serie de equipos aislados entre sí y dedicados al control de una máquina o parte de un proceso. Estos equipos pueden ser ordenadores de diseño y gestión, actuadores, sensores, PLCs, controles numéricos, etc.

Una de las características importantes de los nuevos sistemas de producción es la capacidad que poseen muchos de sus componentes de generar información relacionada con el proceso que llevan a cabo en el sistema productivo. Hasta hace algún tiempo no se había sentido la necesidad de que los datos y recursos manejados por un equipo fueran útiles y necesarios para otros equipos del sistema. Sin embargo, los requisitos que se exigen a los nuevos sistemas de producción hacen que la generación y el tratamiento de esta información sea totalmente necesaria, lo que a su vez exige vías de comunicación entre los diferentes dispositivos inteligentes que intervienen en el proceso.

El desarrollo de las comunicaciones y su aplicación a la industria, ha permitido la implantación de redes industriales que facilitan la comunicación entre estas islas automatizadas. Por todo ello, las comunicaciones constituyen un elemento fundamental en los nuevos entornos de fabricación.

En general los objetivos fundamentales de los sistemas de comunicación son:

- Intercambio de datos, tal como transferencia de archivos entre ordenadores, envió de correo electrónico, transmisión de voz e imágenes, etc.

- Compartir recursos de tal manera de aprovechar al máximo equipos o periféricos como impresoras, unidades de almacenamiento, etc.

---

1 A lo largo de este capítulo, el término industria y el término fabrica se utilizarán de forma indistinta para identificar el lugar físico donde se produce algún objeto, material o servicio.

Centrándose en el ámbito industrial, los sistemas de comunicación tienen como objetivos:

- Coordinar acciones de unidades automatizadas y controlar la transferencia de componentes, a través del intercambio de datos entre las diferentes unidades (autómatas programables o PLCs, PCs industriales) que controlan el proceso productivo.

- Monitorizar y modificar estrategias de control desde el puesto de operación, que puede estar situado en la propia planta o en cualquier otro lugar mediante una conexión a través de redes de datos públicas o privadas.

- Proveer los recursos necesarios para aumentar la confiabilidad y seguridad en los procesos de producción mediante: detección temprana de condiciones de alarma, supervisión y control continúo de procesos de alto riesgo, verificación del estado de las instalaciones y seguimiento de las condiciones de operación de estaciones remotas.

- Proveer servicios de transmisión de voz e imágenes.

- Integración completa del proceso productivo (desde el operario hasta los gestores o clientes)

Como se puede observar, desde el punto de vista industrial, la necesidad de comunicación no se restringe únicamente a la producción.

Diferentes departamentos de la industria pueden participar en la red de comunicaciones para permitir un control global del sistema.

De este modo, no sólo se controla el propio funcionamiento de la planta de fabricación, sino que en función de las decisiones tomadas en las capas administrativas de la empresa, podría actuarse directamente sobre la producción.

Por lo tanto, la red integrada de comunicación industrial debe estructurarse en base a una arquitectura bien definida y bajo las premisas de racionalización, conectividad, calidad y confianza, lo que se estudiará en detalle en la siguiente sección.

9.4. MODELO DE INTEGRACIÓN DE REDES DE COMUNICACIÓN

9.4.1. Integración de la red de comunicaciones

La integración de los diferentes equipos y dispositivos existentes en una industria se hace dividiendo las tareas entre grupos de procesadores con una organización jerárquica. Así, dependiendo de la función y el tipo de conexiones, se suelen distinguir cinco niveles en una red industrial, los cuales se pueden observar en la Figura 9.1.

- **N1 - Nivel de entrada/salida**: es el nivel más próximo al proceso. Este nivel esta constituido básicamente por unidades de captación de señales, de actuación y de entrada/salida de datos del proceso o de un operador local. Normalmente estos tipos de redes se caracterizan, por tener mensajes con una longitud reducida (a través de los cuales se transfieren medidas, acciones de control, etc.), por la fiabilidad e integridad de los mensajes, por la eficiencia del protocolo utilizado, por la velocidad de transmisión y por la técnica de acceso al medio de los dispositivos participantes.

- **N2 - Nivel de campo**: integra pequeños automatismos (PLCs compactos, PIDs, multiplexores de e/s, etc.) en subredes o 'islas'. En el nivel más alto de estas redes se puede encontrar uno o varios autómatas modulares actuando como maestros de la red o maestros flotantes. En estos dos primeros niveles se emplean los denominados buses de campo[2]. Los buses de campo constituyen el nivel más simple y próximo al proceso dentro de la estructura de las comunicaciones industriales, los cuales serán estudiados en detalle en el siguiente capítulo.

- **N3 - Nivel de control de proceso**: Este nivel esta constituido por unidades de control (con CPU y programas propios) tales como autómatas, reguladores de procesos, controladores de robots, controladores numéricos, etc., los cuales se encargan del control

[2] Un bus de campo es, en líneas generales, un sistema de dispositivos de campo (sensores y actuadores) y dispositivos de control, que comparten un bus digital serie bidireccional para transmitir información entre ellos.

automático de ciertas partes de la planta. La integración en red de estas unidades permite el intercambio de datos e información útil para el control global del proceso. En este nivel es donde se suelen emplear las redes de tipo LAN (MAP o Ethernet).

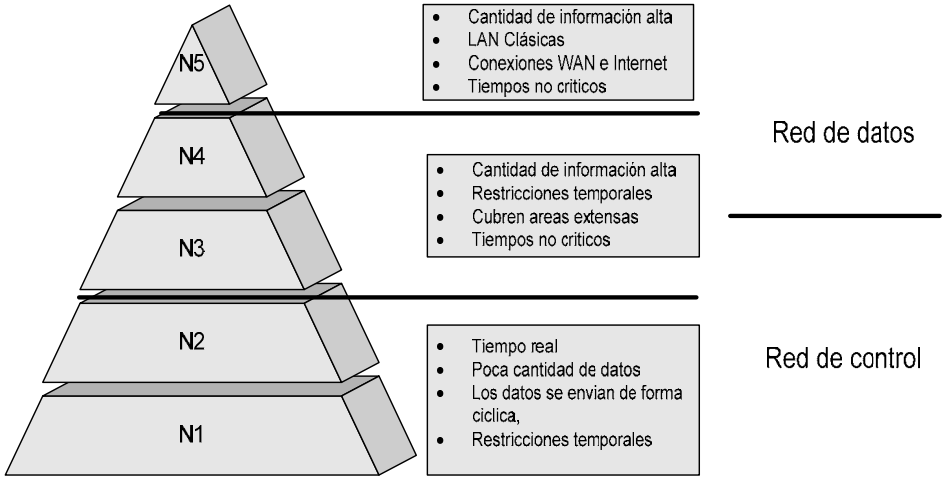


FIGURA 9.1. Jerarquía de comunicaciones industriales.

- **N4 - Nivel de control de producción**: Este nivel incluye una serie de unidades destinadas al control global del proceso, tales como ordenadores de proceso, terminales de diálogo, terminales de enlace con otros departamentos de la empresa, etc. Desde estas unidades se tiene acceso a la mayor parte de las variables del proceso, generalmente con el propósito de supervisarlas, presentarlas, registrarlas y/o almacenarlas, cambiar consignas, alterar programas y obtener datos para su posterior procesamiento.

- **N5 - Nivel de gestión o dirección:** Este nivel incluye la comunicación con ordenadores de gestión y se encarga del procesamiento de los datos, obtenidos en el nivel anterior, y su uso en análisis estadísticos, control de fabricación, control de calidad, gestión de existencias y dirección general. En algunos casos, las unidades de este nivel pueden disponer de conexiones a redes más amplias de tipo WAN propietarias y/o estándares de difusión de Internet.

El conjunto de dispositivos y redes de comunicación a nivel de entrada/salida, campo y de control de proceso dio lugar a la denominación de

Proceso Integrado por Computador o CIP (*Computer Integrated Process*), mientras que los sistemas y redes en los niveles de control de producción y de dirección acuñaron el nombre de Fabricación Integrada por Computador o CIM (*Computer Integrated Manufacturing*).

La arquitectura CIM fue derivada del proyecto del protocolo de automatización y fabricación (MAP), emprendido por un grupo de compañías norteamericanas conducida por General Motors y comenzado en el año 1980. El propósito principal era especificar un estándar abierto para el sistema de comunicación de la fábrica que permitiría interoperabilidad entre los componentes de diferentes vendedores.

Cada uno de los niveles de CIM, además de llevar a cabo tareas específicas, realiza un tratamiento y filtrado de la información que es transmitida en sentido ascendente o descendente a través de la pirámide. De esta forma se limita el flujo de información a los estrictamente necesarios en cada nivel. También existe un tráfico en sentido horizontal dentro de cada nivel, con distintas distinciones en cada uno de ellos.

Esta estructura no es universal, varía con el tamaño del sistema de fabricación y sus características particulares. Además, para cualquiera de los niveles, no hay un estándar universalmente aceptado que cubra todos los aspectos desde el nivel físico al de aplicación (referido al modelo OSI de ISO).

9.4.2. Redes de control y redes de datos

En el esquema piramidal presentado en el apartado anterior (Fig. 9.1), existen diferentes niveles de comunicación, cada uno de ellos con diferentes necesidades. Se pueden distinguir básicamente dos tipos de redes: redes de control y redes de datos o de ofimática. Las redes de control están ligadas a la parte baja de la pirámide, mientras que las redes de datos están ligadas a las partes alta de la pirámide.

Las redes de datos están orientadas al transporte de grandes paquetes de datos, que aparecen de forma esporádica. En este nivel la velocidad de transferencia de información puede ser muy alta del orden de los 100 Mbps, por lo que se requiere un amplio ancho de banda.

Las redes de control están orientadas al transporte de un gran número de pequeños paquetes, intercambiados con frecuencia entre un elevado número de equipos que forman la red y que la mayoría de veces trabajan en tiempo real. La

velocidad de transferencia de información es variable, pudiendo extenderse desde 300 bits por segundo hasta 20 Mbps.

9.5. COMUNICACIÓN Y SISTEMAS DE TIEMPO REAL

Realizar una tarea de forma correcta y a tiempo se ha convertido en un paradigma de la sociedad moderna. De hecho, en las sociedades modernas se está llegando a un punto donde una gran cantidad de acciones deben ser tomadas a tiempo para mantener los estándares de vida actual. Por ejemplo, en el campo de los sistemas industriales, nuevos productos deben ser diseñados y producidos para alcanzar el mercado a tiempo, plantas de producción necesitan reconfigurar la operación de la planta a tiempo para mantener los requisitos de cambio, el estado de una red de distribución de potencia debe ser monitoreada a una tasa que permita reaccionar a tiempo a picos de demanda, la concentración de la contaminación debe ser monitoreada para permitir acciones correctivas a tiempo, plantas químicas hacen uso de materiales tóxicos las cuales requieren un estricto monitoreo y control tal que cualquier fallo pueda ser detectada y las medidas de seguridad puedan ser tomadas a tiempo.

Todos estos ejemplos tienen en común que cuando las restricciones temporales en las cuales deben producirse las salidas no son satisfechas, pueden traer graves consecuencias económicas, sociales o ambientales.

9.5.1. Sistemas de tiempo real

Los sistemas de tiempo real son sistemas de computación que interaccionan repetidamente con su entorno físico y responden a los estímulos que reciben del mismo dentro de un plazo determinado. Para que el funcionamiento sea correcto no basta con que las acciones sean correctas, sino que tiene que ejecutarse dentro de un intervalo de tiempo específico.

Típicamente un sistema de tiempo real (Fig. 9.2), consiste en dos subsistemas que interactúan y cooperan entre ellos: un sistema a ser controlado (ambiente o entorno físico) y un sistema que controla (sistema de control).

Las interacciones entre los dos subsistemas están descritas por tres operaciones: muestreo, procesamiento y respuesta. Los subsistemas de computación continuamente procesan los datos muestreados del entorno físico y producen una apropiada respuesta que es enviada al entorno físico. Las tres

operaciones deben ser realizadas dentro de tiempos específicos; esto es lo que constituye las restricciones de tiempo sobre el sistema. Si una reacción ocurre demasiado tarde podría ser peligrosa para el sistema.

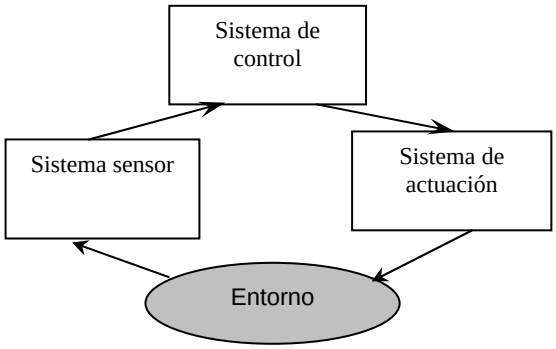


FIGURA 9.2. Entorno de un sistema de tiempo real.

Hoy en día, la computación de tiempo real juega un papel importante en nuestra sociedad, puesto que un gran número de sistemas complejos dependen en parte o completamente de sistemas de computación de tiempo real. Algunos ejemplos de aplicaciones que requieren computación de tiempo real son: control de automóviles, control de procesos de producción complejos, control de plantas nucleares y químicas, control de tráfico aéreo, sistemas multimedia, automatización industrial, sistemas de telecomunicación, sistemas de misiones espaciales, robótica, sistemas militares, etc.

La mayoría de los sistemas de tiempo real son componentes de otros sistemas, en los que realizan funciones de control, en este caso se habla de sistemas empotrados o embebidos (*embedded systems*).

9.5.2. Comunicación en tiempo real

Actualmente, a nivel de campo dentro de un proceso industrial es cada vez más frecuente encontrar sensores, controladores y actuadores interconectados por una red de comunicación formando un lazo de control[3]. Los intercambios de datos entre nodos se realizan a través de la red y, debido a la naturaleza de

[3] Esta arquitectura de control da lugar a lo que se conoce con el nombre de sistemas distribuidos de control en tiempo real.

tiempo real de las aplicaciones que funcionan en el sistema, tales intercambios deben finalizar antes de un plazo de tiempo dado. Esta clase de comunicación con restricciones de tiempo se conoce como *comunicación en tiempo real*. El conjunto de todos los nodos interconectados en la red, junto con los servicios y protocolos usados para permitir el intercambio correcto y a tiempo de los datos, conforman lo que se conoce como *sistema de comunicación en tiempo real*.

En el caso de sistema distribuido de control en tiempo real si las salidas correctas no son producidas a tiempo, podrían producirse excesivos retardos de procesamiento y de comunicación, y el proceso bajo control podría hacerse inestable y poner en peligro toda la integridad del sistema.

Los sistemas de comunicación de tiempo real constituyen el esqueleto o backbone de un sistema distribuido de tiempo real y por lo tanto estos afectan directamente el comportamiento temporal total del sistema. Además, es de hacer notar que la pérdida de comunicación da lugar a la pérdida de todos los servicios del sistema, por lo cual este es un componente crítico del sistema.

9.5.3. Comunicación en los sistemas distribuidos de control en tiempo real

El intercambio de datos en un sistema distribuido se realiza utilizando mensajes. Los mensajes contienen tanto los datos a transmitir así como la información de control necesaria para la transmisión apropiada de los datos desde el remitente hasta su destinatario. En general cada mensaje a transmitir se divide en paquetes de tamaño fijo, de forma que la comunicación entre dos tareas involucra la transmisión de un cierto número de paquetes. Este proceso de comunicación entre tareas se realiza en las siguientes etapas:

- *Generación y colocación en cola del mensaje*. La actividad que envía el mensaje debe crear el mensaje a transmitir y, en caso necesario, partirlo en paquetes de tamaño fijo. Finalmente, cada paquete debe ser puesto en la cola de transmisión.

- *Acceso al dispositivo de comunicación*. Una vez el paquete se encuentra en la cola, deberá esperar a que el dispositivo de comunicación quede libre y listo para realizar la transmisión del paquete.

- *Transmisión del mensaje por el enlace físico*. Desde el recurso procesador origen al recurso procesador destino. El tiempo de

transmisión de cada paquete vendrá determinado por su longitud y por la velocidad de transmisión del dispositivo.

- *Recepción y composición del mensaje*. Una vez que se hayan recibido todos los paquetes correspondientes se deberá formar el mensaje original y notificar a la tarea destino.

A la hora de analizar el sistema distribuido de tiempo real se deberá considerar el efecto de cada una de estas etapas en la respuesta completa. En muchos sistemas se pueden considerar los tiempos de generación y colocación en colas del mensaje como parte de la ejecución de la tarea emisora, de forma que se deberán sumar estos dos tiempos al tiempo de ejecución de la tarea origen. De igual forma, el tiempo de recepción y composición del mensaje se añadirá al tiempo de ejecución de la tarea destino. El impacto del acceso al dispositivo de comunicación y de la propia transmisión del mensaje es más complejo, ya que debe tenerse en cuenta la influencia de otros mensajes que quieran ser transmitidos por el mismo medio de comunicación.

Cuando se dispone de una red de comunicación apta para tiempo real, el acceso y la transmisión de un mensaje por una red de comunicación se puede analizar de una forma muy parecida a como se analiza el acceso y ejecución de una tarea en un procesador. Esto hace que, desde el punto de vista del análisis, no se distingue entre la ejecución de una tarea en un procesador o la transmisión de un mensaje por un dispositivo de comunicación.

Evidentemente, los recursos más importantes y sobre los que se centra la mayor parte del esfuerzo de investigación son el procesador y la red de comunicación. Dado que ambos recursos se pueden considerar con un comportamiento similar, se centrará el estudio en el procesador, esto es: bajo qué condiciones el procesador es asignado a un conjunto de tareas (o una red de comunicación a un conjunto de mensajes). Por lo tanto, a lo largo de este capítulo los términos tareas y mensajes serán utilizados de forma indistinta.

9.6. SISTEMAS DISPARADOS POR TIEMPO Y DISPARADOS POR EVENTOS

Tal y como se indicó anteriormente, un sistema distribuido de tiempo real interactúa con su entorno físico con el propósito de realizar un control sobre este. Estos sistemas están basados en uno o varios procesadores conectados

entre sí a través de un bus o una red de comunicación que interactúan con el entorno. Esta interacción se realiza mediante sensores que dan información del estado actual del sistema a controlar y mediante actuadores que pueden modificar o actuar sobre algún aspecto del mismo. Esto significa que el sistema de control debe estar enterado del estado del entorno físico en ciertos instantes de tiempo. Esto se logra manteniendo una estructura de datos dentro del sistema de cálculo que refleja el estado del entorno.

Puesto que el sistema de cálculo está distribuido, la base de datos está también distribuida entre los diferentes nodos del sistema. Esta base de datos se caracteriza por tener una validez en el tiempo, lo que significa que después de cierto intervalo del tiempo la misma no es válida; esta característica es conocida como exactitud temporal. La exactitud temporal puede ser llevada a cabo a través de dos aproximaciones: sistemas disparados por eventos (event-triggered systems) y sistemas disparados por tiempo (time-trigegered systems).

- **Sistemas disparados por eventos:** De acuerdo a la aproximación manejada por eventos, el sistema de control es notificado de algún cambio significativo en el estado del ambiente (evento externo) o en el estado interno del controlador (acontecimiento interno). A partir de la recepción de tal acontecimiento el sistema de control toma las acciones apropiadas. Si no sucede ningún acontecimiento, no se realizará ninguna acción de control.

  Existen muchos sistemas donde las acciones son realizadas en respuesta a algún evento. Por ejemplo: el apagado de una bomba o el cierre de una válvula cuando el nivel de líquido en un tanque alcance un valor determinado, etc. Los sistemas basados en eventos también son usados extensamente para indicar condiciones de alarma e iniciar acciones de emergencia. Las especificaciones de sistemas basados en eventos usualmente incluyen un requerimiento que el sistema debe responder dentro de un máximo tiempo dado para un evento particular.

- **Sistemas disparados por tiempo:** En la aproximación manejada por tiempo las acciones de control se activan en instantes de tiempo predeterminados. Estos instantes de activación se producen normalmente a intervalos de tiempo regulares, de forma que el sistema está formado por acciones activadas periódicamente.

La aproximación que se seleccione para llevar a cabo el control tiene un impacto considerable sobre sus propiedades temporales. La aproximación

disparada por eventos es más eficiente en recursos cuando no ocurren cambios frecuentes en el estado del entorno. Sin embargo, su funcionamiento temporal, en el peor caso, depende del número de eventos que puedan llegar simultáneamente al sistema informático. Por otra parte, la aproximación manejada por tiempo tiene un comportamiento temporal más estable debido al conocimiento a priori de los instantes de la activación de las acciones. Por lo tanto la primera aproximación es más adecuada para la supervisión de acontecimientos esporádicos tales como alarmas y averías, mientras que el último es más adecuado para realizar control continuo utilizando un muestreo periódico.

9.6.1. Mensajes críticos y no críticos

Una de las restricciones típicas que pueden especificarse en sistemas de tiempo real son las restricciones de tiempo. Una típica restricción de tiempo sobre una tarea o mensaje es el plazo (d), el cual representa el tiempo antes del cual un proceso debe completar su ejecución.

Dependiendo de las consecuencias de la pérdida de un plazo, las tareas o mensajes de tiempo real son clasificadas en:

- **Críticos:** Una tarea o mensaje se dice que es crítico si la pérdida de su plazo puede causar consecuencias catastróficas sobre el sistema. Sistemas de tiempo real críticos son sistemas capaces de manejar tareas de tiempo real críticas. Típicamente los sistemas de tiempo real críticos involucran vidas humanas, tal como sistemas de control de vuelo, sistemas de control de procesos químicos, sistemas de monitoreo de pacientes, o ambientes críticos en tiempo tal como sistemas de control de robots.

- **Acríticos:** Una tarea o mensaje se dice que es acrítico si la pérdida de su plazo decrementa el rendimiento del sistema pero no llega a poner en peligro su correcto comportamiento. Sistemas de tiempo real acríticos son sistemas capaces de manejar tareas de tiempo real acríticas. Ejemplo de sistemas de tiempo real acríticos son sistemas de adquisición de datos remotos, los sistemas de reserva de billetes aéreos, máquinas contadoras automáticas, etc.

Además del plazo existen otros parámetros que en general caracterizan una tarea de tiempo real τ_i. Estos parámetros se muestran en la Figura 9.3 y se describen en la Tabla 9.1.

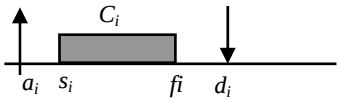


FIGURA 9.3. Parámetros que caracterizan una tarea o mensaje de tiempo real.

TABLA 9.1. Descripción de los parámetros de una tarea de tiempo real

Parámetro		Descripción
Tiempo de llegada	a_i	Tiempo en el cual la tarea está lista para su ejecución, también es referido como el tiempo de liberación o tiempo de solicitud.
Tiempo de cálculo o cómputo	C_i	Tiempo necesario de recurso para ejecutar la tarea sin interrupción.
Plazo	d_i	Tiempo antes del cual una tarea deberá ser completada.
Tiempo de comienzo	s_i	Tiempo en el cual la tarea comienza su ejecución.
Tiempo de finalización	f_i	Tiempo en el cual la tarea finaliza su ejecución.

Otro parámetro importante que caracteriza una tarea de tiempo real es el retraso (J_i) en la activación de una tarea. Entre el instante en que ocurre el evento y el instante en que el sistema se da por enterado de esa ocurrencia puede haber cierto desfase. El retraso en la operación suele depender de las ejecuciones de otras tareas o mensajes concurrentes y por ello suele ser fluctuante. La Figura 9.4 muestra este retraso en la ejecución de una tarea.

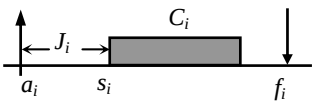


FIGURA 9.4. Retraso o jitter en una tarea de tiempo real.

9.6.2. Mensajería periódica y aperiódica

El intercambio de datos en un sistema distribuido se realiza utilizando mensajes. El intercambio de mensajes entre el emisor y el receptor son generados por la ocurrencia de algún evento. Dependiendo del patrón de llegada de los eventos, la regularidad del intervalo de tiempo entre mensajes o tareas puede variar. En particular las tareas pueden definirse en:

- **Tareas periódicas**: consisten de una secuencia infinita de actividades idénticas, llamadas instancias, que son activadas regularmente a un ritmo constante. La activación de la primera instancia es llamada fase. Si ϕ_i es la fase de la tarea periódica τ_i, el tiempo de activación de la k-esima instancia está dada por $\phi_i+(k\text{-}1)T_i$, donde T_i es el período de la tarea. La Figura 9.5, muestra un ejemplo de esta secuencia de tareas.

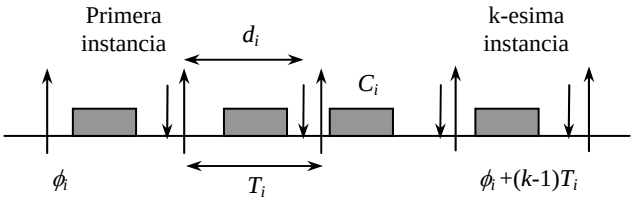


FIGURA 9.5. Secuencia de tareas o mensajes periódicos.

- **Tareas aperiódicas**: consisten de una secuencia infinita de instancias cuyas activaciones no son regulares. La Figura 9.6, muestra un ejemplo de una secuencia de tareas aperiódicas.

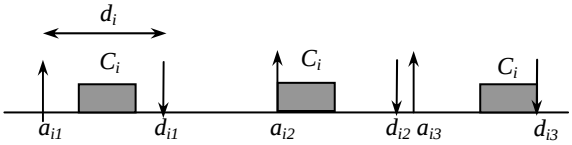


FIGURA 9.6. Secuencia de tareas o mensajes aperiódicos.

9.7. PLANIFICACIÓN DE MENSAJES DE TIEMPO REAL

Un aspecto fundamental a tratar en el desarrollo de sistemas distribuidos de tiempo real es el compartir recursos entre diferentes dispositivos. En los sistemas distribuidos de tiempo real es común que varias tareas necesiten hacer

uso de un mismo recurso de forma concurrente. Si el recurso es multitarea, por ejemplo las memorias multipuerta, no existe ningún problema. Sin embargo, la situación más común es que sólo una tarea pueda hacer uso de un recurso a la vez. Los procesadores y las redes de comunicación son un claro ejemplo.

9.7.1. Políticas de planificación

La planificación del sistema de tiempo real consiste en la definición de las reglas de uso de cada uno de los recursos disponibles. Un sistema de tiempo real se considera planificable si, en función de una política de planificación elegida, es capaz de satisfacer todos los requisitos temporales impuestos. El conjunto de reglas que en algún tiempo determina el orden en el cual las tareas son ejecutadas es llamado un *algoritmo de planificación.*

Los objetivos que persigue toda política de planificación de tiempo real son:

- Ser predecible, con objeto de asegurar tiempos de ejecución finitos. El sistema debe garantizar un mínimo nivel de rendimiento y ser capaz de predecir las consecuencias de alguna decisión de planificación.

- Ser capaz de gestionar el uso de diferentes recursos compartidos. Diferentes recursos pueden requerir diferentes políticas de planificación, incluso formando parte del mismo sistema de tiempo real.

- Debe garantizar el tratamiento de eventos (periódicos y no periódicos).

- Posibilidad de recuperación ante fallos software o hardware.

- Garantizar también la ejecución de tareas sin requisitos temporales, procurando además que los tiempos de respuesta sean reducidos.

- Alcanzar una utilización de recursos altas.

- Que sea sencillo de implementar en aplicaciones reales. Preferiblemente que esté disponible comercialmente.

Como se mencionó anteriormente, el objetivo de un sistema de tiempo real es garantizar la ejecución de las tareas cumpliendo las restricciones temporales; la forma de conseguir esto es teniendo un sistema predecible. Por lo tanto la

predicibilidad constituye una de las principales características que debe tener un sistema de tiempo real.

9.7.2. Test de planificabilidad

El único mecanismo fiable para determinar si la política o políticas de planificación elegidas aseguran el cumplimiento de los plazos consiste en la realización a priori de un test de planificabilidad. Un test de planificabilidad para sistemas de tiempo real consiste en el cálculo de los tiempos de respuesta en el peor caso para cada una de las tareas que conforman el sistema. Si los tiempos de respuesta en el peor de los casos de las tareas son siempre menores que los plazos de ejecución, significa que el sistema siempre cumplirá los requisitos impuestos en cualquier condición.

Dicho test de planificabilidad debe ser al menos suficiente, esto es, si el test dice que el sistema cumple con los requisitos temporales, entonces inexorablemente los cumple bajo cualquier situación. Un test de planificabilidad suficiente debe determinar las cotas superiores de los tiempos de respuesta de las tareas, mediante los cuales se puede garantizar el correcto funcionamiento del sistema. Un test suficiente puede ser pesimista, si se considera como no planificable un sistema que realmente si lo es, como consecuencia de que los tiempos de respuesta en el peor caso se hayan estimado por exceso.

Un planificador se dice que es factible si todas las tareas pueden ser completadas de acuerdo a un conjunto de restricciones específicas. Un conjunto de tareas se dice que es planificable si existe al menos un algoritmo que pueda producir un planificador factible.

9.7.3. Planificación con desalojo

La noción de desalojo es también muy importante en el campo de la planificación de tareas. Una tarea se dice que es desalojada por otras tareas si puede ser interrumpida en algún momento, tal que una tarea más importante que llegue al sistema pueda inmediatamente tomar el recurso. La tarea desalojada es reasumida más adelante tal que se cumpla su objetivo global y sus restricciones temporales. Si las tareas son con desalojo o no depende de la aplicación. Las tareas asociadas a procesos sin interrupción no pueden tolerar el desalojo.

La mayoría de los algoritmos de planificación actuales utilizan el desalojo ya que esta clase de planificación es más eficiente que la sin desalojo. De hecho,

varios estudios demuestran que muchos problemas de planificación que no pueden ser resueltos sin desalojo pueden tener solución si el desalojo es permitido.

9.7.4. Planificación sin desalojo

En la planificación sin desalojo, una vez que una tarea es iniciada, la misma es ejecutada en el recurso (CPU, red de comunicación) hasta su finalización.

Una de las suposiciones básicas en los buses de campo de un sistema de control distribuido, es que el desalojo no es permitido a nivel de transacción. Por lo tanto, cuando una transacción comienza, esta debe ser procesada totalmente antes de pasar a la siguiente transacción. Sin embargo, esto no constituye un problema en términos del planificabilidad ya que la duración de las transacciones por lo general son considerablemente más pequeñas que el período más corto.

9.8. PRIORIDADES ESTÁTICAS Y DINÁMICAS

La prioridad es un mecanismo elemental para planificar la ejecución de un conjunto de tareas. La prioridad es un atributo de las tareas normalmente ligado a su importancia relativa en el conjunto de tareas. Por lo tanto a cada tarea τ_i se le asigna una prioridad *pi* la cual indica la importancia que tiene la tarea τ_i con respecto a las otras tareas en el sistema. Las prioridades pueden ser asignadas a las tareas estáticamente o dinámicamente. Si en un tiempo *t*, $p_a > p_b$ significa que la ejecución de τ_a es más importante que la de τ_b; así, τ_b puede ser retrasada a favor de τ_a.

La planificación estática se basa en un conocimiento a priori completo de los requisitos del sistema y sus restricciones temporales tales como plazos, tiempos de cómputo y de comunicación, restricciones de precedencia y tiempo de liberación. Esta planificación se realiza en tiempo de compilación, esto es, una vez conocido el sistema y antes de su ejecución. La planificación estática produce una planificación fija, la cual se mantiene durante todo el tiempo de funcionamiento del sistema.

Opuesto a la planificación estática se encuentra la planificación dinámica, en la cual los requisitos del sistema y sus restricciones temporales varían con el tiempo y por lo tanto la planificación cambia en el tiempo.

La planificación dinámica es más eficiente que la planificación estática, desde el punto de vista de que consigue mayor utilización de los recursos. Sin embargo, la implementación del planificador estático es mucho más sencilla.

9.8.1. Planificación cíclica

El mecanismo más sencillo de planificación estática lo constituye la planificación cíclica o ejecutivos cíclicos. Esta planificación permite planificar un conjunto de tareas periódicas, independientes y con plazo igual a su período.

Los planificadores cíclicos están definidos por tablas de planificación o planes de ejecución, las cuales se construyen a partir del conjunto de tareas y sus restricciones temporales. Así, en base a los tiempos de ejecución y los períodos del conjunto de tareas, se diseña un plan de ejecución fijo, el cual está formado por un ciclo principal $T_p=mcm(T_i)$ que a su vez está compuesto de ciclos más pequeños T_s, $T_p=k\ T_s$ llamados ciclos secundarios, en el cual se ejecutan actividades correspondientes a las diferentes tareas.

La tabla de planificación contiene un plan principal que define la secuencia de tareas a ejecutar durante un período fijo de tiempo llamado ciclo principal. A su vez, el plan principal se divide en uno o más planes secundarios, los cuales incluyen la secuencia de tareas que deben ejecutarse durante un período fijo de tiempo llamado ciclo secundario.

El funcionamiento del planificador es muy sencillo, ya que sólo tiene que ir leyendo las entradas correspondientes en la tabla de ejecución. Desde este punto de vista es además muy eficiente en tiempo de ejecución, ya que la carga que supone sobre la ejecución de las tareas es mínima.

El principal inconveniente que presenta esta planificación es la poca flexibilidad a la hora de modificar alguno de los parámetros de las tareas, pues ello conlleva el rehacer todo el plan de planificación. Actualmente este tipo de planificación es utilizada en la industria.

A continuación se incluye un ejemplo.

La Tabla 9.2, muestra los parámetros (tiempo de computo C y período T) de tres tareas a ser planificados utilizando un planificador cíclico.

TABLA 9.2. Parámetros de tareas

	τ_1	τ_2	τ_3
C_i (ms)	5	10	40
T_i (ms)	20	40	80

A partir de los parámetros del sistema se tiene:

- Ciclo principal: T_p=*mcm*(T_i) = *mcm*(20, 40, 80) = 80 ms.
- Ciclo secundario: T_p=*k* T_s compuesto por 4 ciclos secundarios de T_s = 20 ms cada uno.

La Figura 9.7, muestra una planificación cíclica para este conjunto de tareas. Donde se puede observar que la tarea τ_3 ha sido divida en 4 subtareas (τ_{31}= 5 ms, τ_{32}= 15 ms, τ_{33}= 5 ms, τ_{34}= 15 ms).

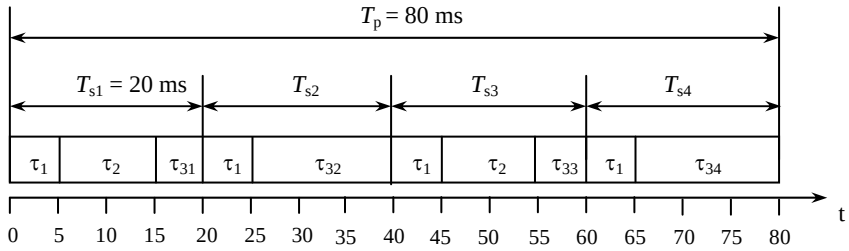


FIGURA 9.6. Planificación cíclica para el conjunto de tareas.

9.8.2. Planificación estática

Como se mencionó anteriormente la prioridad permite establecer la importancia relativa de una tarea respecto a un conjunto de tareas. A partir de la prioridad asignada a las tareas se resuelven los conflictos de utilización del recurso. Cuando hay varias tareas que quieren ejecutar, el planificador elige de entre todas ellas aquella con prioridad más alta y le asigna el recurso.

Si la prioridad de una tarea no cambia durante la ejecución el sistema, se habla de prioridades fijas. Al contrario, si la prioridad puede variar en tiempo de

ejecución, en función del estado de operación del sistema, se habla de prioridades dinámicas.

El principal algoritmo de planificación basado en asignación de prioridad estática lo constituye el algoritmo de planificación Rate Monotonic (RM), el cual asigna una prioridad más alta a las tareas con menor período. El algoritmo Deadline Monotonic (DM) es una mejora al algoritmo RM. Estos algoritmos serán estudiados más en detalle en las secciones 9.8.4 y 9.8.5.

9.8.3. Planificación dinámica

Esta política de planificación al igual que la anterior, trata con la asignación de prioridades a las tareas para la utilización de recursos. Sin embargo, al contrario que en la asignación estática, la prioridad de cada tarea no permanece fija una vez establecida, sino que puede variar en tiempo de ejecución, dependiendo del estado de ejecución de las tareas que requieren el uso del recurso.

El principal algoritmo de este tipo lo constituye el algoritmo EDF (*Earliest Deadline First*), que asigna prioridades en función de los instantes en que se cumplen los plazos de las tareas. Se le asigna la mayor prioridad a la tarea que tiene el plazo más cercano de finalización. Este algoritmo de planificación será estudiado en la sección 9.8.6.

9.8.4. Algoritmo de planificación *Rate Monotonic*

El algoritmo de planificación *Rate Monotonic* asigna prioridades a las tareas de acuerdo a su tasa de solicitud, es decir tareas con períodos cortos tendrán una alta prioridad (esto es, como una función monótona de la tasa de solicitud). Puesto que los períodos son constantes RM es un asignador de prioridades fijo. El algoritmo RM permite el desalojo, es decir, una tarea que llega con un período más corto puede desalojar y adelantar una tarea que esté ejecutándose.

Test de planificabilidad

Un conjunto de n tareas periódicas caracterizadas por (C, T, D), que se ejecutan sobre un único recurso será planificable bajo el algoritmo de planificación Rate Monotonic si cumple que:

$$U = \sum_{i=1}^{n} \frac{C_i}{T_i} \leq U(n) = n(2^{1/n} - 1) \tag{9.1}$$

donde, U es el factor de utilización total del recurso y U(*n*) es el límite de utilización máximo del recurso para *n* tareas.

Este test condiciona la planificabilidad del sistema a que la utilización total del procesador no sobrepase un valor máximo establecido, dependiente del número total de tareas. El límite de utilización varia entre el 83% para *n*=1 y el 69% para valores elevados de *n*. Este test es suficiente pero no necesario, de forma que puede haber conjuntos de tareas que sobrepasen el límite de utilización establecido por el test y que aún pueden ser planificables.

Ejemplo

La Tabla 9.3, muestra los parámetros (tiempo de computo *C* y período *T*) de dos tareas a ser planificados utilizando un planificador *Rate Monotonic*.

TABLA 9.3. Parámetros de tareas

	τ_1	τ_2
C_i (ms)	2	3
T_i (ms)	5	6

A partir de los parámetros del sistema se tiene que:

- Factor de utilización *U* es:

$$U = \sum_{i=1}^{2} \frac{C_i}{T_i} = \frac{2}{5} + \frac{3}{6} = \frac{27}{30} = 0.9$$

  Esto significa que este conjunto de tareas requiere el 90% del tiempo de recurso para ejecutarse.

- Límite de utilización máximo del recurso *U*(*n=2*) es:

$$U(n = 2) = 2(2^{1/2} - 1) = 0.83$$

Puesto que $U > U(n)$, la planificabilidad del conjunto de tareas no puede ser garantizada bajo el algoritmo de planificación Rate Monotonic. Sin embargo, la Figura 9.7, muestra que este conjunto de tareas puede ser planificado bajo RM. Ya que la tarea τ_1 tiene el período menor, tiene asignada la mayor prioridad.

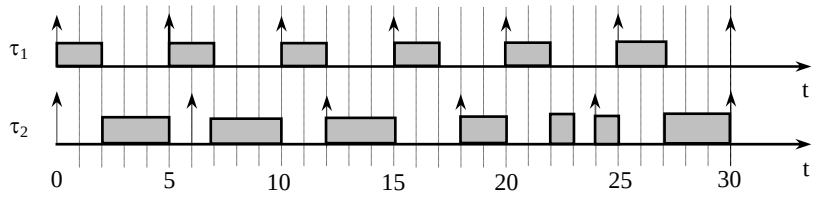


FIGURA 9.7. Planificación producida por el algoritmo *Rate Monotonic*.

9.8.5. Algoritmo de planificación *Deadline Monotonic*

El algoritmo de planificación *Deadline Monotonic* (DM) es una extensión del algoritmo de planificación Rate Monotonic en donde las tareas pueden tener un plazo relativo menor que su período.

De acuerdo al algoritmo DM, a cada tarea se le asigna una prioridad inversamente proporcional a su plazo relativo. Así, en algún instante, la tarea con el plazo relativo más corto se ejecuta. Puesto que el plazo relativo es constante, DM es un asignador de prioridad estático. Como RM, DM usa desalojo en su planificación.

Este método se basa en el cálculo de los tiempos de respuesta (R) en el peor caso para cada tarea. El peor tiempo de respuesta, para tareas con alta prioridad, es igual a su tiempo de cómputo (esto es, R=C). Las tareas con prioridad baja y media sufrirán interferencia producidas por las tareas con alta prioridad. Para una tarea con una prioridad no alta, el tiempo de respuesta en el peor caso es:

$$R_i = C_i + I_i \tag{9.2}$$

donde I_i es la máxima interferencia que la tarea τ_i puede experimentar en un intervalo de tiempo $[t, t + R_i)$. La máxima interferencia ocurre cuando todas las tareas con alta prioridad se liberan al mismo tiempo que la tarea τ_i (esto es, en un instante crítico). El método se basa en que las tareas se ordenan por plazos relativos creciente, tal que $i < j \Leftrightarrow d_i < d_j$ la tarea τ_j tendrá mayor prioridad que la tarea τ_i.

La máxima interferencia que la tarea τ_i puede experimentar en un intervalo $[0, R_i)$ por las tareas de alta prioridad está dada por:

$$I_i = \sum_{j \in ap(i)} \left\lceil \frac{R_i}{T_j} \right\rceil C_j \tag{9.3}$$

donde *ap*(*i*) es el conjunto de tareas con mayor prioridad a la tarea τ_i.

Por lo tanto, el tiempo de repuesta en el peor de los casos para la tarea τ_i será:

$$R_i = C_i + \sum_{j \in ap(i)} \left\lceil \frac{R_i}{T_j} \right\rceil C_j \tag{9.4}$$

Como se puede observar, no existe una solución simple para esta ecuación puesto que R_i aparece a ambos lados de la misma. Sin embargo se ha demostrado que esta ecuación puede resolverse de forma iterativa mediante la siguiente expresión:

$$\begin{aligned} R_i^{n+1} &= C_i + I_i^n, \ \forall i \quad I_i^0 = 0 \\ R_i^{n+1} &= C_i + \sum_{j \in ap(i)} \left\lceil \frac{R_i^n}{T_j} \right\rceil C_j \end{aligned} \tag{9.5}$$

La dependencia monótona de la ecuación respecto del término R_i garantiza la convergencia del algoritmo, siempre y cuando la utilización sea menor del 100%.

La iteración comienza asignando un valor inicial al tiempo de respuesta $R_i^0 = C_i$ y finaliza cuando $R_i^n = R_i^{n+1}$. Si la ecuación no tiene solución entonces los valores de *R* continuarán creciendo y $R_i^{n+1} > T_i$.

Ejemplo

La Tabla 9.4, muestra los parámetros (tiempo de cómputo *C* y período *T*) de tres tareas a ser planificados utilizando un planificador *Deadline Monotonic*. Todas las tareas se liberan simultáneamente en $t = 0$. Los plazos de las tareas son igual a sus períodos.

TABLA 9.4. Parámetros de tareas

	τ_1	τ_2	τ_3
C_i (ms)	3	3	5
T_i (ms)	7	12	20

El primer paso para la planificación de estas tareas por el algoritmo DM es aplicar el test basado en análisis de tiempo para predecir el tiempo de respuesta en el peor caso de cada tarea y luego comparar estos valores con sus plazos para verificar la planificabilidad del conjunto de tarea.

La tarea con más alta prioridad es la tarea τ_1, ya que esta tarea tiene el menor plazo relativo (7 ms). Por lo tanto, ésta tendrá un tiempo de respuesta igual a su tiempo de computación, $R_1 = C_1 = 3$. Puesto que esta tarea tiene la mayor prioridad, ésta interferirá en la ejecución de las otras tareas, por lo tanto haciendo uso de la ecuación recurrente para R, se puede determinar los tiempos de respuesta para las tareas τ_2 y τ_3:

Tarea τ_2

$$R_2^1 = C_2 + I_2^0 = 3$$

$$R_2^2 = C_2 + I_2^1 = 3 + \left\lceil \frac{R_2^1}{T_1} \right\rceil C_1 = 3 + \left\lceil \frac{3}{7} \right\rceil 3 = 6$$

$$R_2^3 = C_2 + I_2^2 = 3 + \left\lceil \frac{R_2^2}{T_1} \right\rceil C_1 = 3 + \left\lceil \frac{6}{7} \right\rceil 3 = 6$$

Puesto que $R_2^2 = R_2^3 = 6$, el tiempo de respuesta de la tarea τ_2 se ha podido encontrar.

Tarea τ_3

$$R_3^1 = C_3 + I_3^0 = 5$$

$$R_3^2 = C_3 + I_3^1 = 5 + \left\lceil \frac{R_3^1}{T_1} \right\rceil C_1 + \left\lceil \frac{R_3^1}{T_2} \right\rceil C_2 = 5 + \left\lceil \frac{5}{7} \right\rceil 3 + \left\lceil \frac{5}{12} \right\rceil 3 = 11$$

$$R_3^3 = C_3 + I_3^2 = 5 + \left\lceil \frac{R_3^2}{T_1} \right\rceil C_1 + \left\lceil \frac{R_3^2}{T_2} \right\rceil C_2 = 5 + \left\lceil \frac{11}{7} \right\rceil 3 + \left\lceil \frac{11}{12} \right\rceil 3 = 14$$

$$R_3^4 = C_3 + I_3^3 = 5 + \left\lceil \frac{R_3^3}{T_1} \right\rceil C_1 + \left\lceil \frac{R_3^3}{T_2} \right\rceil C_2 = 5 + \left\lceil \frac{14}{7} \right\rceil 3 + \left\lceil \frac{14}{12} \right\rceil 3 = 17$$

$$R_3^5 = C_3 + I_3^4 = 5 + \left\lceil \frac{R_3^4}{T_1} \right\rceil C_1 + \left\lceil \frac{R_3^4}{T_2} \right\rceil C_2 = 5 + \left\lceil \frac{17}{7} \right\rceil 3 + \left\lceil \frac{17}{12} \right\rceil 3 = 20$$

$$R_3^6 = C_3 + I_3^5 = 5 + \left\lceil \frac{R_3^5}{T_1} \right\rceil C_1 + \left\lceil \frac{R_3^5}{T_2} \right\rceil C_2 = 5 + \left\lceil \frac{20}{7} \right\rceil 3 + \left\lceil \frac{20}{12} \right\rceil 3 = 20$$

Puesto que $R_2^5 = R_2^6 = 20$, el tiempo de respuesta de la tarea τ_3 se ha podido encontrar.

La Tabla 9.5, muestra el conjunto de tareas y el tiempo de respuesta para cada tarea, a partir de la cual se puede observar que cada tarea tiene un tiempo de respuesta menor o igual a su plazo, por lo tanto este conjunto de tareas es planificable por DM.

TABLA 9.5. Tiempo de respuesta de las tareas

	τ_1	τ_2	τ_3
C_i (ms)	3	3	5
T_i (ms)	7	12	20
R_i (ms)	3	6	20

La Figura 9.8, muestra la planificación de este conjunto de tareas utilizando el algoritmo de planificación DM. Analizando los plazos relativos de cada una se tiene que: la tarea τ_1 tiene el plazo relativo más corto por lo tanto tiene asignada la más alta prioridad, la tarea τ_2 se le asigna una prioridad intermedia y la tarea τ_3 se le asigna la prioridad más baja.

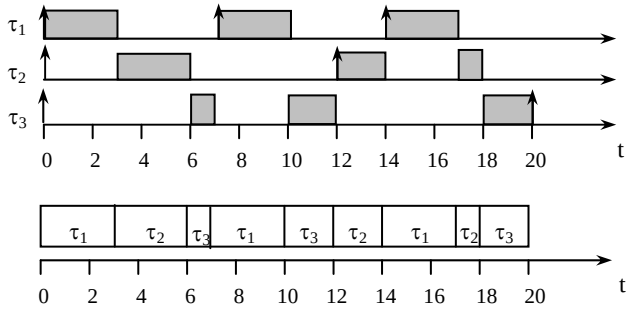


FIGURA 9.8. Planificación producida por el algoritmo *Deadline Monotonic*.

9.8.6. Algoritmo de planificación *Earliest Deadline First*

Earliest Deadline First (EDF) es un planificador dinámico que asigna mayor prioridad a las tareas con plazo de finalización más próximo. EDF utiliza desalojo y puede ser usado para planificar tanto tareas periódicas como aperiódicas.

Para este caso la menor cota para el factor de utilización del procesador es 1, por lo tanto las tareas pueden utilizar el 100% del procesador y aún ser planificable.

Test de planificabilidad

Un conjunto de *n* tareas periódicas caracterizadas por (*C*, *T*, *D*), que se ejecutan sobre un único recurso será planificable bajo el algoritmo de planificación EDF si cumple que:

$$U = \sum_{i=1}^{n} \frac{C_i}{T_i} \leq 1 \tag{9.6}$$

En este caso la menor cota para el factor de utilización del procesador es 1, por lo tanto las tareas pueden utilizar el 100% del recurso y aún ser planificable.

Ejemplo

La Tabla 9.6, muestra los parámetros (tiempo de cómputo *C* y período *T*) de tres tareas a ser planificados utilizando un planificador Earliest Deadline First. Todas las tareas se liberan simultáneamente en $t = 0$. Los plazos de las tareas son igual a sus períodos.

TABLA 9.6. Parámetros de tareas

	τ_1	τ_2	τ_3
C_i (ms)	2	2	4
T_i (ms)	6	8	12

A partir de los parámetros del sistema se tiene que:

- Factor de utilización *U* es:

$$U = \sum_{i=1}^{3} \frac{C_i}{T_i} = \frac{2}{6} + \frac{2}{8} + \frac{4}{12} = \frac{11}{12} = 0.916$$

Esto significa que este conjunto de tareas requiere el 91,6% del tiempo de recurso para ejecutarse.

Puesto que U<1 el conjunto de tareas puede ser planificado bajo el algoritmo de planificación EDF. La Figura 9.9, muestra la planificación EDF para este conjunto de tareas. En la Tabla 9.7 se detalla el funcionamiento del algoritmo EDF para cada instante de tiempo *t*.

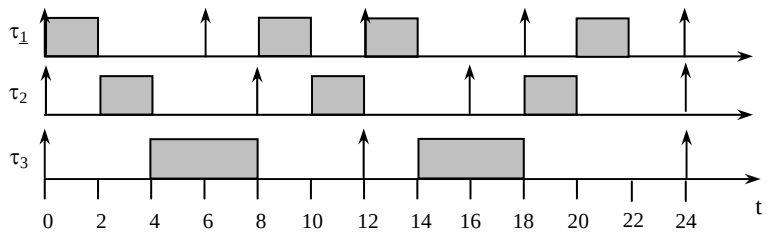


FIGURA 9.9. Planificación producida por el algoritmo *Earliest Deadline First*.

9.9. ANÁLISIS DE TIEMPO REAL EN REDES DE COMUNICACIÓN

El análisis de sistemas distribuidos de control es más complejo respecto al caso de un procesador estudiado previamente.

En el caso de los sistemas distribuidos de control es necesario considerar el análisis del sistema de comunicación entre los diferentes procesadores involucrados en el proceso. Si se quiere enviar un mensaje de un determinado tamaño a través de una red, se debe conocer el tiempo total que empleará el mensaje en llegar a su destino. Como se indicó anteriormente se supondrá que los mensajes se envían en diferentes paquetes de tamaño fijo y que, en la implementación del sistema de tiempo real, se emplean redes de comunicación que planifican la transmisión de los paquetes con prioridades asignadas.

TABLA 9.7. Secuencia de eventos - planificación EDF

Tiempo (*ms*)	*Eventos*
$t = 0$	Todas las tareas están listas para ser ejecutadas, puesto que τ_1 tiene el plazo de finalización más corto (6 ms) la misma comienza su ejecución.
$t = 2$	Las tareas τ_2 y τ_3 están listas para ser ejecutadas, en este caso la tarea τ_2 es ejecutada ya que tiene el plazo de finalización más corto (8 ms)
$t = 4$	Sólo la tarea τ_3 está lista para ser ejecutada, por lo tanto la misma comienza a ejecutarse en este instante de tiempo.
$t = 8$	Las tareas τ_1 y τ_2 están listas para ser ejecutadas, puesto que la tarea τ_2 tiene el plazo de finalización más corto (12 ms), la misma comienza su ejecución
$t = 10$	Sólo la tarea τ_2 está lista para ser ejecutada, por lo tanto la misma comienza a ejecutarse en este instante de tiempo.
$t = 12$	Las tareas τ_1 y τ_3 están listas para ser ejecutadas, puesto que la tarea τ_1 tiene el plazo de finalización más corto (18 ms), la misma comienza su ejecución
$t = 14$	Solo la tarea τ_3 está lista para ser ejecutada, por lo tanto la misma comienza a ejecutarse en este instante de tiempo.
$t = 18$	Las tareas τ_1 y τ_2 están listas para ser ejecutadas, y ambas tareas tiene igual plazo de finalización (24 ms). En este caso se selecciona la tarea τ_2 para su ejecución ya que la misma tenía más tiempo en espera para su ejecución.

Dado que la red no permite desalojo de mensajes, se debe considerar el bloqueo del dispositivo de comunicación debido a la transmisión de mensajes. La duración de este bloqueo será igual al tiempo de transmisión de un paquete (B_m). Para calcular el tiempo total que tarda el mensaje en llegar completamente a su destino se debe considerar la interferencia que producen los mensajes de mayor prioridad que circulen por la misma red. Si los plazos de ejecución son menores o iguales que los correspondientes períodos, el tiempo de respuesta se obtiene de la expresión:

$$R_m = C_m + B_m + \sum_{j \in ap(i)} \left\lceil \frac{R_m}{T_j} \right\rceil C_j \tag{9.8}$$

donde:

R_m tiempo de respuesta peor caso de cada mensaje.

C_m es el tiempo de transmisión de peor caso del mensaje, suponiendo que fuera el único a transmitir por ese canal de comunicación.

$ap(i)$ es el conjunto de mensajes con mayor prioridad al mensaje τ_i.

Se puede observar que esta expresión sólo difiere en el término B_m de la expresión del tiempo de respuesta para el caso de DM.

9.10. CONCLUSIONES

Este capítulo ha presentado una introducción a los conceptos y definiciones que concierne a los sistemas distribuidos de tiempo real. Comenzando desde una perspectiva general de las comunicaciones y su aplicación a la industria, para después profundizar en el problema de tiempo real y enfocarse finalmente en los problemas de comunicación de tiempo real.

En los sistemas de tiempo real, los recursos más importantes y sobre los que se centra la mayor parte del esfuerzo de investigación son el procesador y la red de comunicación. Dado que ambos recursos se pueden considerar con un comportamiento similar, este capítulo se ha centrado en el recurso procesador, esto es: bajo qué condiciones el procesador se asigna a un conjunto de tareas (o una red de comunicación a un conjunto de mensajes).

Los conceptos y definiciones de tiempo real presentados en este capítulo pueden ser aplicados a ciertos sistemas de comunicación de tiempo real, como es el caso de los buses de campo.

En el análisis de sistemas de comunicación de tiempo real a diferencia de los sistemas de tiempo real sobre un procesador, es necesario considerar el análisis del sistema de comunicación entre los diferentes procesadores involucrados en el proceso. Por otro lado, los servicios y protocolos utilizados en el medio o red de comunicación deben ser tales que permitan un intercambio correcto y a tiempo de los mensajes, de tal manera que el tiempo total que emplea un mensaje en llegar a su destino puede ser determinado con exactitud.

9.11. BIBLIOGRAFIA

J. Blazewicz y otros. Scheduling in computer and manufacturing systems. Ed. Springer Verlag, 1993.

G.C. Buttazzo. Hard real-time computer systems. Predictable scheduling algorithms and applications. Ed. Kluwer Academic Publishers, 1997.

J.A. Stankovic. Real-time computing systems: The next generation. IEEE Tutorial on Hard Real-Time Systems. IEEE Computer Society Press, New York, 1988.

9.12. EVALUACIÓN

9.12.1. Evaluación objetiva

1. Indique en que niveles de la pirámide jerárquica de comunicaciones se emplean los buses de campo:

   A) Nivel de control de proceso y Nivel de control de producción

   B) Nivel de gestión y Nivel de campo

   C) Nivel de campo y Nivel de entrada/salida

   D) Nivel de entrada/salida y Nivel de gestión

2. Los sistemas de tiempo real se caracterizan principalmente por:

   A) Producir una respuesta correcta y rápida en el tiempo

   B) Producir una respuesta correcta

   C) Producir una respuesta correcta en un intervalo de tiempo específico

   D) Producir una respuesta rápida

3. Una tarea o mensaje cuya pérdida de plazo produce graves consecuencias en el sistema, se denomina:

   A) Tarea o mensaje periódico

   B) Tarea o mensaje acrítico

   C) Tarea o mensaje aperiódico

   D) Tarea o mensaje crítico

4. Dentro de las características que debe poseer un sistema de tiempo real, ¿Cuál de las siguientes debería ser obligatoria?:

   A) Alcanzar altas utilizaciones de recurso (procesador o red de comunicación)

   B) Fácil implementación y mantenimiento

   C) Ser predecible

   D) Poder tratar tanto eventos periódicos como aperiódicos

5. Un sistema de tiempo real donde la prioridad de las tareas cambian durante la ejecución del mismo, se dice que tiene una política de planificación:

   A) Estática

   B) Con desalojo

   C) Dinámica

   D) Sin desalojo

6. El conjunto de reglas que permiten determinar el orden en el cual se deben ejecutar las tareas se denomina:

   A) Test de planificabilidad

   B) Planificación factible

   C) Planificación no factible

   D) Algoritmo de planificación

7. El algoritmo de planificación que asigna prioridades a las tareas de acuerdo a su tasa de solicitud o período se denomina

   A) Algoritmo de planificación Rate Monotonic

   B) Algoritmo de planificación Deadline Monotonic

   C) Algoritmo de planificación Earliest Deadline First

   D) Algoritmo de planificación dinámico

8. Una característica que distingue un sistema distribuido de tiempo real es

   A) Permite desalojo

   B) No permite el desalojo

   C) Permite únicamente tareas aperiódicas

   D) Permite únicamente tareas periódicas

9. Dado el siguiente conjunto de tareas definidas por $\tau_i = (C_i, T_i)$ con $d_i = C_i$, $\tau_1 = (2, 6)$, $\tau_2 = (1, 4)$, $\tau_3 = (3, 10)$. El factor de utilización U de recurso que tendrá este conjunto de tareas es:

   A) 70,0 %

   B) 80,5 %

   C) 88,3 %

   D) 85,0 %

10. Para el conjunto de tareas especificadas en la pregunta anterior, si se utiliza el algoritmo de planificación Rate Monotonic, el orden de ejecución de las tareas en orden de prioridad ascendente es:

    A) τ_1, τ_2, τ_3

    B) τ_2, τ_1, τ_3

    C) τ_3, τ_2, τ_1

    D) τ_2, τ_3, τ_1

TEMA 10

INTRODUCCIÓN A LOS BUSES DE CAMPO Y A LOS SISTEMAS DISTRIBUIDOS

10.1. Introducción y orientaciones para el estudio
10.2. Objetivos
10.3. Los buses de campo. Definición
10.4. Los sistemas distribuidos
10.5. Estructuración de los sistemas distribuidos industriales
10.6. Características de distribución
10.7. Características de la mensajería de los buses de campo
10.8. Normalizaciones en los buses de campo
10.9. Modelo general de operación de los buses de campo
10.10. Funcionalidades comunes
10.11. Conclusiones
10.12. Bibliografía
10.13. Evaluación

10.1. INTRODUCCIÓN Y ORIENTACIÓN PARA EL ESTUDIO

Los buses de campo han provocado una revolución en cómo se controlan los sistemas industriales, pues a su sencillez estructural se añade la facilidad de utilización y el coste reducido. Los sistemas industriales están formados por máquinas de producción, células, equipos, actuadores y sensores de gran variedad de características. Su control avanzado exige que exista una conectividad entre todos estos elementos siguiendo alguna de las arquitecturas habituales, como las arquitecturas centralizadas o distribuidas.

La sencillez estructural se refiere a que la implantación de un bus de campo permite simplificar y racionalizar la estructura de soporte a la comunicación entre los distintos equipos de una instalación industrial. Se pasa de una instalación formada por un cableado complejo, voluminoso, con cables dedicados, poco mantenible, con gran número de puntos de conexión, a una instalación simple, limpia, fiable y de gran rendimiento. Téngase en cuenta que en una instalación industrial con conexionado punto a punto, cada enlace extremo a extremo tiene de media 13 conexiones y cada una de ellas es un posible punto de avería. La reducción del número de conexiones beneficia la fiabilidad del conjunto.

La facilidad de utilización se refiere a que la instalación se puede utilizar con procedimientos sistemáticos, se puede configurar el sistema desde un configurador central y se puede reconfigurar fácilmente cuando se requiere hacer cambios en la instalación. Además ésta se puede testear de manera sencilla, sin necesidad de recorrer toda la instalación para detectar posibles fallos o averías.

El coste reducido es una consecuencia directa de la simplificación de la instalación, pero también de la facilidad de su supervisión y mantenimiento.

En este capítulo se muestran las principales características de los sistemas distribuidos industriales. Por un lado se muestran las funcionalidades de dichos sistemas desde el punto de vista estructural.

Por otro lado se indican las características de los buses de campo en su aplicación al control distribuido.

10.2. OBJETIVOS

El objetivo principal de este capítulo es proporcionar los conocimientos básicos de los buses de campo en su aplicación en el control distribuido de sistemas industriales. Para alcanzar este objetivo se definen como objetivos el estudio de los sistemas distribuidos industriales, de las características que se requieren para el control de dichos sistemas, y en las funcionalidades de los buses de campo, que forman parte de los sistemas distribuidos.

10.3. LOS BUSES DE CAMPO. DEFINICIÓN

El término genérico de “bus de campo” se refiere a un sistema de comunicación, que engloba a un conjunto de redes de comunicación especialmente adaptado a la interconexión de sistemas y equipos industriales. Su implantación ha permitido sustituir las conexiones punto a punto analógicas tradicionales (por bucle de corriente, por nivel de tensión, de presión, etc.) por redes de transmisión digital bidireccionales, multipunto, sobre estructura de red serializada y formada por un número escaso de conductores (2 o 3 hilos).

Los nodos que forman parte del sistema distribuido deben incorporar una capacidad de conversión a formato digital de los datos a transmitir y deben disponer de un subsistema de comunicación que permita coordinarse con los demás nodos de la red. Este subsistema de comunicación suele ser muy sencillo aunque puede estar a su vez ligado a un procesador local que le permita ejecutar funciones de distintos niveles de complejidad.

Un bus de campo puede actuar como mero soporte a la comunicación de las lecturas de los sensores y como transporte de las órdenes a los actuadores desde un controlador remoto. Pero además, la posibilidad de disponer de inteligencia local permite a los nodos ejecutar distintos niveles de tareas, desde el control local coordinado con otros controladores a tareas avanzadas de control, supervisión, y diagnóstico del sistema.

10.4. LOS SISTEMAS DISTRIBUIDOS

Los sistemas distribuidos (SD) están formados por dispositivos autónomos inteligentes que cooperan con objetivos concretos. Estos dispositivos ejecutan tareas que se coordinan entre si intercambiando información por medio de una

red de comunicación. Cuando en el sistema distribuido existen nodos que ejecutan tareas de control entonces se tiene un sistema distribuido de control. En este caso, la red de comunicación se denomina bus de campo.

Por su característica distribuida, un SD puede disponer de diversas tecnologías tanto en el *hardware*, *software*, enlaces, componentes, etc. Sus ámbitos de aplicación son muy variados y en cada uno de ellos existen variaciones de acuerdo con los objetivos concretos del SD. Por ello, el estudio de los sistemas distribuidos contempla un gran número de posibles escenarios. Los sistemas empotrados suelen formar parte de sistemas distribuidos, así como los sistemas de control y supervisión (SCADA), los sistemas Batch, y muchas aplicaciones específicas en los sectores de automoción, fabricación, transporte aéreo, ferroviario, marino, etc.

10.5. ESTRUCTURACIÓN DE LOS SISTEMAS DISTRIBUIDOS INDUSTRIALES

A los sistemas distribuidos industriales se les suele asociar el modelo de referencia de automatización de la fabricación (ISO-RM-FA, *Internacional Standards Organization Reference Model for Factory Automation*) (Tabla 10.1). Este modelo tiene una estructura piramidal con funcionalidad de capas, correspondientes a los distintos niveles jerárquicos y abstractos de las organizaciones industriales. Es conveniente señalar que este modelo no es único sino que distintos autores utilizan variaciones del mismo según los aspectos que quieran resaltar (por ejemplo, el del capítulo anterior).

La estructura jerarquizada y la división de responsabilidades permiten modelar el sistema distribuido de distintas formas en función de la propia organización industrial. Así, en niveles altos de la estructura jerárquica los procesos y tareas tienen funcionalidades globales de empresa, por ejemplo los ERP como SAP-R3 (Planificadores de Recursos de Empresa). A niveles intermedios se realizan tareas de asignación y coordinación de células, máquinas y almacenes. En los niveles inferiores se realiza el control directo de equipos y máquinas, la supervisión de componentes, la entrada y salida de variables directas de proceso.

Uno de los factores que permiten asegurar la productividad a largo plazo se basa en la previsión de un alto grado de flexibilidad que permita afrontar de manera rápida y económica los cambios que se produzcan tanto en la planta

(incorporación o retirada de líneas de fabricación, de máquinas, etc.) como de cambios en el entorno o en el producto. Se podrá alcanzar este alto grado de flexibilidad si tanto los sistemas como los equipos están sujetos a unas normalizaciones que les permita ser suficientemente versátiles y modulares como para que cualquier cambio local no exija cambios estructurales y, como consecuencia, se pueda implantar de manera rápida y sencilla.

TABLA 10.1. Modelo de referencia de fabricación ISO-RM-FA

Nivel	Jerarquía	Control	Responsabilidad	Funciones Básicas
6	EMPRESA	GESTIÓN CORPORATIVA	marcar objetivos gestión global	gestión financiera, marketing y ventas, investigación y desarrollo
5	PLANTA	PLANIFICACIÓN DE LA PRODUCCIÓN	funciones de planificación y organización de producción	diseño de producto, gestión e ingeniería de producción; abastecimiento, gestión de recursos gestión de mantenimiento
4	SECCIÓN/ ÀREA	ASSIGNACIÓN Y SUPERVISSIÓN DE MATERIALES Y RECURSOS	coordinación de la producción, soporte de tareas y asignación de los recursos a las tareas	gestión de producción, gestión de almacén, abastecimiento, gestión de mantenimiento, transportes, tratamiento de residuos
3	CÉLULA	COORDINAR MÚLTIPLES MÀQUINAS Y OPERACIONES	secuenciamiento y supervisión de tareas de célula y servicios auxiliares	producción a nivel de célula coordinación de máquinas en la célula
2	ESTACIÓN	COMANDOS DE SECUENCIA Y MOVIMIENTO DE MÀQUINAS	dirección y coordinación de la actividad de los equipos	producción a nivel de estación coordinación de subsistemas, planificación de tareas
1	EQUIPO	ACTIVAR SECUENCIAS Y MOVIMIENTOS	realización de los comandos en los equipos	producción a nivel de equipo, lecturas de sensores, órdenes a accionadores

10.6. CARACTERÍSTICAS DE DISTRIBUCIÓN

La automatización de un sistema industrial requiere disponer de unas metodologías que permitan alcanzar el objetivo principal de la automatización: que se cumplan unos objetivos de eficacia, de calidad, de tiempo y de respeto al medio, con un coste mínimo. Para conseguirlo, hace falta proceder siguiendo unos esquemas que tienen aspectos teóricos y prácticos. Además, también se requiere conocer los sistemas industriales, sus opciones y posibilidades.

Un Sistema Distribuido de Control es la distribución de los elementos y funciones del sistema de control por toda la planta, con tres aspectos de valoración: distribución física, distribución funcional i distribución de riesgo.

- La distribución física indica la localización de los elementos y funciones de control en la planta. El objetivo es de situarlos cerca de los elementos bajo control, formando parte del control integral de planta por medio de una infraestructura de intercomunicación.

- La distribución funcional se refiriere a como se sitúan las funciones de control, de supervisión, interfaces, etc. en el entorno físicamente distribuido.

- La distribución de riesgo muestra como se delimitan los riesgos al poder aislar estos tanto en la distribución física como en la distribución funcional.

La Figura 10.1 muestra una división jerarquizada basada en los flujos de información, en modo funcional. En ella debe observarse que las necesidades de respuesta temporal suelen ser fuertemente variables en las distintas capas. (MES, *Manufacturing Execution System*; ERP, *Enterprise Resource Planning*; PLC, *Programmable Logic Controller*).

En los sistemas industriales los computadores son elementos claves en la implementación de los sistemas de automatización, constituyendo una herramienta imprescindible para el control de los dispositivos y procesos de fabricación. La productividad es la razón última de toda organización industrial. El grado de productividad de una organización industrial depende, entre otros factores, del grado de automatización que en ella se ha incorporado. La automatización se basa en el uso y la manipulación de la información generada por controlar el proceso productivo.

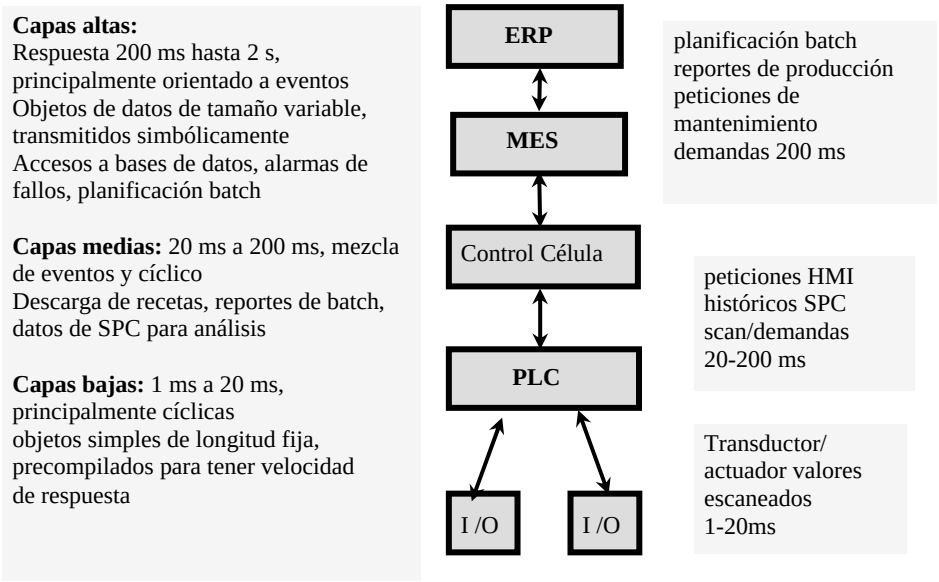


FIGURA 10.1. Flujo de Información en un sistema distribuido.

Todo proyecto de automatización global está formato por un conjunto de dispositivos, máquinas y controladores que están distribuidos en toda la corporación, y por lo tanto la correcto organización y explotación del sistema de comunicación entre todos los elementos de la corporación permitirá la optimización de sus prestaciones y recursos, la reducción de costes de fabricación y por lo tanto, una más alta productividad.

La eficacia y eficiencia de este sistema distribuido dependerá en gran parte de la infraestructura de comunicaciones, o sea, de qué redes de comunicaciones (industriales) proporcionan los servicios deseados. No todas les redes son iguales y por ello se deberá escoger la mejor solución para que todos los requisitos de comunicación se satisfagan debidamente.

En la Figura 10.2 se ilustra la posible infraestructura de comunicaciones que requeriría una organización industrial siguiendo el modelo de referencia de automatización para la fabricación ISO RM FA. En ella se pueden distinguir las diferentes redes de comunicaciones industriales que formarían el esqueleto de la pirámide de la producción, así como también la interconexión de estas redes mediante pasarelas.

Por extensión se suelen denominar indistintamente Redes Industriales o Buses de Campo a toda una familia de subredes, como puede observarse en la figura anterior. En particular se suelen distinguir las redes de bit, orientadas a la transmisión de bits individuales o pequeñas agrupaciones de bits (Interbus, AsI); redes de dispositivos que pueden transmitir desde bits a secuencias de bytes (Interbus, CAN, DeviceNet, Profibus-DP, etc.); redes de fábrica, orientadas a transmitir información general y enlace con los sistemas de gestión (Profibus-FMS, Ethernet-IP, ..). Estas divisiones no son exclusivas sino que a menudo dependen de las denominaciones que toman los distintos grupos de usuarios o de fabricantes y también de los entornos en los cuales estas redes se aplican.

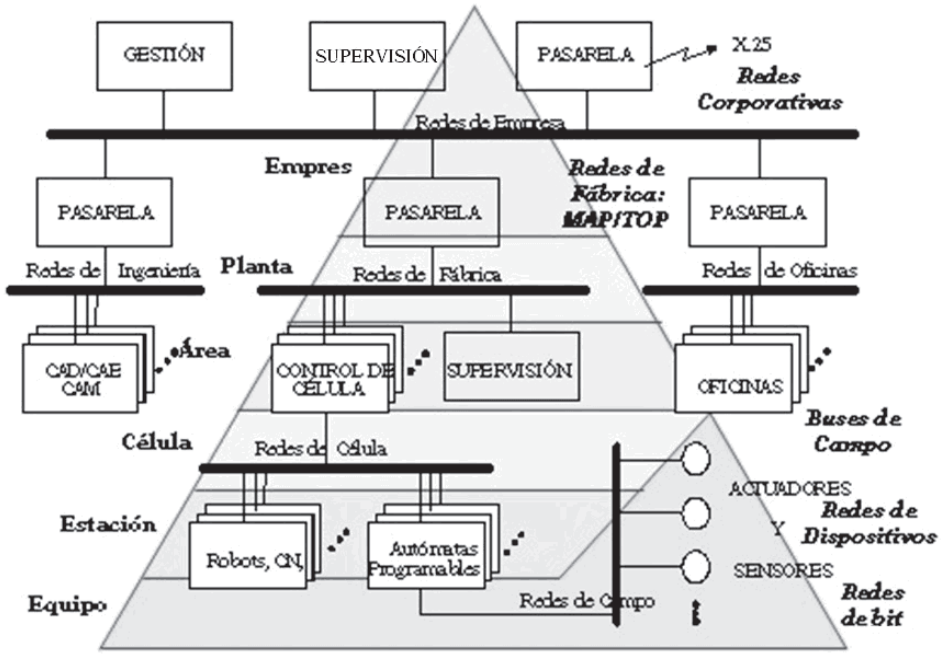


FIGURA 10.2. Flujo de Información en un sistema distribuido.

10.7. CARACTERÍSTICAS DE LA MENSAJERÍA DE LOS BUSES DE CAMPO

Los buses de campo, como integrantes de la estructura de comunicaciones de una empresa industrial, deben cumplir requerimientos temporales, de periodicidad de transmisión y de volumen de datos.

- Intercambio de información fuera de línea (*off-line*): en estas operaciones no hay límite temporal en las transacciones, puesto que no afectan al funcionamiento del sistema.

- Intercambio de información en línea (*on-line*): en estas operaciones se pide que haya interactividad, con posibles retardos en la respuesta, normalmente pequeños (por ejemplo, consulta de estados, carga de programas, etc.).

- Intercambio de información con requisitos de tiempo real: en estas operaciones se pide que haya límite estricto en las transacciones. O sea, se requiere que el tiempo transcurrido entre la generación de un mensaje y su respuesta esté estrictamente dentro de un intervalo de tiempo especificado (por ejemplo, lectura de un sensor). Estas operaciones son las propias de los sistemas de tiempo real, dónde la corrección del sistema no sólo depende de los resultados lógicos obtenidos, sino también del instante en el cual estos se han obtenido.

- Intercambio esporádico de información: en este tipo de comunicación, las transacciones que se hacen no siguen ningún patrón de comportamiento, sino que se realizan esporádicamente, de forma automática (por ejemplo, activación de una alarma) o manual (por ejemplo, carga de un programa en un PLC).

- Intercambio periódico de información: esta comunicación se da entre equipos que comunican siguiendo un cierto patrón de comportamiento de tipo periódico, por medio de tablas simples o múltiples (por ejemplo, lectura de un sensor).

- Intercambio de información de grandes dimensiones: a menudo se producen transacciones de grandes paquetes entre nodos de una misma capa, normalmente en las capas altas (por ejemplo, transferencia de un diseño CAD en capa de planta) o transacciones entre capas (por ejemplo, carga de programas entre un diseño CAM y una máquina, etc.).

- Intercambio de información de pequeñas dimensiones: estos intercambios acostumbran a producirse en las capas inferiores de la pirámide de la productividad (por ejemplo, órdenes a actuadores, lectura de sensores y dispositivos).

10.8. NORMALIZACIONES EN LOS BUSES DE CAMPO

La parte esencial de las funcionalidades de los buses de campo existe desde que se desarrollaron las primeras aplicaciones de control distribuido hacia la década de 1950. Se crearon soluciones a medida para interconectar dispositivos a computadores, principalmente en el seno de fabricantes de equipos industriales. A medida que las tecnologías digitales fueron avanzando y que cada fabricante utilizaba sus soluciones particulares, los integradores de sistemas se encontraron con dificultades de compatibilidad al enlazar equipos de distinta procedencia en aplicaciones de control distribuido. Las organizaciones de normalización industrial (ISO, Cenelec, IEC, etc.) promovieron la búsqueda de soluciones comunes y documentadas que permitiera tanto a fabricantes como integradores el uso de métodos comunes.

Desde 1980 se han propuesto diferentes soluciones para buses de campo con espíritu abierto, o sea, con especificaciones completas disponibles para los posibles usuarios (en contraposición a las especificaciones cerradas normalmente propietarias de algún fabricante). Cada una de estas soluciones tiene un conjunto de propiedades que les proporciona un buen servicio para aplicaciones en casos determinados. Pero a su vez, suelen ser incompatibles entre si al estar definidas con tecnologías y protocolos dispares. En 1999 se tomo la solución de compromiso de aceptar un conjunto de propuestas como las soluciones normalizadas de los buses de campo (para más detalle consúltese el capítulo 7 de [ZUR-05]). La propuesta IEC-61158 incluye los perfiles Fieldbus Foundation H1, HSE y H2, ControlNet, Ethernet/IP, Profibus-DP, Profibus-PA, PROFInet, P-net, WorldFIP, Interbus, Swiftnet. A partir de esta fecha se ha estado trabajando en perfiles adicionales, algunos de ellos orientados a entornos específicos (aviónica, médica, marina, ferrocarril, etc.) mientras que otros se refieren a tecnologías emergentes, como es el caso de comunicación sin cable.

10.9. MODELO GENERAL DE OPERACIÓN DE LOS BUSES DE CAMPO

El modelo de operación de los buses de campo se basa en la especificación de tres de las capas definidas en el modelo ISO/OSI-RM, las correspondientes a la capa 1 (física), capa 2 (enlace) y capa 7 (aplicación) (Fig. 10.3). Pero además los buses de campo suelen añadir una capa superior (capa de usuario) que incorpora funcionalidades de interfase con las aplicaciones distribuidas.

La evolución hasta llegar a este modelo fue consecuencia de las distintas propuestas que se implementaron en las etapas iniciales de las comunicaciones entre computadores, entre las que cabe citar SNA (*Systems Network Architecture)* de IBM Corp., DNA (*Digital Netwoks Architecture*) de Digital Eq. Corp., XNA (*Xerox Network Architecture*) de Xerox Corp., que a su vez inspiraron el modelo ISO/OSI de interconexión de sistemas abiertos. En especial, y por lo que respecta a las comunicaciones entre equipos industriales, se tomaron como referencia los modelos desarrollados por Modbus, Bitbus, Hart y MAP. Merece la pena indicar que tanto algunos fabricantes de máquinas y dispositivos industriales como de integradores de automatización habían desarrollado sus propios sistemas de comunicación, frecuentemente sistemas propietarios y sin capacidad de interconectividad con otros sistemas.

FIGURA 10.3. Modelo de capas de los buses de campo.

Modbus (Modicon-1979) proporciona servicios de comunicación muy simples en los que se contemplan transacciones orientadas a bits (lectura-escritura de bits individuales o por grupos) y a carácter, en las transacciones entre un maestro controlador/supervisor y una zona de memoria del esclavo. Su simplicidad y al ser el protocolo abierto y público permite una implementación en nodos simples con microprocesadores de baja potencia y coste.

Bitbus (Intel-1983) (IEEE-1118) define una arquitectura propiamente distribuida. Los nodos se conectan a subredes que a su vez se interconectan mediante repetidores formando una estructura de árbol. Internamente Bitbus define unas estructuras abstractas en los nodos y unos servicios de comunicación orientados a tareas en los que las transacciones pueden ser de control de nodo (identificación, reset, cambio de tarea, estado, etc.) y de movimiento de datos (lectura, escritura). Entre los servicios de mensajería se definieron servicios genéricos (identificación de dispositivo, tarea 0, etc.) que

todos los dispositivos conectados a la red debían poder responder, y servicios específicos, dentro de tareas dependientes de la aplicación.

Hart (Rosemount-1987) (*Highway Addressable Remote Transducer*) es un buen ejemplo de transición de dispositivos industriales con comunicación analógica (4-20 mA) hacia la comunicación digital. La señal analógica en mA tiene superpuesta una modulación de corriente en baja frecuencia, que corresponde a los bits individuales de los mensajes digitales. Su capacidad de comunicación es muy reducida (tres o cuatro transacciones por segundo y hasta 15 nodos), por lo que es aplicable a dispositivos de procesos lentos, con constantes de tiempo de decenas de segundos o varios minutos. Pero los servicios de comunicación son extensos, hasta 510 servicios distintos. El protocolo, de tipo maestro-esclavo, especifica un subconjunto de transacciones Universales que cualquier nodo conectado a la red ha de poder servir (identificación, estado, variable primaria, rangos, etc.). Otras transacciones se especifican como Comunes, mientras que se pueden definir por el usuario otro subconjunto de transacciones Específicas dependiendo del equipo remoto.

MAP (*Manufacturing Automation Protocol*) fue una iniciativa promovida por General Motors en 1980 con el objetivo de normalizar las redes de las fábricas al constatar que los costes de interconexión de máquinas de distintos fabricantes podían llegar a alcanzar el mismo monto de las propias máquinas. Se creó un grupo de trabajo de normalización que tomó como referencia los trabajos previos en ISO/OSI y los propios desarrollos de fabricantes de máquinas industriales. Se diseñó una topología de tipo bus en dos niveles, figura 10.4, el primero para enlazar los centros neurálgicos de la empresa y el segundo mediante redes de célula en cada una de las fábricas, secciones y células. La decisión del método de acceso en el bus tenía que ser determinista por lo que se optó por el Paso de Testigo sobre Bus, posteriormente normalizada como IEEE 802.4-1984. MAP especificó todos y cada uno de los niveles de ISO/OSI, aunque contempló dispositivos de célula con funcionalidades reducidas y con definición de las capas física, enlace y aplicación (Mini-MAP) (Tabla 10.2). La parte más original de MAP se encontraba en la capa de aplicación. En ella se definen hasta 86 servicios de comunicación agrupados en distintas funcionalidades, desde E/S, supervisión, coordinación, usuario, ficheros, etc. Pero estos servicios estaban acompañados de unas funcionalidades abstractas definidas en la especificación de la mensajería de fabricación, MMS. En ella no sólo se indica de manera precisa cada transacción sino que se indica cual debe ser el comportamiento del sistema frente a cada transacción. Para ello se recurrió a la formalización mediante metodologías orientadas a objetos, definiendo abstractamente las estructuras de datos de comunicación, los

servicios que se pueden realizar sobre ellos y el comportamiento o las respuestas admisibles en dichas transacciones. Los resultados se recogieron en la especificación MAP 3.0 (1984), luego refrendados y ampliados en 1992.

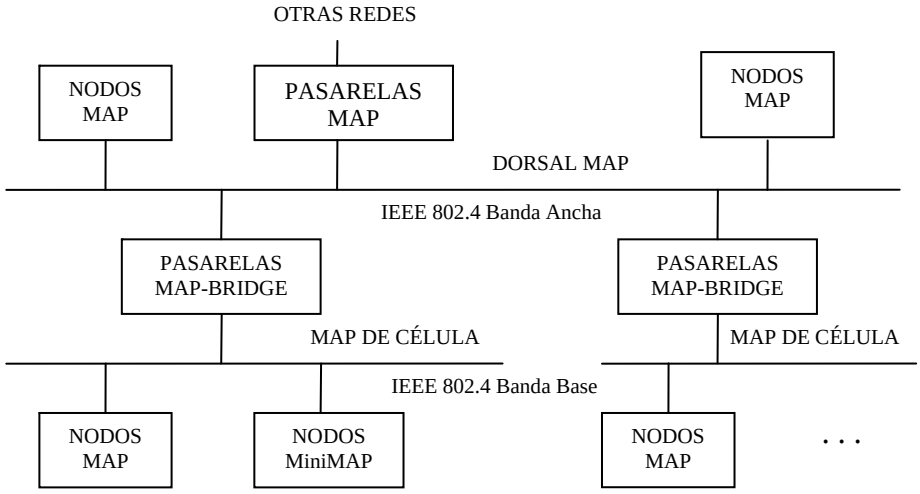


FIGURA 10.4. Arquitectura del modelo MAP.

TABLA 10.2. Modelo de capas en MAP3.0

ISO/OSI	MAP 3.0	MiniMAP
7 Aplicación	MHS DS FTAM NM MMS	MMS
6 Presentación	ISO 8822 / 8823	-
5 Sesión	ISO 8326 / 8327	-
4 Transporte	T clase 4	-
3 Red	CLNP	-
2 Enlace	LLC tipo 1 MAC TB 802.4 (MAC Eth 820.3)	LLC tipo 1 MAC TB 802.4
1 Física	Banda ancha 10 Mbps	Banda Base 5 Mbps

Entre las ampliaciones se definieron los estándares acompañantes a MMS (*Companion Standards*), que tenían como objetivo el desarrollo de modelos MMS completos para cierto tipo de máquinas comunes en las fábricas (robots, controles numéricos, autómatas programables, etc.). Cada uno de estos CS incluía las características comunes y ampliadas de estos tipos de máquinas, de tal manera que cualquier máquina de este tipo pudiera instanciarse dentro del CS. Así, desde el punto de vista de comunicación (y por supuesto, automatización), todos los robots y sus objetos internos se llaman de la misma manera (si se han instanciado, o sea si están disponibles en el robot en particular) y por ello se facilita tanto la programación como la explotación remota de estas máquinas.

Como ya se ha indicado, los buses de campo actuales han tomado como referencia los trabajos y desarrollos previos, y se ha llegado a un consenso de modelo de capas simple y con funcionalidades a la medida de las aplicaciones sobre las que tiene que trabajar.

La reducción de las capas lleva asociada dos grandes consecuencias: Una reducción del coste computacional del nodo, al reducirse el número de funciones que se deben ejecutar en cada mensaje recibido o transmitido. Por otro lado, permite una simplicidad del *hardware* de los nodos que redunda en que estos pueden ser de bajo coste económico. Las capas no presentes dejan sus funcionalidades debido a que los buses de campo habitualmente constan de un único segmento y las extensiones se realizan con repetidores o pasarelas sencillas. Si se requiere alguna de las funcionalidades de estas capas, se suelen incluir en la capa de enlace o en la de aplicación.

La capa física se diseña con características que están especialmente adaptadas a su uso en entornos duros de fabricación y de industria. Esta capa se diseña con criterios de robustez frente a radiaciones electromagnéticas, seguridad intrínseca, e incluso de uso de par trenzado, común en la industria.

En la capa de enlace se utilizan variaciones de las subcapas MAC habituales en las redes locales, desde acceso aleatorio al medio (en CAN y sus extensiones) a paso de testigo (Profibus, P-net, Interbus) reserva temporal (TTP), o sondeo (Profibus-DP y PA, ASi). En esta capa se realizan tareas de asignación de direcciones, enlace entre nodos adyacentes, detección de errores, y confirmación de recepción si está activada esta funcionalidad.

La capa de enlace en CAN está basada en el protocolo aleatorio de acceso al medio pero a diferencia de Ethernet, el inicio de los mensajes siempre está sincronizado a nivel de bit. En la primera parte de la transmisión se realiza una

contienda entre los posibles mensajes sincronizados, la cual finaliza con un único mensaje transmitido, mientras los demás participantes en esta contienda se retiran hasta la próxima sincronización de mensaje.

Profibus define nodos Maestros y nodos Esclavos. Su capa de enlace utiliza un protocolo de paso de testigo entre los distintos nodos Maestros, mientras que los Maestros, cuando adquieren el testigo, se enlazan con los esclavos mediante consulta directa y funcionalidad maestro-esclavo. La Figura 10.5 muestra el modelo básico de Profibus DP y el enlace DP-PA. En este último sólo se realiza una conversión de la capa física (cambio de frecuencia y transmisión por corriente) pero la funcionalidad maestro-esclavo se mantiene.

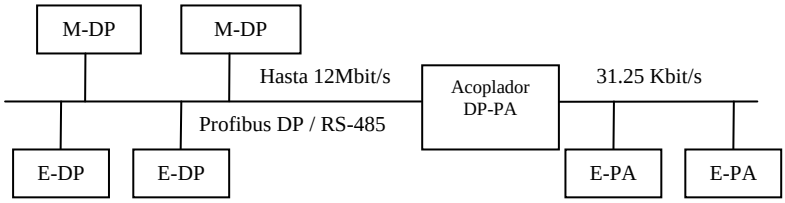


FIGURA 10.5. Modelo básico de Profibus DP/PA.

P-net asigna cíclicamente una fracción de tiempo de transmisión a cada nodo. Si un nodo detecta que puede empezar a transmitir pero no necesita usar el canal, el silencio en la red (durante un tiempo prefijado) se interpreta como que este nodo cede el enlace al siguiente nodo de la cadena. La figura 10.6 muestra el mecanismo de cesión virtual de testigo frente al silencio en la red.

La capa de enlace de Interbus es un ejemplo de paso de testigo sobre anillo lógico, pero a diferencia de otros anillos, la capa física se estructura en modo bus (forzando que cada nodo tenga tanto el circuito de ida como de vuelta en el anillo y éste se cierra sobre si mismo en el último nodo de la red, Fig. 10.7). Aunque la funcionalidad es de paso de testigo, la operación cíclica universal y la forma de construcción del mensaje convierten esta capa de enlace en un transporte de datos de y hacia un nodo maestro en el que en cada ciclo se actualiza una copia de los objetos de comunicación en un registro de entrada de tantos bits de ancho como bits en la red y además, en cada ciclo se copian los bits del registro de salida en cada uno de los nodos de la red. El resultado es que las aplicaciones externas al bus ven y depositan los bits distribuidos en la memoria formada por los registros de entrada y salida del nodo maestro. El proceso en paralelo de enlace hace las actualizaciones entre el nodo maestro y los nodos distribuidos.

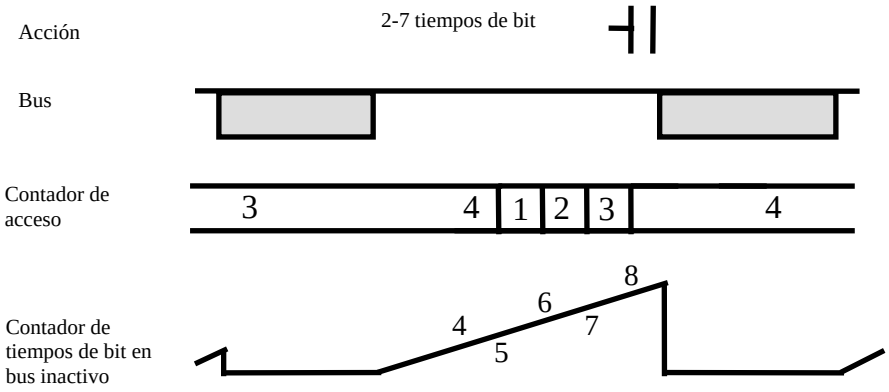


FIGURA 10.6. Modelo de acceso en una red P-net de cuatro nodos.

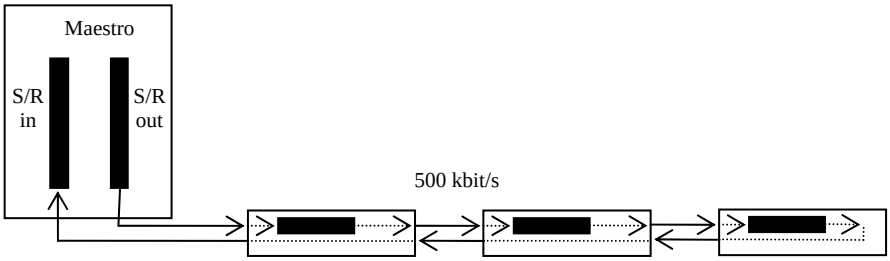


FIGURA 10.7. Topología de bus físico sobre anillo lógico en una Interbus.

La capa de aplicación utiliza básicamente las funcionalidades de Send y Receive de la capa de enlace para proveer servicios de mensajería compleja a los usuarios de la capa externa. En ciertos buses de campo esta capa se omite, dejando que la capa externa gestione todos los mensajes, independientemente del tipo y funcionalidad, mediante acceso directo a la capa de enlace (CAN, Interbus). Otros buses de campo incluyen una capa de aplicación explícita que gestiona desde las conexiones, el tipo de mensajería (control, supervisión, datos) y otras funcionalidades de esta capa. Así CANOpen, CAN-King, SDS, DeviceNet, NMEA2000 y otros buses de campo definen capas de usuario por encima de la capa de enlace de CAN.

La capa de usuario tiene como objetivo el facilitar que dispositivos y nodos de distintos fabricantes sean interoperables, o sea, que se puedan interconectar desde el punto de vista funcional. Para ello, deben conformar las tres capas inferiores, pero además las aplicaciones deben ver estos nodos de manera única.

El modelo que sigue esta capa incluye un conjunto de objetos de descripción e intercambio de datos: Perfiles de dispositivos; Descripciones de dispositivos; Lenguaje de descripción de dispositivos; Bloques funcionales.

- Los *perfiles de dispositivos* (*profiles*) consisten en una definición precisa de normalizaciones de dispositivos para aplicaciones específicas, a la manera de los estándares acompañantes a MMS, en la forma de la definición abstracta de uno o varios tipos de dispositivos, con indicación de los objetos de comunicación en ellos definidos, de las operaciones que se pueden realizar sobre ellos y de su comportamiento estático y dinámico. La definición de objetos lleva implícita tanto su codificación como sus formatos y los tipos asociados. Los perfiles se refieren a dispositivos como válvulas, codificadores, sensores, motores, etc. centrándose para cada uno de ellos en las partes comunes y permitiendo asociar a cada dispositivo la parte de perfil que tiene activado.

- La *descripción de dispositivos* consiste en la descripción de como cada dispositivo presenta sus datos. Incluye el nombre, el fabricante, el software interno, las revisiones, los datos y objetos de comunicación, la estructura operativa, los procedimientos, los formatos, los tipos, y otros. Esta descripción se deja accesible a todos los otros dispositivos de la red por medio de disquetes (una BD central) o puede estar integrada dentro de la propia ROM del dispositivo

- El *lenguaje de descripción de dispositivos* proporciona la semántica de definición y exportación de las descripciones de dispositivos, por medio de una gramática formal

- Los *bloques funcionales* son funciones de automatización compactas que realizan funciones determinadas generalmente asociadas a la automatización. Como ejemplo, un bloque funcional PID consiste en una función compleja que realiza el control proporcional, integral y derivativo sobre unas variables de planta, accesibles en el propio nodo que realiza la función, o en otros nodos conectados al bus de campo. Este bloque funcional se configura en las variables de operación para cada aplicación concreta. El resultado es que esta función no hace falta programarla (externamente) sino solamente configurarla (en la capa de usuario).

En esta capa también se realizan tareas de gestión, configuración, supervisión e identificación. En ella se definen desde bloques funcionales como

gramáticas y bases de datos de descripción de dispositivos. Las bases de datos de descripción de dispositivos hacen un seguimiento de los nodos conectados al bus de campo, identificándolos, configurándolos y siguiendo su operación (ficheros GSD en Profibus).

10.10. FUNCIONALIDADES COMUNES

Las propiedades de comunicación de los distintos buses de campo varían entre ellos al estar fundamentadas en distintos principios de comunicación. Las partes comunes se refieren principalmente a las características de predictibilidad, especialmente en algunos buses que permiten evaluar las latencias máximas, como enlace con las características de tiempo real. Un sistema de comunicaciones es de tiempo real cuando todos los protocolos involucrados en sus transacciones están acotados en el tiempo. Cuando se conoce el peor tiempo de una transacción se puede evaluar el coste temporal del protocolo asociado y, si el bus de campo lo admite, se puede planificar la mensajería conjunta con garantías de cumplimiento de todos los plazos. En la evaluación de estos plazos intervienen los métodos que se utilizan para servir las transacciones. Se suelen considerar tres modelos de transacción principales, aunque existen variantes de los mismos. Corresponden a los modelos Maestro-Esclavo, Productor-Consumidor y Cliente-Servidor (Fig. 10.8).

Maestro-Esclavo: Una entidad gobierna todos los servicios de transacción. Orden ----------> Respuesta
Productor-Consumidor: Cada entidad produce información, que adquieren los consumidores. Consumo <---------- Producción
Cliente-Servidor: Dos entidades cooperan para proporcionar servicios de transacción. El cliente realiza una petición que el servidor procesa y sirve. Petición ----------> Indicación Confirmación <---------- Respuesta

FIGURA 10.8. Modelos de transacción en los buses de campo.

El modelo Maestro-Esclavo consiste en que uno de los nodos actúa como Maestro y es el que dirige todas las transacciones, mientras que los nodos Esclavos no tienen iniciativas de comunicación sino que sólo responden a los comandos del nodo Maestro. El nodo Maestro centraliza todas las operaciones y por ello mantiene la integridad de todo el sistema, mediante una planificación en tablas o cíclica interacciona con el sistema, mientras que los nodos Esclavos, al ser reactivos, pueden requerir una funcionalidad limitada y por ello, económica. Este modelo es adecuado para mensajería cíclica y de carácter centralizado, aunque puede absorber mensajería esporádica desde los nodos Esclavos sujeta a las esperas de búsqueda exploratoria del nodo Maestro. Profibus DP y PA, P-net, Interbus y AsI entre otros utilizan este modelo de manera directa o indirecta.

El modelo Productor-Consumidor consiste en la definición en los distintos nodos de la red de las características de producción de datos y, en cada uno de los nodos, del consumo de los datos cuando circulan por la red. Los nodos productores tienen asociado un procedimiento de producción de mensajes que puede ser disparado por tiempo o por evento. En la producción del mensaje, éste se propaga por toda la red y todos los nodos tienen la opción de copiar el contenido de dicho mensaje. Habitualmente esta mensajería está orientada al contenido, o sea, el mensaje indica los objetos de comunicación que éste traslada. Por ello, los nodos consumidores utilizan los objetos de comunicación a la medida que estos se propagan por la red. La naturaleza distribuida de este modelo de comunicación requiere que se planifiquen de manera conjunta los instantes de producción en cada nodo, cuando la mensajería es periódica, o la saturación de los nodos, cuando la mensajería es esporádica, a fin de garantizar la planificabilidad de la red. Por otro lado, todos los nodos de la red participan en las transacciones y por ello, su complejidad suele ser algo superior a la de los nodos puramente esclavos. CAN, WorldFIP y Fieldbus Foundation son algunos de los buses de campo que utilizan este modelo.

El modelo Cliente-Servidor caracteriza las transacciones entre iguales. En la red se encuentran nodos Clientes, nodos Servidores y nodos mixtos. Los clientes piden servicios que los servidores, cuando tienen oportunidad los sirven. La complejidad de los nodos depende del grado y tipo de servicio a suministrar. Los servicios de mensajería se organizan a partir de cuatro primitivas: En el lado del cliente se inicia la transacción mediante el envío de la primitiva de Petición (*Request*). El servidor recibe la primitiva de Indicación (*Indication*), que le indica el tipo de servicio a proporcionar. Cuando éste tiene preparado el servicio retorna una primitiva de Respuesta (Response), que cuando la recibe el cliente corresponde a la primitiva de Confirmación (*Confirm*).

Estos modelos de transacción se utilizan ampliamente en los buses de campo de manera que en ciertas etapas se utiliza uno de los modelos y en otras, otro de los modelos. Así, en la etapa de configuración de la red se suele utilizar el modelo Cliente-Servidor mientras que durante la etapa de explotación se utiliza alguno de los restantes.

10.11. CONCLUSIONES

En la actualidad el uso de los buses de campo se ha consolidado de manera que las nuevas implantaciones de automatización incluyen capacidad de comunicación tanto en el ámbito de automatización (señales y mando) sino también de coordinación y gestión. Pero no siempre los diseñadores disponen de suficiente información o conocimientos como para sacar todo el provecho que estas tecnologías admiten. Por ello es conveniente no solamente conocer cada una de las tecnologías asociadas a los distintos buses de campo sino también los aspectos comunes entre ellas, que permitan extraer parámetros de diferenciación y de relación mutua.

En este capítulo de buses de campo se han expuesto los aspectos comunes y abstractos de la comunicación en entornos industriales, haciendo hincapié en las funcionalidades requeridas en cada una de las capas del modelo de capas, y además justificando la inclusión de una nueva capa, la de usuario, que pretende organizar y simplificar las aplicaciones externas que hacen uso de estos buses de campo.

En el desarrollo de estos modelos de comunicación han influido por un lado las experiencias de las tecnologías anteriores pero específicas para el entorno de fabricación (en especial el modelo MAP y el MMS) y además las aportaciones de los fabricantes de equipos y sistemas de automatización (Rockwell Automation, Siemens, Foxboro, Schneider, etc.) que en su momento buscaron soluciones propietarias para las comunicaciones de planta (generalmente de manera privada e interna a sus propios dispositivos pero sin posibilidades de intercambio de datos eficiente con dispositivos de otros fabricantes) pero con una gran experiencia práctica y por ello una garantía de operación.

10.12. BIBLIOGRAFIA

R. Zurawski. The industrial communication technology handbook. Ed. CRC Press, 2005.

N.P. Mahalik. Fieldbus technology: Industrial network standards for real-time distributed control. Ed. Springer, 2003.

K. Bende. PROFIBUS. The fieldbus for industrial automation. Ed. Prentice-Hall, 1993.

MAP/TOP 3.0. Manufacturing Automation Protocol/Technical and office protocol specifications, General Motors/Boeing Co, August. ISO DIS 9506 Manufacturing Message Specification Part 1 & 2 December 1988; Parts 3 - 7: MMS Companion Standards, 1991.

J.R. Pimentel. Communication networks for manufacturing. Ed. Prentice-Hall, 1989.

10.13. EVALUACIÓN

10.13.1. Ejercicios resueltos

Ejercicio 10.1. Sistema Distribuido de Control

Un cierto bus de campo consiste en una capa física con topología de bus, una capa de enlace que opera en modo maestro-esclavo, y una capa de aplicación que realiza tareas de control y supervisión mediante la lectura de los dispositivos sensores y los comandos de control a los dispositivos actuadores. El Sistema Distribuido (SD) está formado por cinco bucles de realimentación, cada uno de ellos formado por un sensor, un actuador y una tarea de control en el nodo maestro. Indique cómo sería la operación cíclica de control de los cinco bucles de realimentación

Solución: En el nodo maestro y en los nodos esclavos se realiza la siguiente secuencia de acciones:

a) El nodo maestro inicia un ciclo con el envío de un mensaje de consulta de la variable controlada al sensor del primer bucle y se queda a la espera de la respuesta del nodo sensor.

b) El sensor, cuando recibe la orden de consulta del maestro, prepara el valor de la variable controlada y retorna un mensaje con dicho valor.

c) El maestro calcularía el valor de la variable manipulada mediante la ejecución del algoritmo de control y, a continuación, enviaría una orden

al dispositivo actuador con un mensaje de actuación. El nodo maestro se queda a la espera de recibir la confirmación del nodo esclavo.

d) El nodo actuador recibiría la orden, ejecutaría la acción de control y retornaría al nodo maestro un mensaje de confirmación de la ejecución de la orden.

e) Se realizarían las acciones a, b, c, d, para cada uno de los bucles realimentados.

f) El nodo maestro se queda a la espera de iniciar un nuevo ciclo, normalmente sincronizado por un reloj que debe indicar un periodo más largo que el máximo tiempo de ejecución de todas las acciones previas.

10.13.2. Evaluación objetiva

1. El medio de transmisión usual en la capa física de un bus de campo consiste en:

   A) Un par de cables para cada dispositivo conectado

   B) Un par trenzado de hilos en conexión topológica de bus

   C) Un medio sin hilos (*wireless*)

   D) Señales transmitidas por variaciones de presión en un fluido compresible

2. Los buses de campo son:

   A) Mecanismos de transporte de pasajeros en una planta industrial

   B) Conducciones para distribución de materia prima en los almacenes de las fábricas

   C) Sistemas de comunicación para transmisión de mensajes entre dispositivos y máquinas industriales

   D) Redes de comunicación de las oficinas técnicas en las empresas de transporte

3. Un sistema distribuido industrial es:

   A) Un sistema de gestión de pedidos en una gran empresa industrial

   B) Un sistema de análisis de los riesgos de fabricación

   C) Un sistema industrial en el que la fabricación se realiza en un número finito de naves industriales

   D) Un conjunto de máquinas, dispositivos y equipos que se hallan interconectados en una organización industrial

4. La mensajería en un bus de campo consiste en:

   A) El intercambio de información entre proveedores y almacén

   B) El intercambio de datos operativos entre controladores, dispositivos, supervisores y máquinas en un entorno industrial

   C) La preparación de un mensaje único que recoja toda la información producida en una fábrica

   D) La creación de estadísticas de funcionamiento de las máquinas

5. El modelo de capas de un bus de campo:

   A) Es un modelo abstracto de comunicación que agrupa las funcionalidades y los protocolos de una forma común

   B) Es un modelo genérico que agrupa la estructura funcional y las apariencias visibles (forma, color) de los dispositivos industriales

   C) Es el resultado de la iniciativa de algunos fabricantes para que los usuarios no compren equipos de la competencia

   D) Consiste en el mínimo número de funciones y servicios de comunicación del bus de campo

6. MAP, en el contexto de las comunicaciones industriales se refiere a:

   A) Métodos de Análisis de Protocolos

   B) Mensajería Abstracta y su Programación

   C) Protocolos para Automatización de la Fabricación

   D) Mensajería Abierta con Protocolos

7. Se define el modelo cliente servidor como:

   A) Una entidad gobierna todos los servicios de transacción

   B) Cada entidad produce información, que adquieren los consumidores

   C) Dos entidades cooperan para proporcionar servicios de transacción. El cliente realiza una petición que el servidor procesa y sirve

   D) Una entidad produce información para el resto de entidades

8. El modelo de operación de los buses de campo se basa en:

   A) El modelo ISO, capa 1, capa 2 y capa 7

   B) El modelo ISO, capa 1, capa 3 y capa 7

   C) El modelo ISO, capa 1, capa 2 y capa 6

   D) El modelo ISO, capa 1, capa 2

9. Las capa de usuario incluye un conjunto de objetos de descripción e intercambio de datos, ¿Cuál de los siguientes objetos de descripción no esta incluida en dicha capa?

   A) Bloques funcionales

   B) Leguaje de descripción

   C) Lenguaje de descripción de dispositivos

   D) Objetos de control de errores

10. ¿Cuál de las siguientes características no es de Profibus?:

    A) Define nodos Maestros

    B) Define nodos Servidores

    C) Define nodos Esclavos

    D) Su capa de enlace utiliza un protocolo de paso de testigo entre los distintos nodos Maestros

ANEXO

SOLUCIONES A LAS PREGUNTAS DE EVALUACIÓN OBJETIVA

SOLUCIONES A LAS PREGUNTAS DE EVALUACIÓN OBJETIVA

TEMA	1	2	3	4	5	6	7	8	9	10
1)	C	B	D	D	B	C	B	C	C	B
2)	A	C	C	C	C	B	D	B	C	C
3)	C	D	D	A	A	B	C	A	D	D
4)	D	B	D	C	B	A	B	B	C	B
5)	A	B	C	A	A	D	A	A	C	A
6)	B	D	B	B	C	C	A	D	D	C
7)	A	A	D	B	D	C	A	B	A	C
8)	B	B	D	A	D	D	C	C	B	A
9)	D	A	B	C	C	C	C	A	C	D
10)	A	D	C	B	D	B	A	D	B	B

52520UD01A01
62306UD01A01

Se presentan dos libros sobre el tema de la aplicación industrial de las comunicaciones, **Comunicaciones Industriales: Principios básicos**, y Comunicaciones Industriales: Sistemas Distribuidos y Aplicaciones, publicados en la Editorial de la UNED, como aportación de un grupo de profesores de cinco universidades públicas españolas (Universidad Nacional de Educación a Distancia, Universidad Politécnica de Valencia, Universidad Politécnica de Cataluña, Universidad de Vigo y Universidad de Oviedo) coordinados con la experiencia en trabajo colaborativo desde la propia UNED, para aportar los conocimientos más avanzados dentro de una profesión en la que, hasta ahora, las comunicaciones han sido una de las materias menos desarrollada. Ambos libros presentan un objetivo común: cubrir los temas de las comunicaciones industriales, pero de una forma complementaria entre sí.

Las comunicaciones industriales son una de las áreas en auge dentro del amplio mundo de las comunicaciones, al unirse en un mismo entorno los temas empresariales (y más en detalle, los temas de fabricación más ligados a la industria en su más amplio y básico aspecto) y las comunicaciones, como completo (y complejo) soporte para la implantación tecnológica dentro de la empresa.

El primer libro, **Comunicaciones Industriales: Principios básicos**, pretende aportar a su lector de los conocimientos previos necesarios para poder entender, conocer y asimilar las comunicaciones, desde las bases de las comunicaciones analógicas y digitales, la codificación, medios y modos de transmisión, los protocolos y estándares existentes, los componentes y elementos de las redes locales, amplias o inalámbricas, sin olvidarse de la seguridad. Por último, este primer libro tiende un puente al segundo, estableciendo los principios de las comunicaciones industriales, los sistemas de tiempo real, los buses de campo y los sistemas distribuidos.

El segundo libro, Comunicaciones Industriales: Sistemas Distribuidos y Aplicaciones, presenta una visión más aplicada para un lector ya «introducido» en las comunicaciones, pero interesado en entrar como complemento en el mundo de las comunicaciones industriales.

ISBN: 978-84-362-5460-0
52520
9 788436 254600

UNIVERSIDAD NACIONAL DE EDUCACIÓN A DISTANCIA