REDES DE COMPUTADORAS

Julio Johnny Regalado Jalca
Vicente Fray Romero Castro
Marieta Del Jesús Azúa Menéndez
Leonardo Raúl Murillo Quimiz
Galo Roberto Parrales Anzúles
Yanina Holanda Campozano Pilay
Ángel Leonardo Pin Pin

Ingeniería y Tecnología

3ciencias

REDES DE COMPUTADORAS

Julio Johnny Regalado Jalca

Vicente Fray Romero Castro

Marieta Del Jesús Azúa Menéndez

Leonardo Raúl Murillo Quimiz

Galo Roberto Parrales Anzúles

Yanina Holanda Campozano Pilay

Ángel Leonardo Pin Pin

Editorial Área de Innovación y Desarrollo, S.L.

Quedan todos los derechos reservados. Esta publicación no puede ser reproducida, distribuida, comunicada públicamente o utilizada, total o parcialmente, sin previa autorización.

© del texto: **Los autores**

ÁREA DE INNOVACIÓN Y DESARROLLO, S.L.
C/ Els Alzamora, 17 - 03802 - ALCOY (ALICANTE) info@3ciencias.com

Primera edición: **abril 2018**

ISBN: **978-84-948577-2-0**

DOI: **http://dx.doi.org/10.17993/IngyTec.2018.32**

AUTORES

Julio Johnny Regalado Jalca. Ingeniero en Computación y Redes por la Universidad Estatal del Sur de Manabí, Magister en Educación Informática por la Universidad de Guayaquil. Investiga temas relacionados al Desarrollo de Software para Ingeniería Civil, Seguridad Informática, Redes Convergentes y Telecomunicaciones. Actualmente profesor de la Universidad Estatal del Sur de Manabí Carrera de Ingeniería Civil. Ecuador.

Vicente Fray Romero Castro. Ingeniero en Sistemas por la Universidad Laica Eloy Alfaro de Manabí, Magister en Sistemas de Información Gerencial por la Escuela Superior Politécnica del Litoral. Investiga temas relacionados a Tecnologías de Desarrollo de Software, Inteligencia de Negocios y metodologías Orientadas a Objetos. Actualmente Docente en la Universidad Estatal del Sur de Manabí.

Marieta del Jesús Azúa Menéndez. Ingeniería en Computación y Redes en la Universidad Estatal del Sur de Manabí, Tecnólogo Programador en Sistemas por la Universidad Técnica de Manabí, Magister en Educación Informática por la Universidad de Guayaquil, Maestrante en Matemática y computación por la Universidad Internacional de la Rioja. Investiga temas relacionados a Telecomunicaciones, Informática, Técnicas Multivariantes, matemáticas y computación. Actualmente profesor de la Universidad Estatal del Sur de Manabí Carrera de Laboratorio Clínico. Ecuador.

Leonardo Raúl Murillo Quimiz. Ingeniero en Computación y Redes por la Universidad Estatal del Sur de Manabí, Magister en Educación Informática por la Universidad de Guayaquil. Investiga temas relacionados con redes y telecomunicaciones. Actualmente profesor y coordinador del área de seguimiento a graduados de la Universidad Estatal del Sur de Manabí. Ecuador.

Galo Roberto Parrales Anzúles, Licenciado en Análisis de Sistemas por la Universidad Estatal del Sur de Manabí, Magister en Educación Informática por la Universidad de Guayaquil. Investiga temas relacionados con Redes y Telecomunicaciones, Plataformas Educativas E – Learning. Actualmente Docente de la Universidad Estatal del Sur de Manabí. Ecuador.

Yanina Holanda Campozano Pilay. Ingeniera en Computación y Redes por la Universidad Estatal del Sur de Manabí, Magister en Gerencia Educativa. Actualmente Docente de la Universidad Estatal del Sur de Manabí. Ecuador.

Ángel Leonardo Pin Pin. Obtuvo su pre-grado de Ingeniería en Computación y Redes en la Universidad Estatal del Sur de Manabí, su pos-grado o maestría en educación informática la realizo en la universidad de Guayaquil, ha realizado los cursos de CISCO – ESPOL, para fundamentar investigaciones en temas relacionados con redes y telecomunicaciones, protocolos LAN y WAN. Actualmente profesor de la Universidad Estatal del Sur de Manabí carrera laboratorio Clínico.

ÍNDICE GENERAL

PRÓLOGO.....**11**
CAPÍTULO I. REDES DE COMPUTADORAS.....**13**
1.1 Introducción.....13
1.2 Definiciones y conceptos de transmisión de datos.....13
1.3 Definición y concepto de Teleinformática.....13
1.4 Sistema de comunicación de datos.....14
1.5 Serie – Paralelo.....15
1.5.1. Multiplexación.....15
1.6 Canal de comunicaciones.....18
1.7 Tipos de transmisión.....18
1.8 Medios de Transmisión.....19
1.9 Medios guiados.....20
1.10 Medios no guiados.....22
1.11 Transmisión analógica y digital de datos.....23
1.12 ETD – ETCD.....24
1.12.1 Equipos terminales de datos (ETD).....25
1.12.2 Equipos terminales de circuitos de datos (ETCD).....25
1.13 Línea de comunicaciones.....25
1.13.1 Enlace de datos.....25
1.13.2 Circuito de datos.....26
1.14 Problemas en la transmisión.....26
1.15 Técnicas de detección de errores.....28
CAPÍTULO II. MÓDEMS.....**31**
2.1 Interfaz ETD/ETCD.....31
2.2 El Módem.....33
2.3 Modulación.....33
2.4 Tipos de modems.....36
2.4.1 El modem serie externo.....36
2.4.2 El modem interno.....37
2.4.3 El modem USB 39
2.4.4 El modem en formato PC Card.....40
2.5 Órdenes AT.....46
CAPÍTULO III: TEORÍA DE LA INFORMACIÓN.....**49**
3.1 Introducción a la Teoría de la Información.....49
3.2 Cantidad de información: Capacidad de un mensaje de despejar incertidumbre.....50
3.3 Fuentes discretas sin memoria.....51
3.4 Eficiencia de un código.....55
3.5 Código de Shannon-Fano.....59
3.6 Código de Huffman.....61
3.7 Códigos cíclicos.....76
CAPÍTULO IV. COMPRESION DE DATOS.....**77**
4.1 Introducción.....77
4.2 Tipos de compresión.....77
4.3 Tipos de compresión sin pérdidas.....78
4.4 Compresores basados en diccionario.....80

4.4.1 Selección del tamaño del diccionario.....81
4.4.2 Selección de las cadenas para completar el diccionario.....82
4.5 Compresión RLE.....84
4.6 LZ77.....85
4.7 Compresión con pérdidas.....87
4.8 Transformación Discreta por Coseno (DCT).....88
4.9 Estándar JPEG.....89
CAPÍTULO V. NOCIONES DE CRIPTOGRAFIA.....93
5.1 Introducción.....93
5.2 La seguridad de datos en los negocios.....94
5.3 Conceptos básicos de Criptografía.....95
5.4 Generalidades sobre los algoritmos criptográficos.....99
5.4.1 Algoritmos simétricos.....99
5.4.2 Algoritmos de clave pública.....99
5.5 Seguridad de un sistema criptográfico.....100
5.6 Generalidades sobre los algoritmos DES, IDEA y AES.....102
5.7 Firmas digitales.....111
CAPÍTULO VI. LOS PROTOCOLOS DE COMUNICACIÓN.....113
6.1 Introducción.....113
6.2 Concepto de protocolo.....113
6.3 Modelo de referencia OSI.....116
6.4 Niveles del modelo de referencia OSI.....116
CAPÍTULO VII. EL NIVEL DE ENLACE DE DATOS.....119
7.1 Introducción.....119
7.2 Protocolos de enlace de datos.....119
7.3 Funciones de un protocolo a nivel de enlace de datos.....119
7.4 Fases de un protocolo a nivel de enlace de datos.....122
7.5 Protocolos orientados a carácter.....122
7.6 Protocolos orientados a bit.....124
CAPÍTULO VIII. INTRODUCCIÓN A LA ARQUITECTURA TCP/IP.....125
8.1 ¿Qué es INTERNET?.....125
8.1.1 Historia de Internet.....125
8.2 Modelo OSI.....127
8.3 Arquitectura TCP/IP.....128
8.4 Retos a través del desarrollo de Internet.....128
8.5 Características distintivas de TCP/ IP.....129
8.6 Estructura de capas.....130
8.7 IP Protocolo Internet.....133
8.8 Fragmentación.....135
8.9 Direccionamiento IP.....143
8.10 Enrutamiento de Datagramas IP.....147
BIBLIOGRAFÍA.....163

Índice de Tablas

Tabla 1: Matriz de coeficientes....90
Tabla 2: Cuantización....91
Tabla 3: DIF=DC-PREV....91

Índice de Figuras

Figura 1: Transmisión serie paralelo....15
Figura 2: Medio de transmisión....16
Figura 3: Transmisión asíncrona....17
Figura 4: Transmisión síncrona....18
Figura 5: Transmisión simplex, semidúplex y dúplex....18
Figura 6: Conector RS232 DB-25....32
Figura 7: Diferentes métodos de modulación....35
Figura 8: Modem serie externo....36
Figura 9: Modem interno....37
Figura 10: Potencialidad de un código....70
Figura 11: Espacio de codificación....71
Figura 12: Límite de Harmming....72
Figura 13: Potencialidad de un código....72
Figura 14: Limite de Hamming relación entre K,N B....73
Figura 15: Caso particular de corrección de errores simple....73
Figura 16: Código algebraico....75
Figura 17: Descripción del algoritmo....103
Figura 18: Función de cifrado f....105
Figura 19: Trama de datos ETD....119
Figura 20: Segmentación y Bloque....120
Figura 21: Trama de datos ETD....120
Figura 22: Estructura de capas OSI 130
Figura 23: Arquitectura de niveles....132
Figura 24: Encapsulación en Ethernet....134
Figura 25: Ruteros responsables de la fragmentación....136
Figura 26: Formato del datagrama....137
Figura 27: Clase de Opción....140
Figura 28: Registro de Ruta....141
Figura 29: Direcciones clase C....145

PRÓLOGO

Actualmente las redes de computadoras (también llamadas red de ordenadores, red de comunicaciones de datos, red informática) conforman un conjunto de equipos informáticos y software conectados entre sí con la finalidad de compartir información, recursos y ofrecer servicios, lo que ha tomado gran importancia mundial, es totalmente indispensable para las organizaciones puesto que brinda mucha facilidad en el acceso a la información y una gran capacidad de almacenaje de datos. Es por este motivo que resulta muy grato dar a conocer este libro que se titula **"REDES DE COMPUTADORES**". Con este libro se pretende que el lector pueda despejar toda clase de interrogantes acerca de las redes de computadoras y le permita familiarizarse rápidamente con su contenido, no solo teórico sino también práctico mediante una serie de ejercicios.

Este libro está dirigido a los estudiantes universitarios que cursan carreras relacionadas con las tecnologías de la información, los sistemas de información o con la informática. El libro está pensado como texto para un curso de un cuatrimestre sobre las redes de computadoras, también se dirige a los profesionales como: analistas de sistemas, programadores de aplicaciones y personas en general que están relacionadas con el uso de redes o interesadas en aprender esta tecnología.

Cada capítulo es sumamente importante en el proceso de aprendizaje. El objetivo del libro es lograr que cada lector enriquezca sus conocimientos con los temas plasmados en cada capítulo, y servir como ayuda didáctica a cientos de docentes en el área de redes de computadoras.

LOS AUTORES

CAPITULO I. REDES DE COMPUTADORAS

Conceptos básicos de la comunicación de datos

1.1 Introducción

En esta época del desarrollo tecnológico, se puede expresar que la informática y las comunicaciones se encuentran en un grado tan alto de integración que es muy difícil determinar con exactitud cuál es la frontera entre estas disciplinas.

Las tecnologías usadas para satisfacer los problemas de comunicaciones y los de informática son exactamente las mismas.

La teleinformática, expresión que proviene de las palabras telecomunicaciones e informática, es el resultado de la irrupción de técnicas conocidas como de procesamiento distribuido, donde la importancia del puesto de trabajo individual, desplazó al criterio de cómputos centralizado, preparado para el procesamiento de datos.

1.2 Definiciones y conceptos de transmisión de datos

Se entiende por transmisión de datos al movimiento de información codificada, de un punto a uno o más puntos, mediante señales eléctricas, ópticas, eletroópticas o electromagnéticas.

Este requerimiento, originado en las organizaciones gubernamentales, industriales, comerciales, bancarias, empresariales, militares, etc., ha nacido por la necesidad de poner a disposición de ellas en un punto remoto la capacidad de proceso de un ordenador, ubicado en un punto que podríamos llamar central.

Ese punto puede estar dentro de la propia organización, próximo o alejado del ordenador central.

La diferencia importante reside en la distancia y la geografía del problema a considerar, pues en función de estos parámetros, puede ser necesario o no el uso de redes de comunicaciones.

1.3 Definición y concepto de Teleinformática

La palabra Teleinformática está constituida por la contracción de las palabras telecomunicaciones e informática. En ella se reúnen los aspectos técnicos de ambas especialidades.

Se puede definir la teleinformática como “la ciencia que estudia el conjunto de técnicas que es necesario usar para poder transmitir datos dentro de un sistema informático o entre puntos de él situados en lugares remotos o usando redes de telecomunicaciones”.

Lo que se intenta con la teleinformática es lograr que un ordenador pueda dialogar con equipos situados geográficamente distantes, reconociendo las características esenciales de la información como si la conexión fuera local, usando redes de telecomunicaciones.

1.4 Sistema de comunicación de datos

Los sistemas teleinformáticos o también llamados “sistemas funcionales de comunicación de datos” son formas de trabajo que en general responden a necesidades concretas de los usuarios informáticos que trabajan en la modalidad “fuera de planta o remota”.

Los principales **objetivos** que tiene que satisfacer un sistema teleinformático son los siguientes:

- Reducir tiempo y esfuerzo.
- Capturar datos en su propia fuente.
- Centralizar el control.
- Aumentar la velocidad de entrega de la información.
- Reducir costos de operación y de captura de datos.
- Aumentar la capacidad de las organizaciones, a un costo incremental razonable.
- Aumentar la calidad y la cantidad de la información.
- Mejorar el sistema administrativo.

Casos en que los sistemas informáticos, son especialmente aptos para que utilicen técnicas teleinformáticas:

- Cuando se desea reducir un elevado volumen de correo, de llamadas telefónicas o de servicios de mensajería.
- En los casos en que se efectúen muy a menudo operaciones repetitivas, tales como crear o copiar a ser procesados o directamente duplicar información ya procesada.
- Cuando sea necesario aumentar la velocidad de envío de la información, mejorando las funciones administrativas.
- En la ejecución de operaciones descentralizadas.
- Para mejorar el control, descentralizando la captura de datos y centralizando su procesamiento.
- En los casos en que es necesario disminuir riesgos en el procesamiento de la información, debido a problemas técnicos del hardware.
- Cuando sea menester mejorar la actividad de planificación en la organización.

Cada sistema teleinformática cubre un conjunto de necesidades y, por lo tanto, posee especiales características que diferencian unos de otros.

1.5 Serie – Paralelo

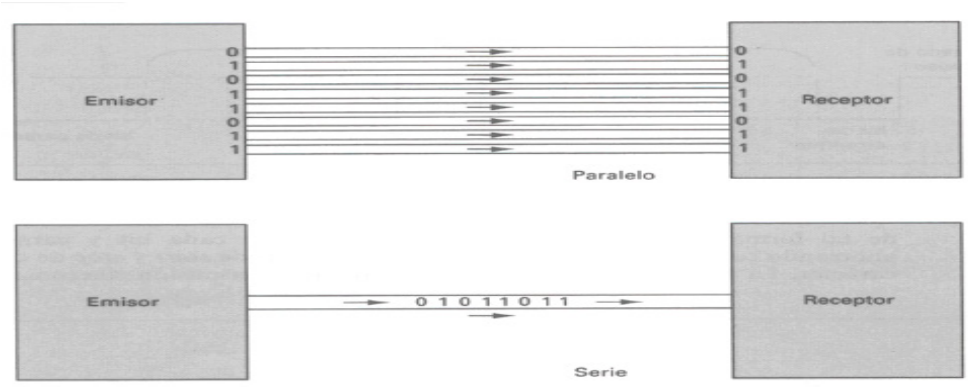


Figura 1: Transmisión serie paralelo

La transmisión en modo paralelo es aquella en la que los bits que componen cada Byte o carácter se transmiten en un solo ciclo de n bits. Posee las siguientes características:

- Este modo es que se usa en los ordenadores para realizar la transferencia interna de los datos.
- En estos casos se transmite cada conjunto de n bits, seguido por un espacio de tiempo y luego nuevamente otro conjunto de n bits, y así sucesivamente.
- En la transmisión en paralelo se pueden usar dos formas de transmisión distintas. Una es disponer de n líneas diferentes a razón de una por bit a transmitir: la otra, es usar una única línea, pero enviando cada bit mediante un procedimiento técnico que se denomina **multiplexación.**

1.5.1. Multiplexación

Algoritmos de reparto de canal. Técnica que permite a múltiples usuarios compartir una única línea.

Es deseable en aquellas situaciones donde se emplean líneas de transmisión de larga distancia (o de gran costa, aunque no sean de gran longitud), conseguir el máximo aprovechamiento de la capacidad del medio de transmisión. Básicamente, consiste en compartir un mismo medio de transmisión entre varias comunicaciones, con lo que se divide el coste asociado a cada comunicación individual. El esquema general del proceso se muestra en la siguiente figura:

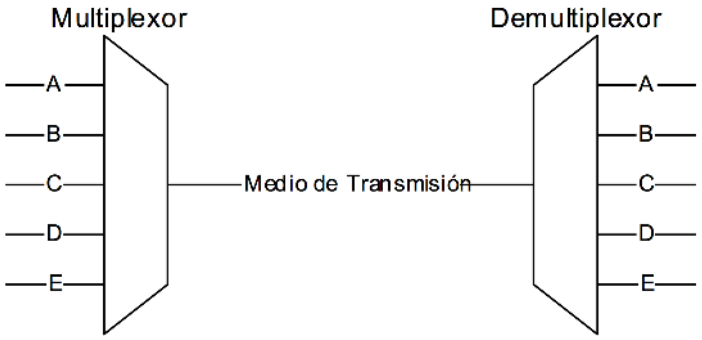


Figura 2: Medio de transmisión.

El proceso de multiplexación es reversible y por tanto permite la transmisión simultánea de varias estaciones (A, B, C, D) por el mismo medio de transmisión.
Los dos métodos básicos que nos permiten realizar la multiplexación, son la multiplexación por división del tiempo (MDT) y la multiplexación por división de la frecuencia (MDF).
En la multiplexación por división de la frecuencia se divide el ancho de banda del canal entre las señales a multiplexar. Cada estación puede transmitir simultáneamente a las demás. En este tipo de multiplexación, el ancho de banda del canal debe exceder a la suma de los anchos de banda de las señales que se transmiten, puesto que es necesario dejar bandas de seguridad entre las frecuencias asignadas a los diversos canales. Las señales transmitidas a través del medio deben ser señales analógicas. El espectro de frecuencia se divide entre los canales lógicos, con cada usuario en posesión única de alguna banda de frecuencia.
En el segundo método, se asigna a las diferentes estaciones un turno de transmisión rotativo, durante un quantum de tiempo transmite una estación, luego la siguiente, etc. Luego realmente cada estación utiliza el canal alternativamente, no de forma simultánea. Los usuarios esperan su turno, obteniendo cada uno, de forma periódica la banda entera durante un breve lapso.

- Cuando se usa la transmisión en paralelo, se emplean generalmente altas velocidades, dado que esa es precisamente, una de sus características más importantes: enviar más bits en menor tiempo posible. En estos casos las velocidades se miden en Bytes o caracteres por segundo.
- En general no se usa este tipo de transmisión, cuando las distancias superan las decenas de metros debido a que el tiempo de arribo de los bits difiere de una línea a otra, situación ésta que se agrava con el aumento de la distancia.

La transmisión en modo serie es aquella en las que los bits que componen cada carácter se transmiten en n ciclos de 1 bit cada uno. Posee las siguientes características:

- En este caso, se envían un bit detrás de otro, hasta completar cada carácter.
- Este modo de transmisión es el típico de los sistemas teleinformáticos.

En muchas ocasiones, las señales que son transmitidas por los vínculos de telecomunicaciones, al llegar a los equipos informáticos deben pasar al modo paralelo y viceversa. Este proceso de transformación se denomina deserialización y serialización, respectivamente.
La secuencia de los bits transmitidos se efectúa siempre al revés de cómo se escriben las cifras en el sistema de numeración binario. Cuando se transmite con bit de paridad, éste se transmite siempre en último término.

La transmisión en modo serie tiene dos procedimientos diferentes, el denominado asincrónico y el sincrónico.

En el procedimiento asincrónico, cada carácter a ser transmitido es delimitado por un bit denominado de cabecera o de arranque, y uno o dos bits denominados de terminación o de parada. El bit de arranque tiene funciones de sincronización de los relojes del transmisor y del receptor. El bit o bits de parada, se usan para separar un carácter del siguiente.

Figura 3: Transmisión asíncrona.

En el procedimiento sincrónico, existen dos relojes uno en el receptor y otro en el transmisor. La información útil es transmitida entre dos grupos, denominados genéricamente delimitadores.

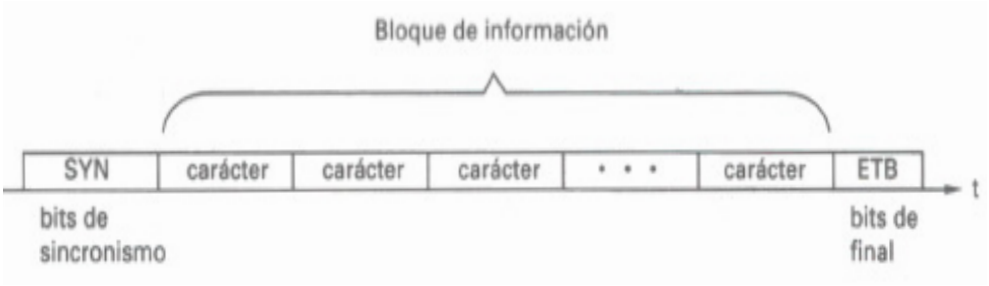


Figura 4: Transmisión síncrona.

1.6 Canal de comunicaciones

Es el recorrido físico que es necesario establecer para que una señal eléctrica, óptica o electro-óptica, se pueda desplazar entre dos puntos (uno llamado fuente y otro colector). Existen otras denominaciones tales como: línea, enlace, facilidad, etc.

Los canales pueden ser analógicos o digitales. Tanto los canales analógicos pueden llevar señales digitales previamente moduladas, como los canales digitales pueden llevar señales analógicas previamente digitalizadas.

1.7 Tipos de transmisión

Los distintos tipos de transmisión de un canal de comunicaciones pueden ser de tres clases diferentes:

- Simplex
- Semidúplex
- Dúplex (o dúplex completo)

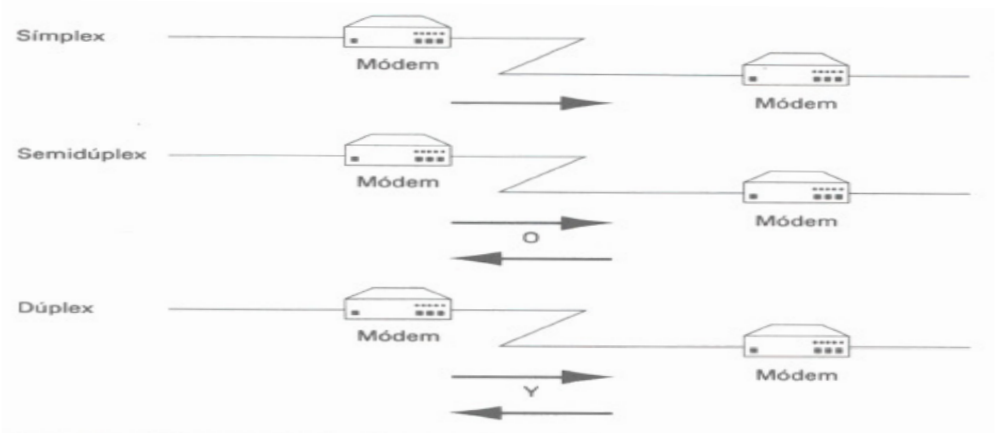


Figura 5: Transmisión simplex, semidúplex y dúplex.

Método simplex
Es aquel en que una estación siempre actúa como fuente y la otra siempre como colector. Este método permite la transmisión de información en un único sentido. Un ejemplo de servicio simplex es el que brindan las agencias de noticias a sus asociados

Método semidúplex
Es aquel en que una estación A en un momento de tiempo, actúa como fuente y otra estación corresponsal B actúa como colector; y en el momento siguiente, la estación B actuará como fuente y la A como colector. Este método permite la transmisión en las dos direcciones, aunque en momentos diferentes. Por ejemplo, la conversación entre dos radioaficionados que están dialogando, pero donde uno espera que el otro termine de hablar para continuar el diálogo; nunca pueden ambos hablar simultáneamente.

Método dúplex
Es aquel en que dos estaciones A y B, actúan como fuente y colector, transmitiendo y recibiendo información simultáneamente. Este método permite la transmisión en las dos direcciones, en forma simultánea. Por ejemplo, la conversación telefónica entre dos personas que no se escuchan y pretenden hablar simultáneamente.

1.8 Medios de transmisión

El medio de transmisión constituye el soporte físico a través del cual emisor y receptor pueden comunicarse en un sistema de transmisión de datos. Distinguimos dos tipos de medios: guiados y no guiados. En ambos casos la transmisión se realiza por medio de ondas electromagnéticas. Los medios guiados conducen (guían) las ondas a través de un camino físico, ejemplos de estos medios son el cable coaxial, la fibra óptica y el par trenzado. Los medios no guiados proporcionan un soporte para que las ondas se transmitan, pero no las dirigen; como ejemplo de ellos tenemos el aire y el vacío.
La naturaleza del medio junto con la de la señal que se transmite a través de él constituyen los factores determinantes de las características y la calidad de la transmisión. En el caso de medios guiados es el propio medio el que determina principalmente las limitaciones de la transmisión: velocidad de transmisión de los datos, ancho de banda que puede soportar y espaciado entre repetidores. Sin embargo, al utilizar medios no guiados resulta más determinante en la transmisión el espectro de frecuencia de la señal producida por la antena que el propio medio de transmisión.

Ancho de banda
Es la capacidad de transmisión de un canal de comunicación. La unidad de medida es bits por segundo (bps) y sus múltiplos (Kbps, Mbps, Gbps). También se expresa

frecuentemente en Bytes por segundo y sus múltiplos (KBps, MBps, GBps). A mayor ancho de banda, mayor cantidad de información se trasmite por unidad de tiempo y, como consecuencia, el proceso de comunicación de datos es más rápido.

1.9 Medios guiados

Pares trenzados

Este consiste en dos alambres de cobre aislados, en general de 1mm de espesor. Los alambres se entrelazan en forma helicoidal, como en una molécula de DNA. La forma trenzada del cable se utiliza para reducir la interferencia eléctrica con respecto a los pares cercanos que se encuentran a su alrededor, (Dos cables paralelos constituyen una antena simple, en tanto que un par trenzado no.)

La aplicación más común del par trenzado es el sistema telefónico, casi todos los teléfonos están conectados a la oficina de la compañía telefónica a través de un par trenzado. La distancia que se puede recorrer con estos cables es de varios kilómetros, sin necesidad de amplificar las señales, pero si es necesario incluir repetidores en distancias más largas. Cuando hay muchos pares trenzados colocados paralelamente que recorren distancias considerables, como podría ser el caso de los cables de un edificio de departamentos que se dirigen a la oficina de teléfonos, éstos se agrupan y se cubren con una malla protectora. Los pares dentro de estos agrupamientos podrían sufrir interferencias mutuas si no estuviera trenzado. En algunos lugares del mundo en donde las líneas telefónicas se instalan en la parte alta de los postes, se observan frecuentemente dichos agrupamientos, como cables con diámetros de varios centímetros.

Los pares trenzados se pueden utilizar tanto para transmisión analógica como digital, y su ancho de banda depende del calibre del alambre y de la distancia que recorre; en muchos casos pueden obtenerse transmisiones de varios megabits/s, en distancias de pocos kilómetros. Debido a su adecuado comportamiento y bajo costo, los pares trenzados se utilizan ampliamente y es probable que se presencia permanezca por muchos años.

El principal factor limitador de las líneas de par trenzado es causado por un fenómeno conocido como efecto piel. Al aumentar la velocidad de transmisión y por tanto la frecuencia de la señal transmitida, la corriente tiende a fluir únicamente por la superficie del cable. Esto tiene el efecto de aumentar la resistencia de los hilos para las señales de alta frecuencia, lo que provoca una mayor atenuación de las señales transmitidas. A altas frecuencias una cantidad creciente de la potencia de la señal se pierde debido a los efectos de la radiación. Por eso, tradicionalmente para aquellas aplicaciones que requieren una velocidad superior a 1Mbps solían emplearse otros medios de transmisión. Una línea de transmisión que minimiza ambos efectos es el cable coaxial.

Cable coaxial de banda base

El cable coaxial es otro medio típico de transmisión. Hay dos tipos de cable coaxial que se utilizan con frecuencia, uno de ellos es el cable de 50 ohms, se utiliza en la transmisión digital; en tanto que el otro tipo, el cable de 75ohms, se emplea en la transmisión analógica.

El cable coaxial consta de un alambre de cobre duro en su parte central, es decir, que constituye el núcleo, el cual se encuentra rodeado por un material aislante. Este material aislante está rodeado por un conductor cilíndrico que frecuentemente se presenta como una malla de tejido trenzado. El conductor externo está cubierto por una capa de plástico protector.

La construcción del cable coaxial produce una buena combinación e un gran ancho de banda y una excelente inmunidad al ruido. El ancho de banda que se puede obtener depende de la longitud del cable; para cables de 1km, por ejemplo, es factible obtener velocidades de datos de hasta 10Mbps, y en cables de longitudes menores, es posible obtener velocidades superiores. Se pueden utilizar cables con mayor longitud, pero se obtienen velocidades muy bajas. Los cables coaxiales se emplean ampliamente en redes de área local y para transmisiones de largas distancia del sistema telefónico.

Cable coaxial de banda ancha

El sistema que considera el otro tipo de cable coaxial emplea la transmisión analógica en el cableado que se utiliza comúnmente para el envío de la señal de televisión por cable, y se le denomina de banda ancha. Aunque el término banda ancha proviene del medio telefónico, en el cual se refiere a frecuencias superiores a los 4kHz, el significado de este término en el medio de redes de ordenadores se asocia a las redes de cables utilizadas para la transmisión analógica. Dado que las redes de banda ancha utilizan la tecnología patrón para envío de señales de televisión por cable, los cables pueden emplearse para aplicaciones que necesiten hasta los 300 Mhz (y en algunos casos hasta los 450 Mhz), y extenderse a longitudes que alcanzan hasta los 100 Km, gracias a la naturaleza analógica de la señal, que es menos crítica que la del tipo digital. Para transmitir señales digitales en una red analógica, cada interfaz debe tener un dispositivo electrónico que convierta en señal analógica el flujo de bits de envío, y otro para convertir la señal analógica que llega en un flujo de bits. Dependiendo del tipo (y precio) de estos dispositivos electrónicos, 1 bps puede llegar a ocupar un ancho de banda que va desde 1 a 4 Hz. Un cable típico de 300Mhz, por lo general puede mantener velocidades de transmisión de datos de hasta 150 Mbps.

Fibras ópticas

Una fibra óptica es un cilindro de pequeña sección (diámetro del orden de 2 a 125 µm), de un medio flexible –cristal, plástico- capaz de conducir un rayo óptico. Las mejores son las de silicio puro, y que tienen menos pérdidas, aunque su precio es

más elevado.
Un cable de fibra óptica consta de tres secciones concéntricas. La más interna, el núcleo, consiste en una o más hebras o fibras hechas de cristal o plástico. Cada una de ellas lleva un revestimiento de cristal o plástico con propiedades ópticas distintas a las del núcleo. La capa más exterior, que recubre una o más fibras, debe ser de un material opaco y resistente.
Un sistema de transmisión por fibra óptica está formado por una fuente luminosa muy monocromática (generalmente un láser), la fibra encargada de transmitir la señal luminosa y un fotodiodo que reconstruye la señal eléctrica.
La luz se propaga en zig-zag debido a los fenómenos de reflexión total que tienen lugar en el interior de la fibra. Por este motivo las pérdidas son muy escasas. Además, las fibras son inmunes a las interferencias electromagnéticas y a su vez no interfieren en otros sistemas. Por lo tanto, resultan extremadamente útiles para la transmisión de señales en medios muy ruidosos. Entre otras de sus ventajas podemos citar su elevado ancho de banda (permite alcanzar velocidades del orden de Gbps sobre decenas de Km) y sus reducidos peso y tamaño.
Los sistemas de fibra óptica suelen resultar más caros que los de cable coaxial y mecánicamente más delicados, por lo que presentan más dificultades en su instalación. Se utilizan en telecomunicaciones a larga distancia, aplicaciones militares, redes locales, distribución de señales de audio/vídeo.

1.10 Medios no guiados

En muchas ocasiones resulta problemática la instalación de un tendido. Los enlaces vía radio emplean la propagación de las ondas electromagnéticas en el espacio y por lo tanto no precisan de ningún tipo de cableado entre emisor y receptor. Dentro de los enlaces vía radio existen diferentes tipos según la banda empleada, exhibiendo diferentes propiedades. Los métodos más usuales son:

Radio enlaces de onda corta
La OC es una banda de radio, comprendida entre 2 y 15 Mhz aproximadamente, (aparece con las siglas SW en los receptores de radio). Poseen un alcance de miles de kilómetros, ya que se reflejan en la ionosfera y además son omnidireccionales, aunque sólo permite reducidas velocidades de transmisión, menores de 1200 bps. Aunque antaño fueron el medio más común, su uso actualmente se encuentra restringido a circunstancias especiales, debido a su limitada capacidad. Se emplea, por ejemplo, para la difusión de noticias de las agencias de teletipos de todo el mundo.

Radio enlaces de VHF y UHF
Estas bandas cubren aproximadamente desde 55 a 550 Mhz. Son también omnidireccionales, pero a diferencia de las anteriores la ionosfera es transparente a ellas. Su alcance máximo es de un centenar de kilómetros, y las velocidades que

permite del orden de los 9600 bps. Su aplicación suele estar relacionada con los radioaficionados y con equipos de comunicación militares, también la televisión y los aviones.

Microondas

Las microondas nos permiten transmisiones tanto terrestres como con satélites. Dada sus frecuencias, del orden de 1 a 10 Ghz, las microondas son muy direccionales y sólo se pueden emplear en situaciones en que existe una línea visual que une emisor y receptor. Los enlaces de microondas permiten grandes velocidades de transmisión, del orden de 10 Mbps.

Como hemos visto, existen diferentes tipos de enlaces vía radio, y cada uno nos permite cubrir un rango de distancias a diferentes velocidades de transmisión. No se olvide que en el caso de los enlaces de microondas las distancias, en tierra, de un enlace suelen ser de unos 30 a 50 Km. máximo. Sin embargo, en el caso de la comunicación con un satélite, si bien las distancias pueden ser de hasta 36.000 Km., sólo durante una pequeña parte del recorrido la señal se atenúa por el efecto de la atmósfera y el resto del trayecto es prácticamente en el vacío, que no atenúa la señal.

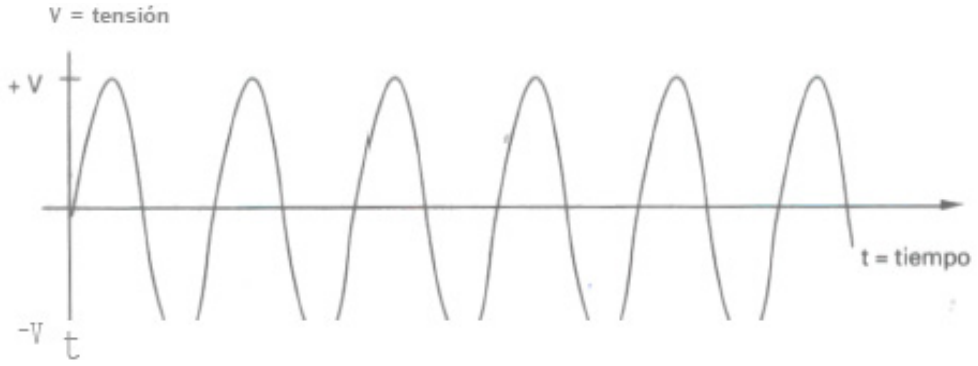
1.11 Transmisión analógica y digital de datos

Para transmitir datos es preciso transformarlos en una señal temporal que los represente y que pueda atravesar un determinado medio de transmisión para llegar al receptor en las mismas condiciones posibles.

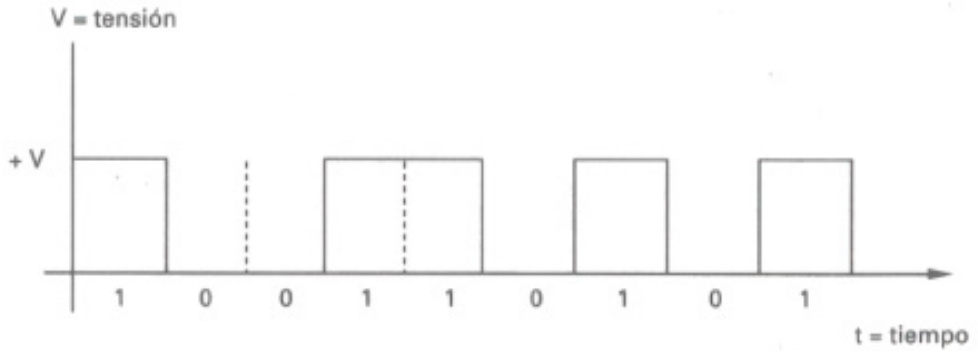
En las redes de ordenadores, los datos a intercambiar siempre están disponibles en forma de señal digital. No obstante, para su transmisión podemos optar por la utilización de señales digitales o analógicas. La elección no será, casi nunca, una decisión del usuario, sino que vendrá determinada por el medio de transmisión a emplear.

No todos los medios de transmisión permiten señales analógicas ni todos permiten señales digitales. Como la naturaleza de nuestros datos será siempre digital, es necesario un proceso previo que adecue estos datos a la señal a transmitir.

Una señal analógica representa una onda electromagnética que varía de forma continua. Dependiendo de su espectro, las señales analógicas pueden transmitirse por una amplia variedad de medios, por ejemplo, cables como el coaxial, la fibra óptica y medios de propagación espacial o atmosférica. Se utiliza en sistemas como telefonía, radio y televisión.

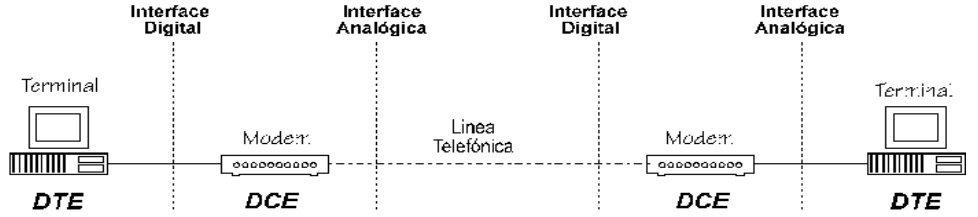


Una señal digital es una secuencia de pulsos de voltaje que pueden transmitirse por medio de un cable; por ejemplo, un nivel de voltaje positivo constante puede representar el uno binario y un nivel de voltaje negativo puede representar el cero binario. Sólo puede ser trasmitida de medios físicos.



1.12 ETD – ETCD

Los equipos terminales de datos producen señales digitales y solamente están en condiciones de recibir y procesar este tipo de señales. Sin embargo, estos datos deben ser transportados por redes de comunicaciones que sólo admiten señales del tipo analógicas. Por lo tanto, deben existir elementos que posibiliten la modificación de las señales digitales mediante un procedimiento que se denomina modulación, a los efectos de que su transmisión sea posible sobre las redes analógicas.



1.12.1 Equipos terminales de datos (ETD)

Se pueden dividir en dos partes: fuente y colector de datos y el controlador de comunicaciones.

La parte que actúa como fuente o colector de datos es, en realidad, el equipo informático propiamente dicho, con aptitud de generar, recibir o procesar información en modo local. Estos, pueden cubrir una gama muy amplia de equipos informáticos, que pueden ir desde terminales no inteligentes hasta grandes ordenadores, pasando por equipos periféricos, como pueden ser unidades de cinta magnética, unidades de discos, lectoras de tarjetas o simples ordenadores personales o miniordenadores.

El controlador de comunicaciones, como su nombre lo indica, posee los elementos encargados de las funciones de comunicación, incluyendo en algunos casos facilidades de detección y corrección de errores, elementos que permitan controlar el diálogo y la interconexión de las interfases con la red.

1.12.2 Equipos terminales de circuitos de datos (ETCD)

Entre los equipos informáticos y las redes analógicas, deben conectarse equipos que actúan como interfases que se denominan “Equipos terminales de circuitos de datos”, comúnmente conocidos como equipos Modems de datos o simplemente Modems. La palabra Modem es una contracción de las palabras Modulador-de Modulador, funciones éstas principales en el trabajo de estos equipos.

Su misión básicamente es la de transformar las señales digitales producidas por el ETD, en señales analógicas capaces de ser transmitidas por la red y viceversa.

Adicionalmente, los equipos modems, realizan otras funciones como, por ejemplo, el establecimiento y liberación del circuito de comunicaciones.

1.13 Línea de comunicaciones

Es el conjunto de medios de transmisión que permite unir los dos equipos de terminación del circuito de datos – ETCD (Modems), cuya constitución dependerá de la distancia, velocidad, etc., y que debe cumplir unas determinadas especificaciones apoyándose en el infraestructura de comunicaciones permanente, pública o privada.

1.13.1 Enlace de datos

Es el circuito de unión entre el equipo ETD que actúa como fuente de datos y el ETD que actúa como colector de datos.

Tanto al iniciarse el circuito como al finalizar las señales que salen o llegan al ETD de este circuito son señales digitales, que pueden entender los equipos informáticos.

1.13.2 Circuito de datos

Es el camino formado por los modems y la línea de comunicaciones. Su misión es entregar en la interfase con el ETD (Colector), las señales bajo la misma forma y con idéntica información que recibió en al interfase con el ETD (Fuente).

1.14 Problemas en la transmisión

La transmisión de una señal supone el paso de la misma a través de un determinado medio, por ejemplo: un cable, el aire, etc. Debido a diferentes fenómenos físicos, la señal que llega al receptor difiere de la emitida por el transmisor.

Si la suma de todos los efectos no produce una gran diferencia entre ambas señales, conseguiremos una transmisión libre de errores. Por el contrario, cuando la señal recibida difiera en exceso de la señal transmitida el receptor puede interpretar incorrectamente la información y decimos entonces que se produce un error de transmisión. Evidentemente no todas las señales sufren los mismos efectos al atravesar los distintos medios de transmisión, luego cuando sea posible, escogeremos el tipo de señales y medios que conduzcan a las mejores condiciones de transmisión.

Veamos ahora algunos de estos problemas de la transmisión.

La atenuación
Consiste en el debilitamiento o pérdida de amplitud de la señal recibida frente a la transmitida. Por ejemplo, sabemos que cualquier sonido se percibe con menor intensidad cuando más alejados nos encontramos de la fuente que lo origina. Efectivamente, la atenuación tiene un efecto proporcional a la distancia. A partir de una determinada distancia, la señal recibida es tan débil que no se puede reconocer mensaje alguno.

Para paliar el efecto de la atenuación se pueden incorporar en el camino de la señal unos dispositivos activos, cuya función es amplificar la señal en la misma medida en que acaba de ser atenuada por el medio, de esta forma se consigue recuperar la señal para que pueda alcanzar más distancia.

Según el tipo de señal, analógica o digital, estos dispositivos tienen un comportamiento distinto y también diferente nombre. Para el caso de señales digitales hablamos de dispositivos repetidores, que son capaces de restaurar la misma señal original. Para las señales analógicas se denominan amplificadores y estos elementos no permiten recuperar la señal original, debido al efecto del ruido

que no se puede aislar de las señales analógicas pero sí de las digitales.

Distorsión por atenuación

Hasta ahora hemos supuesto que la atenuación afecta por igual a todas las señales. Sin embargo, la atenuación es función, además de la distancia, de la frecuencia de las señales que se propagan. Las de mayores frecuencias sufren una mayor atenuación.

Este fenómeno produce, en las señales con diferentes componentes frecuenciales, una atenuación distinta para cada componente de frecuencia, lo que origina que la señal recibida tenga una forma diferente de la transmitida, amen de una menor amplitud. Como la señal recibida se ha deformado con respecto a la transmitida decimos que se ha distorsionado.

Para compensar esta diferente atenuación a distintas frecuencias, los amplificadores pueden incorporar una etapa denominada ecualizador.

El retardo de grupo

Otro de los problemas de la transmisión es el retardo. Sabemos que todas las señales se propagan a una cierta velocidad, que depende del medio y de la naturaleza de la señal. Por ejemplo: el sonido se propaga en el aire aproximadamente a 340 m/s, la luz a 3000.000 km/s, etc.

Luego todas las señales van a tardar un cierto tiempo en recorrer la distancia que separa al emisor del receptor. Además, si en el camino la señal atraviesa determinados circuitos electrónicos, ópticos, o de cualquier otra naturaleza, estos pueden añadir un retardo adicional.

Este fenómeno carece de trascendencia en las transmisiones de voz, ya que el oído humano no es sensible a las diferencias de retardo. Sin embargo, tiene efectos importantes en la transmisión de datos digitales, especialmente a alta velocidad.

La diafonía

La diafonía (crosstalk) Es un fenómeno que todos hemos experimentado en las comunicaciones telefónicas. Consiste en la interferencia de un canal (o cable) próximo con el nuestro, esto produce una señal que es la suma de la señal transmitida y otra señal externa atenuada que aparece de fondo. En una conversación telefónica esto se observa como una segunda conversación que se oye de fondo mezclada con la nuestra.

El motivo de este fenómeno es la influencia mutua entre dos canales de transmisión próximos en frecuencia o que comparten el mismo tendido de cables.

El ruido impulsivo

Otra fuente de problemas en la transmisión es el denominado ruido impulsivo. Consiste en pulsos irregulares de corta duración y relativamente gran amplitud, que son provocados por inducciones, como consecuencia de conmutaciones

electromagnéticas. Este tipo de ruido es debido a causas variadas externas al medio de transmisión. Podemos asociarlo a las interferencias en un receptor de radio cuando se aproxima una motocicleta, o también al encendido de determinados aparatos en un domicilio.
Existen infinidad de dispositivos cuyo encendido o apagado genera un impulso de radio frecuencia capaz de influir a canales de comunicación próximos. El ruido impulsivo es típicamente aleatorio, es decir, se produce de manera inesperada y no suele ser repetitivo.

El ruido térmico
Está presente en todos los dispositivos electrónicos y medios de transmisión y es debido a la agitación de los electrones en un conductor. Es proporcional a la temperatura y se encuentra distribuido uniformemente en todo el espectro de frecuencias. Habitualmente el efecto del ruido térmico es despreciable, excepto en aquellos casos en los que se trabaja con señales muy débiles.

1.15 Técnicas de detección de errores

Anteriormente vimos que existen diversos problemas en la transmisión de señales que pueden ocasionar, errores en la transmisión, especialmente cuando la distancia que separa emisor y receptor es grande. Pueden tomarse medidas para reducirlos, pero no para eliminarlos completamente. Por ello, al realizar una transmisión deben utilizarse técnicas que permitan detectar y corregir los errores que se hayan producido. Estas técnicas se basan siempre en la idea de añadir cierta información redundante a la información que desee enviarse. A partir de ella el receptor puede determinar, de forma bastante fiable, si los bits recibidos corresponden realmente a los enviados.

Paridad
Uno de los métodos más comúnmente empleados para detectar errores, cuando el número de bits de información a transmitir es pequeño y la probabilidad de que ocurra un error es baja, es el uso de un bit adicional de paridad por elemento transmitido.
En el caso más sencillo se añadirá un bit a cada carácter a transmitir cuyo valor será dependiente de su peso (número de unos que contiene el carácter). Si se utiliza paridad par, se añadirá un uno si el carácter original tiene por peso un número impar y un cero en caso contrario. Para paridad impar, el proceso es el inverso. Con este método puede detectarse la ocurrencia de un número impar de errores pero los errores pares no serán detectados. Resulta adecuado, por ejemplo, cuando se utiliza transmisión serie asíncrona, cada carácter transmitido puede constar de siete y ocho bits de datos más uno de paridad.

CRC

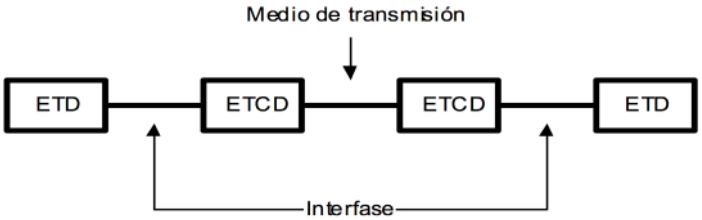
Los códigos de redundancia cíclica, también conocidos como códigos polinomiales constituyen el método de detección de errores más empleado en comunicaciones. Se utiliza con esquemas de transmisión orientados a tramas (o bloques). Permiten sustanciales mejoras en fiabilidad respecto a los métodos anteriores, siendo a la vez una técnica de fácil implementación.

El método se basa en el uso de aritmética polinomial módulo 2 (No hay acarreos en la sustracción ni en la adición y las operaciones suma, resta y OR–exclusivo coinciden). La trama a transmitir –de n bits – representa un polinomio de coeficientes binarios. La idea consiste en añadir una secuencia de k bits, al final de la trama, de manera que la secuencia de k+n bits resultante constituya los coeficientes de un polinomio divisible de forma exacta por un polinomio G(x) determinado previamente por emisor y receptor. Cuando el receptor recibe la trama realiza la división entre G(x), si el resto es distinto de cero ha ocurrido un error de transmisión.

CAPÍTULO II. MÓDEMS

2.1 Interfaz ETD/ETCD

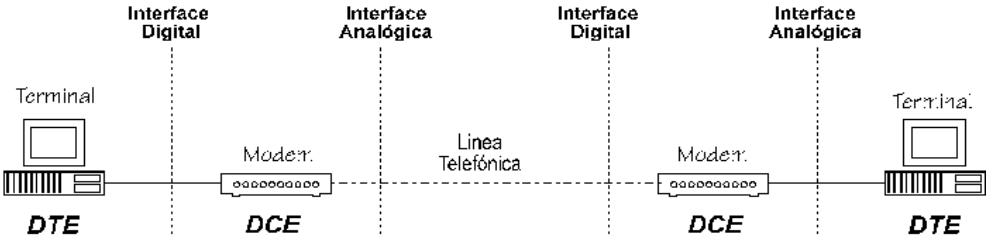
Existen unos dispositivos especializados que realizan la función de conversión de las señales que proporciona el ordenador a las señales que se pueden transmitir con facilidad a través del medio de transmisión escogido. Según la terminología empleada por el UIT-T estos dispositivos reciben el nombre de ETCD (Equipos Terminales de Circuito de Datos) y los ordenadores reciben la denominación de ETD (Equipos Terminales de Datos).



El tipo de ETCD más conocido es el módem, cuya función es precisamente convertir las señales digitales que suministra el ordenador en señales analógicas que puedan ser transmitidas con facilidad largas distancias.

Para facilitar la conexión entre ETD y ETCD se han desarrollado múltiples estándares que determinan todas las características eléctricas, mecánicas y funcionales de la conexión constituyendo lo que denominamos la definición de una interfase. Posiblemente el más conocido y popular es el “Recomended Standard 232”.

Los dispositivos ETD se asocian a dispositivos tales como computadoras, terminales, etc. En cambio los dispositivos ETCD se asocian a modems, impresoras, plotters, etc.



El conjunto de normas para asegurar la compatibilidad entre todo tipo de conexiones físicas ha sido definido por los organismos internacionales UIT e ISO en una serie de recomendaciones y estándares en los que se indican las interfaces para los enlaces de transmisión y los ETD. Estas recomendaciones regulan las características de origen mecánico, eléctrico y funcional para lograr la correcta comunicación entre el ETD y el ETCD.

Características mecánicas

Las características mecánicas son determinadas por el tipo de conector y el sexo (macho o hembra). En la interface digital, el conector de 25 pins es conocido comúnmente como el "conector EIA" (Electronics Industries Association). Actualmente, este es un nombre inapropiado ya que las características mecánicas fueron originalmente definidas por ISO 2110 (International Standar Organization). Es verdad que las características mecánicas son ahora definidas por el estándar EIA 232-D (la cual remplazó al estándar RS-232-C), pero el mismo conector es también especificado para la nueva interface de alta velocidad definida en EIA 530.

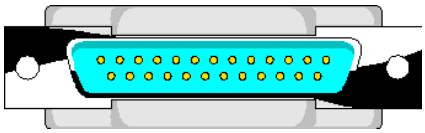


Figura 6: Conector RS232 DB-25.

Características eléctricas

Las características eléctricas son el voltaje, corriente e impedancia en el punto de interconexión. Las interfaces digitales pueden estar balanceadas o sin balancear, alimentadas con voltaje o alimentadas con corriente de acuerdo a sus especificaciones.

Características funcionales

En la interface digital, las descripciones de los circuitos de intercambio de datos, control, señal y tiempo definen las características funcionales, además de los pins del conector en los cuales estas aparecen.

INTERFACES MAS COMUNES

TIPO	PINS	ELÉCTRICAS	MECÁNICAS	FUNCIONALES
Velocidad media y baja, sin balancear	25	EIA RS-232-C	ISO 2110	EIA RS-232-C
	25	EIA-232-D	EIA-232-D	EIA-232-D
	25	CCITT V.28	ISO 2110	CCITT V.24
	25	EIA RS-336-A	ISO 2110	EIA RS-336-A
	37	EIA RS-423	ISO 4902	EIA RS-449
	15	CCITT V.10	ISO 4903	CCITT X.24
Velocidad alta, balanceado	25	EIA 530	EIA 530	EIA 530
	37	EIA RS-422	ISO 4902	EIA RS-449
	34	CCITT V.35	ISO 2593	CCITT V.35
	15	CCITT V.11	ISO 4903	CCITT X.24

NOTA: EIA 232-D remplazó al EIA-232-C. La interfase EIA-530 es un remplazo de 25 pins para el EIA RS-422/EIA RS-4449/ISO 4903

2.2 El Módem

Un módem es un dispositivo que sirve para modular y demodular (en amplitud, frecuencia, fase u otro sistema) una señal llamada portadora mediante otra señal de entrada llamada moduladora.

El modem es un periférico que con el tiempo se ha convertido ya en imprescindible y pocos son los modelos de ordenador que no estén conectados en red que no lo incorporen. Su gran utilización viene dada básicamente por dos motivos: Internet y el fax, aunque también le podemos dar otros usos como son su utilización como contestador automático incluso con funciones de central o para conectarnos con la red local de nuestra oficina o con la central de nuestra empresa.

Aún en el caso de estar conectado a una red, ésta tampoco se libra de éstos dispositivos, ya que en este caso será la propia red la que utilizará el modem para poder conectarse a otras redes o a Internet estando en este caso conectado a nuestro servidor o a un router.

Lo primero que hay que dejar claro es que los modem se utilizan con líneas analógicas, ya que su propio nombre indica su principal función, que es la de modular-demodular la señal digital proveniente de nuestro ordenador y convertirla a una forma de onda que sea asimilable por dicho tipo de líneas.

Es cierto que se suelen oír expresiones como modem ADSL o incluso modem RDSI, aunque esto no es cierto en estos casos, ya que estas líneas de tipo digital no necesitan de ningún tipo de conversión de digital a analógico, y su función en este caso es más parecida a la de una tarjeta de red que a la de un modem.

2.3 Modulación

Cuando las características del medio de transmisión no permiten señales digitales, se hace necesario convertir los datos digitales en una señal analógica susceptible de ser transmitida correctamente. El ejemplo más popular de esto, es la transmisión de datos digitales a través de las redes telefónicas. Estas redes fueron diseñadas para transmitir señales analógicas en el rango del espectro vocal (300-3400) Hz., lo que no resulta adecuado para la transmisión de señales digitales. Sin embargo, puede ligarse un dispositivo digital a la red telefónica por medio de un modem, que convierte los datos digitales en señales analógicas y viceversa.

En telecomunicación el término modulación engloba el conjunto de técnicas para transportar información sobre una onda portadora, típicamente una onda sinusoidal. Estas técnicas permiten un mejor aprovechamiento del canal de comunicación lo que posibilita transmitir más información en forma simultánea, protegiéndola de posibles interferencias y ruidos.

Básicamente, la modulación consiste en hacer que un parámetro de la onda portadora cambie de valor de acuerdo con las variaciones de la señal moduladora, que es la información que queremos transmitir.
Esta señal la transmitimos y el receptor debe realizar el proceso contrario, denominado demodulación para recuperar la información.

El proceso de modulación precisa de dos señales, moduladora y portadora, y genera como resultado una tercera señal, denominada señal modulada. Se define el término modulación como el proceso mediante el cual una señal que contiene información (moduladora) se combina con otra señal (portadora) para dar como resultado una nueva señal (modulada) que contiene la misma información que la primera pero que es el resultado de modificar alguno de los parámetros característicos (amplitud, frecuencia o fase) de la segunda.
Veamos ahora algunos esquemas simples de modulación:

- FSK (Modulación por desplazamiento de la frecuencia): Se modifica la frecuencia de la portadora según el valor de bit a transmitir, durante las celdas de bit con valor uno emplearemos la frecuencia f1 y la f para los bits con valor cero. La señal así resultante será una sucesión de tramas de las dos frecuencias indicadas, y será una señal analógica. Es el método más utilizado en modems de baja velocidad (300 a 1200 baudios) diseñados para operar con conexiones a través de la red telefónica conmutada. Presenta la ventaja de requerir circuitería simple para la demodulación y requisitos de ancho de banda bajos.
- ASK (modulación por desplazamiento de la amplitud): En esta técnica no se modifica la frecuencia de la portadora sino su amplitud. Los dos valores binarios se representan mediante diferentes niveles de amplitud de esta señal. Generalmente una de las amplitudes es cero, o sea el uno binario se representa por la presencia de portadora y el cero por la ausencia de la misma. Es una técnica bastante ineficiente de modulación debido a su mayor sensibilidad al ruido (que modifica al amplitud), y sobre líneas de voz suele utilizarse sólo hasta 1200 bps. Sin embargo, resulta adecuada para la transmisión de datos digitales sobre fibra óptica. En este caso, los bits de valor uno se representan mediante pulsos de luz y los bits a cero por la ausencia de luz.
- PSK (Modulación por desplazamiento de fase): La frecuencia y la amplitud se mantiene constantes y se varía la fase de la portadora para representar los niveles uno y cero con distintos ángulos de fase. Con este tipo de modulación el receptor debe mantener una señal portadora de referencia con la que comparar la fase de la señal recibida. Esto implica circuitos de demodulación complejos. Por este motivo, suele emplearse una forma alternativa de modulación en fase, que emplea desplazamientos en la fase relativos a la señal anterior transmitida. Por ejemplo, un desplazamiento de 90º relativos a la señal actual pueden servir para indicar un cero binario y un cambio den la fase de 270º un uno binario. De esta forma el circuito de demodulación solo necesita determinar los desplazamientos en la fase y no el valor absoluto (PSK diferencial).

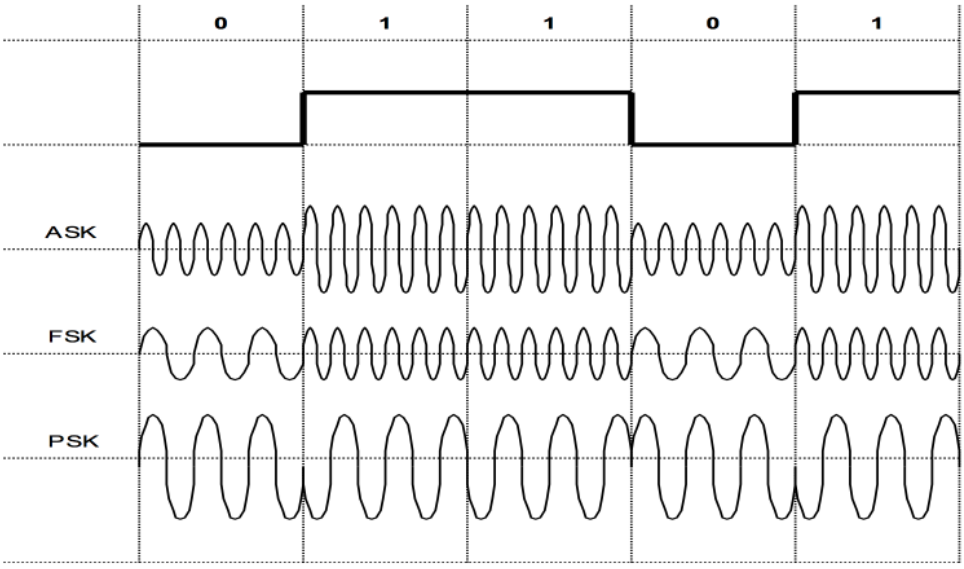


Figura 7: Diferentes métodos de modulación.

Pero en el canal telefónico también existen perturbaciones que el módem debe enfrentar para poder transmitir la información. Estos trastornos se pueden enumerar en: distorsiones, deformaciones y ecos. Ruidos aleatorios e impulsivos. Y por último las interferencias.
Los métodos de modulación y otras características de los módems telefónicos están estandarizados por el UIT-T (el antiguo CCITT) en la serie de Recomendaciones "V". Estas Recomendaciones también determinan la velocidad de transmisión. Destacan:

- V.32. Transmisión a 9.600 bps.
- V.32 bis. Transmisión a 14.400 bps.
- V.34. Transmisión a 33.600 bps. Uso de técnicas de compresión de datos.
- V.90. Transmisión a 56'6 kbps de descarga y hasta 33.600 bps de subida.
- V.92. Mejora sobre V.90 con compresión de datos y llamada en espera. La velocidad de subida se incrementa, pero sigue sin igualar a la de descarga.

Uno de los primeros parámetros que lo definen es su velocidad. El estándar más habitual y el más moderno está basado en la actual norma V.90 cuya velocidad máxima está en los 56 Kbps (Kilobites por segundo). Esta norma se caracteriza por un funcionamiento asimétrico, puesto que la mayor velocidad sólo es alcanzable "en bajada", ya que en el envío de datos está limitada a 33,6 Kbps. Otra consideración importante es que para poder llegar a esta velocidad máxima se deben dar una serie de circunstancias que no siempre están presentes y que dependen totalmente de la compañía telefónica que nos presta sus servicios, pudiendo ser en algunos casos bastante inferiores.
Evidentemente, el modem que se encuentre al otro lado de la línea telefónica, sea

nuestro proveedor de Internet o el de nuestra oficina debe ser capaz de trabajar a la misma velocidad y con la misma norma que el nuestro, ya que sino la velocidad que se establecerá será la máxima que aquel soporte.

Otras normas habitualmente utilizadas son

Norma	Velocidad máxima	Otras velocidades
V.90	56.000 bps	57.333, 54.666, 53.333, 52.000, 50.666, 49.333, 48.000, 46.666, 45.333, 44.000, 42.666, 41.333, 40.000, 38.666, 37.333, 36.000, 34.666 bps
V.34+	33.600 bps	31.200 bps
V.34	28.800 bps	26.400, 24.000, 21.600, 19.200, 16.800 bps
V.32bis	14.400 bps	12.000 bps
V.32	9.600 bps	7.200 bps
V.23	4.800 bps	
V.22bis	2.400 bps	
V.22 y Bell 212ª	1.200 bps	
V.21 y Bell 103	300 bps	

Otros estándares considerados como imprescindibles son los de control de errores y compresión de datos. Los más habituales son: V.42, V.42bis y MNP 2-5.

2.4 Tipos de modems

2.4.1 El modem serie externo

Figura 8: Modem serie externo.

Este es el modem “clásico” por antonomasia y posiblemente aún el más utilizado, a pesar de que la competencia de los modelos basados en USB es cada vez más fuerte. Por tanto, los mejores modelos se suelen encontrar aún en este formato y es ya habitual encontrarse en ellos funciones de contestador automático, fax y central telefónica, actuando incluso en el caso de que nuestro ordenador esté apagado, gracias a la memoria que incorporan. Algunos modelos también integran un altavoz y un micrófono, por lo que se convierten en plenamente autónomos...

Hay que tener en cuenta que la velocidad de comunicación del modem con el puerto serie debe ser bastante mayor de la que éste es capaz de transmitir a través de la línea telefónica, entre otros motivos por la compresión hardware que es capaz de realizar a los datos que le llegan.

Ventajas

- No ocupan ninguna ranura de expansión, lo que es adecuado para ordenadores con nulas o pocas posibilidades de ampliación.
- Sólo utilizan los recursos del propio puerto serie al que están conectados.
- Disponen de indicadores luminosos que nos informan del estado de la conexión y del propio modem.
- Se pueden "reiniciar" sin necesidad de hacerle un "reset" al ordenador o simplemente apagar cuando no lo utilizamos.
- Por último, algunos modelos externos implementan botoncitos adicionales para subir o bajar el volumen del altavoz o para activar las funciones de contestador o incluso implementan un micrófono o un altavoz, que en los modelos internos difícilmente podremos encontrar.

Inconvenientes

- En máquinas muy antiguas nos podemos encontrar con que la velocidad de transferencia del puerto serie limitará la velocidad del modem, por lo que será necesario sustituir nuestra tarjeta I/O por una más moderna.
- Necesitan de una fuente de alimentación externa y ocupan lugar en nuestro escritorio.
- Necesitan un puerto serie libre.

2.4.2 El modem interno

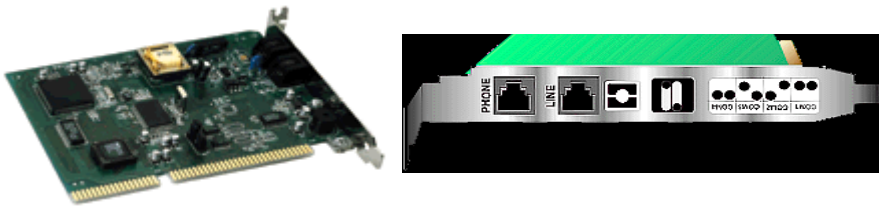


Figura 9: Modem interno.

- En este tipo de configuración normalmente encontramos modelos de gama baja y prestaciones recortadas, como ocurre en el caso de los "Winmodem", también llamados "softmodem" o HSP. Sin embargo esto no es más que una estrategia de los fabricantes debido a que este tipo de modem suelen resultar más económicos que los externos.

- Aquí igualmente podremos hacer una segunda distinción dependiendo del tipo de bus al que vayan conectados. Encontraremos modelos para ranura ISA o para PCI. Debido a que el primero está tendiendo a desaparecer, cada vez es más difícil encontrar modelos para él, siendo lo habitual los dispositivos PCI, que además tienen la ventaja del Plug and Play (PnP) que siempre es una ayuda en el momento de su instalación.
- La mayor parte de la funcionalidad del dispositivo está implementada en la propia placa base y al igual que ocurre en el caso de los Winmodem su funcionamiento está más basado en el software que en el hardware, lo que repercute en un menor precio de coste pero por el contrario su utilización consume ciclos de CPU y su portabilidad está limitada ya que no todos los sistemas operativos disponen del soporte software adecuado para hacerlos funcionar.

Ventajas

- No necesitan una fuente de alimentación externa y no ocupan lugar en nuestro escritorio, lo que normalmente es de agradecer...
- No ocupan ninguno de los puertos serie existentes en nuestra máquina.
- En máquinas muy antiguas no hay que preocuparse de posibles problemas en la velocidad de transferencia por causa de un puerto serie lento debido a la utilización de algún chip anticuado.

Inconvenientes

- Ocupan una ranura de expansión, lo que puede ser contraproducente cuando disponemos de pocas en el interior de nuestra máquina.
- Utilizan recursos que muchas veces son preciosos si el número de dispositivos que tenemos instalado es muy elevado. Esto también puede llevar a problemas de conflictos que nos pueden producir más de un dolor de cabeza aún en el caso de dispositivos PnP.
- No tienen “lucecitas” que nos informen del estado de la conexión y del propio modem. Tampoco se pueden “reiniciar” cuando tenemos algún problema con él, aunque esto último suele ser hoy día bastante raro...
- Por último, algunos modelos externos implementan botoncitos adicionales para subir o bajar el volumen del altavoz o para activar las funciones de contestador o incluso implementan un micrófono o un altavoz, que en los modelos internos difícilmente podremos encontrar.

2.4.3 El modem USB

Este tipo de configuración es la más reciente dentro del mundo de los modem. La principal ventaja la tenemos en el propio método de conexión.

Respecto del modelo externo para puerto serie tiene la ventaja de que no hay que preocuparse por la velocidad de conexión de éste con el ordenador, pues en este caso el caudal proporcionado es más que suficiente. Tampoco es problema el contar con pocos puertos USB, pues siempre podremos adquirir un hub para interconectar más dispositivos. De todas formas para evitar este gasto sería interesante que el propio modem incorporara como mínimo dos conectores, aunque no suele ser lo habitual.

Ventajas

- No ocupan ninguna ranura de expansión, lo que es adecuado para ordenadores con nulas o pocas posibilidades de ampliación, incluso para ordenadores portátiles, aunque hay que tener en cuenta que su consumo normalmente será mayor que el de un dispositivo de tipo PC-Card.
- Sólo utilizan los recursos del propio USB al que están conectados.
- Suelen dispone de indicadores luminosos que nos informan del estado de la conexión y del propio aparato.
- Algunos modelos disponen de un interruptor para apagarlo cuando no lo utilizamos. En todo caso, al igual que ocurre con cualquier otro dispositivo USB, siempre se puede desconectar (y por supuesto conectar) “en caliente”, es decir, con el ordenador en marcha.
- Una ventaja sobre los modem externos serie es que no precisan de ninguna alimentación externa.

Inconvenientes

- Ocupan lugar en nuestro escritorio.
- Necesitan un conector USB libre, bien en el propio ordenador, bien en algún otro dispositivo USB conectado a nuestra máquina que actúe de hub.

2.4.4 El modem en formato PC Card

Este tipo de modem es el adecuado para los ordenadores portátiles, pues tiene las mismas prestaciones que el resto de tipos analizados, pero con el tamaño de una tarjeta de crédito.

Ventajas

- No necesita fuente de alimentación externa y su consumo eléctrico es reducido, aunque no es conveniente abusar de él cuando lo utilizamos en un ordenador portátil usando las baterías.

Inconvenientes

- Requieren una ranura PC-Card libre, normalmente de tipo I (las más estrechas)

El RS – 232D

En agosto de 1969, la asociación de Industrias de Electrónica (EIA, Electronic Industries Association) estableció el conector estándar para comunicación serial EIA RS-232C encargado de controlar el intercambio de datos en forma serial entre el equipo de comunicaciones de datos y el equipo terminal a través de líneas telefónicas.

La última revisión del estándar RS-232, es conocida como RS-232D o, técnicamente EIA 232-D (UIT V.24).

El conector RS-232 tiene dos modalidades conocidas comúnmente como DB25 y DB9, que corresponden a conectores de 25 pins y 9 pins respectivamente. En algunas ocasiones se añaden las letras P o S. por ejemplo: DB25S, DB25P, DB9S y DB9P. La letra P (de Plug) se les asocia a los conectores macho. La letra S (de Socket) para los conectores Hembra.

El RS-232 es una norma para la conexión entre un ETD y un ETCD que define:

- El tipo de conector a emplear
- Las características eléctricas
- Los niveles de tensión
- Las longitudes máximas a distintas velocidades
- Los nombres de las señales que intervienen en el funcionamiento y la estructura del protocolo de comunicación

Esta norma establece una señalización eléctrica bipolar, según la siguiente tabla:

Nivel lógico 0	+15..+3 Voltios
Nivel lógico 1	-15..-3 Voltios

Las velocidades de transmisión que puede soportar este estándar van desde los 0bps hasta los 20Kbps. Con respecto a las distancias máximas se propone que no sean superiores a 15 metros. Aunque un diseño cuidadoso puede permitir distancias muy superiores, hay que suponer que esta limitación teórica se puede manifestar en la práctica en dispositivos que cumplan la norma.

SEÑALES DEL CONECTOR RS-232

El conector RS-232D (DB25) está dividido en cuatro grupos distintos de señales:

- Señales de Datos.
- Señales de Control.
- Señales de temporización.
- Señales de Tierra.

SEÑALES DE DATOS

- TXD (Transmit Data): Los datos son enviados desde la computadora al módem por este pin para transmisión.
- RXD (Received Data): Los datos son recibidos por la computadora.

SEÑALES DE CONTROL

- DTR (Data Terminal Ready): La computadora usa DTR para decirle al módem que está encendida, el software esta cargado, y esta lista para comunicarse.
- DSR (Data Set Ready) o MR (módem Ready): El módem local le dice a la computadora que esta encendido en modo normal de operación.

- DCD (Data Carrier Detect) o CD (Carrier Detect): Una indicación de que el módem remoto esta en línea y esta listo para intercambiar datos. (la presencia de una portadora no implica que la portadora tiene datos). Esta señal es proveída a la computadora desde el módem local.
- RI (Ring Indicator) : El módem detecta timbrado en la línea.
- RTS o RS (Request to Send): La computadora pregunta al módem si esta listo para empezar a transmitir datos.
- CTS o CS (Clear To Send): El módem le dice a la computadora que continúe y empiece a enviar datos.

Mnemónico	Nombre en Inglés	E/S	Descripción
DTR	Data Terminal Ready	S	El ETD indica que está preparado
DSR	Data Set Ready	E	El ETCD indica que está preparado
RTS	Request To Send	S	Solicita permiso para transmitir
CTS	Clear To Send	E	El ETCD autoriza la transmisión solicitada
TXD	Transmitted Data	S	Línea de transmisión de datos serie
RXD	Received Data	E	Línea de recepción de datos serie
GND	Ground	-	Masa de referencia 0v
RI	Ring Indicator	E	Detección de llamada
DCD	Data Carrier Detect	E	Detección de portadora

SEÑALES DE TEMPORARIZACIÓN

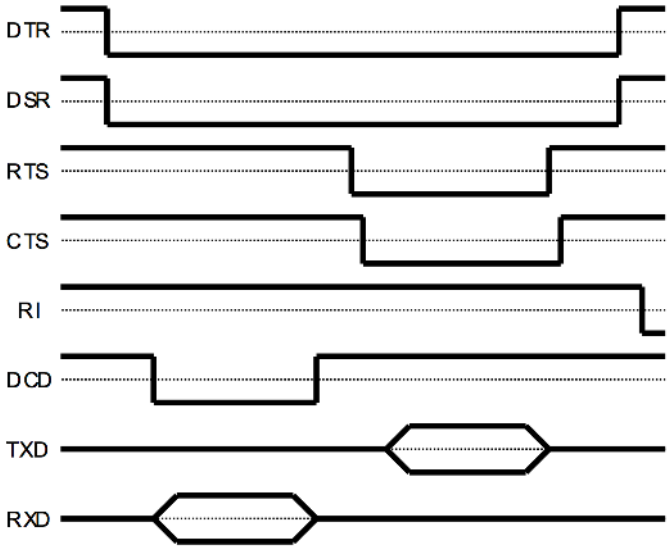
- Estas señales pins 15,17 y 24 comúnmente llamadas “señales de reloj” son usadas en comunicaciones síncronas para permitir a los dispositivos transmisores y receptores se sincronicen. Un lado debe de llevar el tiempo y el otro debe ser esclavo.

SEÑALES DE TIERRA

- PG (Protective Ground): Usada para protección eléctrica.
- SG (Signal Ground): Usada como señal de referencia para las demás señales.

Seguidamente vamos a comentar el funcionamiento conjunto de las señales

descritas en su uso normal, es decir, en una comunicación entre ordenador (ETD) y módem (ETCD).
La siguiente figura detalla la evolución de señales que hemos relacionado anteriormente. No debe olvidarse que estamos viendo sólo un subconjunto de todas las señales que se definen en la norma.



Las señales DTR y DSR son un requisito previo a la comunicación a través de esta interfaz, puesto que informan a cada dispositivo - ETD y ETCD - del estado del otro, indicando si están o no preparados para funcionar.

La señal DCD indica si se detecta portadora en el canal de transmisión. La detección de portadora es un requisito previo a la recepción, debe existir portadora antes y durante la recepción. Cuando desaparece la portadora, sabemos que no se van a recibir más datos.

En canales semi-duplex habrá que esperar a que finalice la recepción antes de poder iniciar una transmisión, lo que no será necesario si el canal es dúplex. Para iniciar una transmisión, el ordenador activará la línea RTS y esperará a que le autorice el módem por medio de la línea CTS. En cuanto esta se active, el ordenador podrá proceder a transmitir los datos.

Ocasionalmente, y siempre en una situación de inactividad de los equipos (estado de reposo), se puede producir la activación de la señal RI, que indica que el módem está recibiendo una llamada por la red. Esta situación podría servir para iniciar el establecimiento de una nueva comunicación.

Solamente 9 de los 25 pins son comúnmente usados en aplicaciones asíncronas (1,2,3,4,5,6,7,8,20). Con aplicaciones síncronas, solo dos más son normalmente usados (15 y 17). Algunas interfaces que utilizan 9 ó más pins, sólo requieren 3 cables entre sus conectores; transmisión, recepción y señal de tierra (2,3 y 7), por lo general en conexiones directas.

Consecuentemente, no hay necesidad de que las señales de control coordinen la compartición de una sola línea.

Las funciones de los pins de un DTE no son las mismas que un DCE (estas están cruzadas), por ejemplo: Una solicitud de envío (RTS) de un DTE es una señal de control de salida que solicita permiso para enviar datos. Por otra parte un RTS de un DCE es una señal de control de entrada donde otro dispositivo RS-232D está solicitando permiso para enviar datos, por lo tanto:

- Un DTE es un dispositivo que transmite datos por el pin 2.
- Un DCE es un dispositivo que transmite datos por el pin 3.

La interface entre la computadora y el módem usan señales lógicas que representan las señales de control. Típicamente, cuando una señal esta baja (tierra o voltaje negativo), la señal está en estado apagado (OFF). Si el voltaje es positivo, se dice que la señal está en estado encendido (ON).

Es importante conocer que el módem debe necesitar la señal DTR de la computadora para permitir el intercambio de información; y que el programa de comunicaciones necesita la señal DSR del módem antes de empezar a intercambiar datos.

Además, solo después de que el módem responde con CTS con una respuesta RTS de la computadora, puede la computadora intercambiar datos. Cuando RTS está arriba, existe portadora. El transmisor manda todos los datos (1s y 0s) hasta que CTS se enciende y en ese momento los datos empiezan a fluir.

SEÑALES DEL MODEM

Estas señales (luces indicadoras) se encuentran en el frente de la mayoría de los modems externos y nos dicen que está sucediendo durante la comunicación. Las posiciones exactas de las luces están en orden y pueden variar dependiendo del módem. Todas las señales están abreviadas a dos caracteres.

- HS (High speed): Indica que el módem está actualmente operando en su más alta razón de transmisión.
- AA (Auto Answer): Indica que el módem responderá automáticamente todas las llamadas que lleguen.
- CD (Carrier Detect): Este luz se prende cada vez que el módem detecta portadora, la cual significa que la conexión con la computadora remota ha sido hecha exitosamente. La luz se deberá apagar cuando una de las

computadoras descuelga la línea y la señal de la portadora termina.

• OH (Off-Hook): Esta luz se enciende cada vez que el módem toma el control de la línea telefónica. Esto es equivalente al descolgar el auricular del teléfono.

• RD (Receive Data): Esta luz parpadea cada vez que el módem transfiere datos a la computadora. Esto sucede cada vez que se están recibiendo datos de la computadora remota.

• SD (Send Data): Esta luz parpadea cada vez que la computadora transfiere datos al módem, siempre que se manden datos a la computadora remota.

• TR (Terminal Ready): Esta luz se enciende cuando el módem detecta una señal DTR desde el programa de comunicaciones. Esta señal informa al módem que un programa de comunicaciones está cargado y listo para correr.

• MR (módem Ready): Esta luz avisa que el módem esta encendido y listo para operar.

PARÁMETROS DE COMUNICACIONES

Antes de probar la interfaz RS-232 se deben verificar los siguientes requisitos para una comunicación serial apropiada.

- Baud Rate: La velocidad de ambos dispositivos debe ser la misma.
- Paridad: Los formatos de verificación de error deben de ser idénticos en ambos dispositivos; las opciones son: sin paridad, con paridad par; con paridad impar. Por lo general se omite la paridad
- Tamaño de Palabra: Número de bits por caracter (varían de 5 a 8 bits), ambos dispositivos deben de usar el mismo número de bits por caracter.
- Número de Bits de paro: Es el número de bits (pueden ser 1o 2 bits de paro).
- Modo: se utilizan los modos de transmisión “Simplex” Transmisión (no muy usada), “Half-Duplex” y “Full-Duplex”. Ambos dispositivos deben de tener el mismo modo.
- Código de Temporización: La mayoría de las computadoras usan temporización asíncrona (derivada de los bits de inicio y de paro de un caracter) pero también existen interfaces síncronas. Aunque ambos dispositivos deben de utilizar la misma forma de transmisión de la temporización asíncrona algunas veces el número de bits de paro puede ser diferente para los dos dispositivos. Se deberá de asegurar de poner el mismo número de bits de paro en ambos extremos las opciones son generalmente 1 o 2. Dependiendo si el módem es temporizado internamente o de forma externa, los dispositivos síncronos requerirán que tome en cuenta pins de temporización.

2.5 Órdenes AT

La mayoría de los modems se controlan y responden a caracteres enviados a través del puerto serie. El lenguaje de comandos para modems más extendido es el de los comandos Hayes que fue inicialmente incorporado a los modems de este fabricante.

Órdenes de comunicación
Los comandos comienzan con las letras AT y siguen con las letras del alfabeto (A...Z).

- ATA: con esta orden el módem queda en espera de una llamada telefónica, comportándose como un receptor (autoanswer).

Cada módem utiliza una serie de órdenes "AT" comunes y otras específicas. Por ello, se deberá hacer uso de los manuales que acompañan al módem para configurarlo adecuadamente.

Registros
Los registros o registros S son porciones de memoria donde se pueden guardar permanentemente parámetros que definen el perfil del módem (profiles). Además de las órdenes "AT", se dispone de esta serie de registros que permiten al usuario la modificación de otras características de su funcionamiento. Al igual que ocurre con las órdenes "AT", existen registros comunes y otros específicos del módem. Para programarlos se usa el comando ATSNúmero del registro=Valor a establecer Se enumeraran los más comunes.

Registro 0: número de llamadas que el módem espera antes de responder (autoanswer). Si su valor es 0, el módem nunca responderá a las llamadas.

Registro 1: contabilizador de llamadas realizadas / recibidas.

Registro 2: código del carácter que se utiliza para activar la secuencia de escape. Suele ser un +.

Registro 3: código del carácter de fin de línea. Suele ser un 13 (enter).

Registro 4: código de carácter de avance de línea, (line feed).

Registro 5: código de carácter de borrado con retroceso (backspace).

Registro 6: tiempo de espera antes de empezar a marcar (s).

Registro 7: tiempo de espera para recibir portadora (s).

Registro 8: tiempo asignado a la pausa del Hayes (la coma, en s).

Registro 9: tiempo de respuesta a la detección de portadora, para activar la DCD (en décimas de segundo).

Registro 10: tiempo máximo de pérdida de portadora para cortar la línea. Aumentando su valor permite al remoto cortar temporalmente la conexión sin que el módem local inicie la secuencia de desconexión. Si es 255, se asume que siempre hay portadora. Este tiempo debe ser mayor que el del registro 9 (en décimas de segundo).

Registro 12: determina el guard time; éste es el tiempo mínimo que precede y sigue a un código de escape (+++), sin que se hayan transmitido o recibido datos. Si es 0, no hay límite de tiempo (S12 x 20 ms).

Registro 18: contiene la duración de los tests.

Registro 25: tiempo para que el módem considere que la señal de DTR ha cambiado.

Registro 26: tiempo de respuesta de la señal CTS ante RTS.

CAPÍTULO III: TEORÍA DE LA INFORMACIÓN

Códigos de representación de la información

La información es transmitida en forma de mensajes.
Por mensaje se entiende el conjunto de símbolos que contienen cierta información.
El conjunto total de símbolos se conoce como alfabeto.

Para poder transmitir un mensaje por un canal de comunicación es en ocasiones necesario codificarlo. La codificación se utiliza ampliamente durante la transmisión, conservación y procesamiento de la información y es necesaria para lograr la mejor adaptación de la fuente de información con el canal de comunicación.

3.1 Introducción a la Teoría de la Información

Para entender el estudio de la codificación es necesario algunas nociones de la teoría de la información por lo que le dedicaremos algún tiempo al estudio de estas teorías.
En 1949 Claude E. Shannon publicó el trabajo titulado “Una teoría matemática de la comunicación”, que sentó las bases sobre la teoría de la información, donde trataba la incertidumbre o las cantidades de desorden en un sistema.

A la teoría de la información, en su planteamiento clásico, pertenecen los resultados del estudio de problemas teóricos relacionados con el aumento de la eficiencia del funcionamiento de los sistemas de comunicación. Esto se expresa en la estimación de la cantidad de información transmitida en los mensajes, en el análisis de las características de información de las fuentes de mensajes y de los canales de comunicación y de la posibilidad de codificar mensajes de tal manera que se aumente la velocidad de transmisión, aún en presencia de perturbaciones.

Esta teoría establece cuántos “bits” de información pueden enviarse en una unidad de tiempo a través de canales de comunicación perfectos o imperfectos, usando las propiedades abstractas de estos canales. Además como medir la capacidad de generación de información de una fuente de mensajes, como codificar los mensajes procedentes de una fuente de manera tal que resulte eficiente para una transmisión sobre un tipo de canal y como detectar y corregir errores en la transmisión.

El bit: una medida universal de la cantidad de información

Otro aspecto a señalar es la relación entre información y la ganancia de conocimiento. Por ejemplo, la afirmación “en Polo Norte hace frío” aporta un conocimiento muy escaso y, en consecuencia, se dice que la información contenida es mínima, ya que actualmente no se puede aprender nada de ahí por ser un fenómeno totalmente determinado. Por tanto, sólo puede haber información relativa a un elemento variable. Este elemento variable se puede presentar en diferentes estados.

Por último toda fuente de información se apoya en un lenguaje escrito (o hablado), por lo que la información se refiere a datos que se pueden escribir.

En resumen, la información debe:

- Poder suministrar conocimiento
- Referirse a elementos variables que se presentan en diferentes estados
- Poder ser escrita

Algunas definiciones

3.2 Cantidad de información: Capacidad de un mensaje de despejar incertidumbre

La cantidad de información juega un papel importante en problemas de transmisión, codificación, etc. Esta cantidad está relacionada con el número de estados posibles y con la posibilidad de que ocurra uno de ellos, de modo que la cantidad de información que se obtiene al recibir un símbolo(o una sucesión de símbolos) está en relación inversa con la probabilidad de ocurrencia del evento representado por dicho símbolo. Así, la cantidad de información en un evento E se expresa:

I(E)=log2 1/P(E) I(mi) = log a 1/P(mi)

Shannon(Norma ISO-2382/XVI) o bit

En un sistema de Comunicaciones binario P(E)=½ . Por tanto, la cantidad de información obtenida al especificar una de las 2 alternativas igualmente probables es 1 Shannon (o 1 bit, que aquí no tiene el sentido de dígito binario).

Considérese el código ASCII, de 7 elementos binarios, y que se recibe una señal binaria que contiene información en ese código. Cada bloque de 7 dígitos binarios recibidos representa un símbolo, que puede ser una letra, un número, etc. Según la definición la cantidad de información que se obtiene al recibir el carácter O, por ejemplo, no es la misma que se obtiene al recibir el carácter W(es decir, de 7 Shannon o 7 bits), ya que las probabilidades de recibir uno u otro es diferente, aunque ambos se representen por 7 bits.

Sólo en el caso en que cada uno de los 27=128 caracteres posibles a codificar sean

equiprobables en cada instante, la cantidad de información será 7 Shannon:

I=log2 1/(1/27)=log2 27= 7 Shannon

En general en sistemas digitales, si cada símbolo o carácter está representado por N bits, y todos los símbolos posibles son equiprobables, se cumplirá que la cantidad de información es:

I=log2 1/(1/2N)=log2 2N = N Shannon/símbolo

P(A) = n/N
P(A): probabilidad del evento A: relación entre la cantidad de resultados favorables al evento A, y la cantidad total de resultados posibles.
Propiedades:

1. $I(E) \geq 0$
2. Si P(E) = 1, entonces I(E)=0
3. Si P(E1) > P(E2), entonces I(E1) < I(E2)
4. Si P(E1,E2)= P(E1).P(E2), entonces I(E1E2)=I(E1)+I(E2) (eventos independientes).

Para comprender los elementos básicos de la teoría de la información es importante tener a disposición una descripción matemáticas de los mecanismos generadores de información (fuentes de información).

Las fuentes de información son dispositivos capaces de generar mensajes**.**

3.3 Fuentes discretas sin memoria

Las fuentes que generan una cantidad limitada de mensajes se llaman discretas. Un mensaje de una fuente discreta es una secuencia de símbolos y los diferentes mensajes se diferencian entre sí por el orden en que se suceden esos símbolos:

s1,s2,...,si,... 1 ≤ i ≤ n

La aparición de los diferentes símbolos de tal fuente sigue una ley probabilística dada. En el caso en que los símbolos de la fuente sean emitidos en forma independiente, o sea

P(si/si-1,si-2,...,s1) = P(si) para toda i,
Se dice que la fuente es discreta sin memoria. Esto quiere decir que la aparición de

un símbolo no depende de otros anteriores. ¿Es el idioma español una fuente de información discreta? ¿Es una fuente discreta sin memoria?

Si P(si) es la probabilidad de ocurrencia del símbolo si de la fuente, se cumple que

$$\sum_{i=1}^{n} P(si) = 1$$

Entropía

Para caracterizar una fuente información en cuanto a su contenido no resulta práctico referirse a cada símbolo emitido. Es más conveniente usar el contenido medio de información por símbolos emitidos cuando se ha realizado un número elevado de pruebas. Este valor promedio de información se conoce como entropía (H).

Entropía: Cantidad promedio de información por símbolo que provee la fuente.

Considérese una fuente de información que emite los símbolos s1 y s2 (binaria) con probabilidades de ocurrencia P(s1) y P(s2). Se cumple que:

$$P(s1) + P(s2) = 1 \text{ y} \qquad I(s1)=\log_2 1/P(s1), \qquad I(s2)=\log_2 1/P(s2)$$

Si K y L son las veces que aparecen los símbolos s1 y s2 en una secuencia larga de N símbolos, entonces la probabilidad de ocurrencia de cada uno de ellos serán:

$$P(s1)= K/N \qquad \text{y} \qquad P(s2)=L/N$$

y la información de la fuente (la información de ambos símbolos) será:

$$I(s)= KI(s1) + LI(s2)$$

Sustituyendo en la última ecuación se tiene:

$$I(s)=N.P(s1).\log_2 1/P(s1) + N.P(s2).\log_2 1/P(s2)$$

Como la entropía H(s) es el contenido medio de la información por símbolos, entonces

$$H(s) = I/N = P(s1)\log_2 1/ P(s1) + P(s2)\log_2 1/ P(s2) \qquad \text{Shannon/símbolo}$$

De modo general si una fuente s emite n símbolos de un alfabeto, entonces la entropía de la fuente s es:

$$\sum_{n}$$

$$H(s) = -\sum_{i=1} P(si) \log_2 P(si) \quad \text{Shannon/símbolo}$$

General

$$H(s) = -\sum_{i=1}^{n} P(si) \log_2 P(si) \quad \text{Shannon/símbolo}$$

Las bases de los algoritmos más utilizados en la teoría de la información son 2, n, 10. En correspondencia con esto las unidades en que se expresa la entropía son: bits. nit, hartley.

$$H(s) = -\sum_{i=1}^{n} P(si) \log_2 P(si) \quad \text{bit/símbolo}$$

$$H(s) = -\sum_{i=1}^{n} P(si) \ln P(si) \quad \text{nat/símbolo}$$

$$H(s) = -\sum_{i=1}^{n} P(si) \log_{10} P(si) \quad \text{hartley/símbolo}$$

$$H(s) = -\sum_{i=1}^{n} P(si) \log_2 P(si) \quad \text{Shannon/símbolo}$$

La expresión anterior coincide en estructura con la entropía de un sistema físico. En ambos casos la magnitud H caracteriza el grado de diversidad de los estados del sistema. Obsérvese que el grado de indeterminación del estado de una fuente de información depende del número de estados (símbolos) y de la probabilidad de ocurrencia de ellos. Cuando las probabilidades de los símbolos son diferentes se reduce la libertad de elección de esos símbolos.

Ejemplo

Si una fuente de información tiene, 2 símbolos cuyas probabilidades son 0,99 y 0,01, la indeterminación de la elección del símbolo será mucho menor que en la fuente cuyos símbolos son equiprobables.

Generalizando: La información es máxima si los eventos son equiprobables

Cuando todos los símbolos tiene la misma probabilidad de ocurrencia, es decir,

$$P(s1)=P(s2)=...=P(sn)=1/n$$

entonces:

$$H(s)= - \sum(1/n)\log2(1/n)= - n(1/n)\log2(1/n)= - \log2(1/n)= \log2\ n$$

y da el valor máximo de la entropía.

Se concluye que la entropía de una fuente es máxima, cuando la distribución de probabilidades de sus símbolos es uniforme y numéricamente es igual al logaritmo de la cantidad de mensajes o símbolos de la fuente.

Ejemplos

Considerando como fuente el alfabeto ruso y suponiendo que las 32 letras de ese alfabeto aparecen con igual probabilidad en un texto usual, entonces se llega a que la entropía de esa fuente es log2 32 = 5 Shannon/símbolo.

Si se considera análogamente alfabeto inglés, con el símbolo espacio, bajo las mismas condiciones, entonces la entropía de esa fuente es log2 27 = 4,76 Shannon/ símbolo.

Sin embargo, se sabe que los símbolos de esos alfabetos (y en cualquier alfabeto) no aparecen con igual probabilidad. Atendiendo a las probabilidades reales, se tiene, para el alfabeto ruso, que la entropía de la fuente es H(s)=4,42 Shannon/símbolo y, en el caso del alfabeto inglés, la entropía de la fuente es H(s)=1,3 Shannon/símbolo. Esto significa que el idioma inglés es redundante en un 70% y que sería posible reconstruir correctamente un texto en inglés si alguna de las letras se perdiera o cambiara a causa del ruido o a la distorsión.

Evidentemente, hay situaciones en transmisión de datos en que es deseable que haya redundancia con el fin de aumentar la probabilidad de recibir un mensaje adecuadamente en un medio ruidoso.

Propiedades de la Entropía

1. $H(s) \geq 0$ La entropía es siempre positiva, ya que los valores de las probabilidades se expresan mediante magnitudes menores que la unidad, por lo tanto sus logaritmos son negativos, de forma que los términos son siempre positivos. Por otra parte no es interpretable desde el punto de vista técnico, que la cantidad de información. Además, es una magnitud acotada.
2. $H(s) = 0$ en el caso de que exista un mensaje con probabilidad de ocurrencia 1 y el resto 0. Esta situación corresponde al caso en que el estado de la fuente está completamente determinado.
3. H(s)tiene su valor máximo en el caso de que las probabilidades de ocurrencia

de los mensajes sean iguales entre sí.

Hmax(s) = log2 n = CI llamada capacidad de símbolo de la fuente

4. La **entropía es una** magnitud acotada

3.4 Eficiencia de un código

La codificación de la fuente que vimos en el esquema de un sistema de Transmisión de Datos es un proceso encaminado a lograr una representación más eficiente de la información.

Como ya vimos codificar es establecer correspondencia entre los símbolos de la fuente y los símbolos del alfabeto de un código.

Como el alfabeto de los símbolos es menor que el alfabeto de la fuente, entonces a cada letra de la fuente le corresponde una cierta secuencia de símbolos, conocida por palabra del código. La cantidad de símbolos en la palabra del código determina lo que se conoce como longitud de la palabra del código.

Par que un código pueda ser utilizado con propósitos de transmisión y almacenamiento de la información se debe cumplir una serie de características siendo la más generalizadora de todas la que el código debe ser instantáneo.

Condiciones de un código

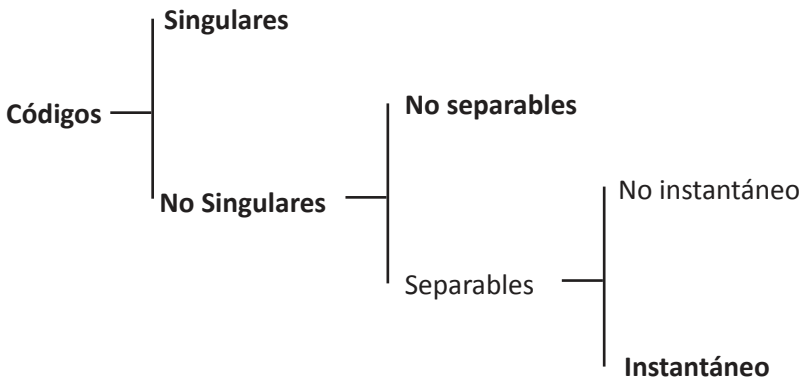
No singular: A cada mensaje se le asocia una secuencia distinta de símbolos del código.

Separable (Únicamente decodificable): No basta que las diferentes palabras del código xi, sean diferentes, sino que la secuencia de palabras del código: xi, xi2, xi3 ... se corresponda una única secuencia de símbolos.

Instantáneo: La condición necesaria y suficiente para que un código sea instantáneo es que sus palabras cumplan la condición de los prefijos (no exista palabra que sea prefijo de otra de longitud mayor).

a	b	c	d
m1-01	m1-0	m1-0	m1-0
m2-01	m2-01	m2-01	m2-10
m3-10	m3-001	m3-011	m3-110

a: Singular
b: No singular, no separable
c: no singular, separable, no instantáneo
d: no singular, separable, instantáneo.



A cada mensaje de la fuente se le asocia una secuencia de los símbolos del código, o sea una palabra del código.

Como vimos en el ejemplo de los alfabetos, el alfabeto Ingles tenia H(S)=1.3 Shannon /símbolo.

Esto significaba que el idioma ingles era redundante en un 70% y que sería posible reconstruir correctamente un texto en ingles si alguna de las letras se perdiera o cambiara a causa del ruido o la distorsión.

Evidentemente hay situaciones en la transmisión de datos que se requiere y es deseable que haya redundancia con el fin de aumentar la probabilidad de recibir un mensaje adecuadamente en un medio ruidoso.

Sin embargo la codificación de la fuente que vimos en el esquema de un sistema de transmisión de datos es un proceso encaminado a lograr una representación más eficiente de la información (eliminar la redundancia)

Además de ser una medida de la incertidumbre o indeterminación, la entropía permite evaluar la eficiencia del código utilizado, como a continuación se expone.

La construcción de códigos se debe regir por algún criterio de eficiencia, la cual pudiera verse en términos de asignar palabras de código cuya longitud fuera dependiente de la probabilidad de ocurrencia del carácter o mensaje. Esto es, a mensajes más probables asignar longitudes de código menores y a mensajes menos probables asignar longitudes de código mayores.

Se define la longitud promedio L de las palabras de código a partir de las probabilidades de ocurrencia del carácter o mensaje:

$$L = \sum_{i=1}^{q} P(si)\, li \qquad \text{símbolos/mensaje}$$

Donde q: cantidad de símbolos del alfabeto fuente (a codificar)

P (si): probabilidad de ocurrencia del símbolo (mensaje)

li : longitud de la palabra de código

Si r: #de símbolos del alfabeto del código, es obvio que la máxima cantidad de información que puede brindar la ocurrencia de cualquier símbolo del código es log r.

Log r: I = log 1/1/r = log r

Si como promedio, la cantidad de símbolos de código por mensaje es L, entonces la cantidad máxima de información que puede obtenerse debido a la ocurrencia de cualquier palabra del código es L log r.

De manera que: H (S)=<L log r

H(S): cantidad promedio de información que aporta la fuente

L log r: Cantidad de información que aporta el mensaje codificado

Esta información deriva el primer teorema de Shannon que afirma: los mensajes de un cierto alfabeto pueden codificarse de tal forma que la cantidad promedio de símbolos empleados para representar un mensaje de la fuente puede hacerse tan semejante como se quiera a la entropía de la fuente pero nunca menor que este valor.

Si se supone que
Ejemplo: sea una fuente s que emite 3 símbolos con probabilidades **P(s1) = ½, P(s2) = ¼ y P(s3)=¼. Su entropía será:**

H(s) = - ½ log2 ½ - ¼ log2 ¼ - ¼ log2 ¼ = 1,5 Shannon/símbolo

Si se aplica el código binario A= {0,10,11}, se tendrá que la longitud media de la palabra de código es:

L= 11.P(s1) + l2.P(s2) + l3.P(s3) = ½ + 2. ¼ + 2. ¼ = 1,5 s í m b o l o s (binarios)/ mensaje

Si se aplica el código binario B= {01,10,11}, la longitud media de la palabra de código L = 2.

Considerando que el alfabeto de código es binario, se define la eficiencia de un código como

□ = H(s)/L

(Esta definición se deriva del Primer Teorema de Shannon, el cual afirma que los mensajes de un cierto alfabeto pueden codificarse de tal forma que la cantidad promedio de símbolos necesarios para representar un mensaje de la fuente puede hacerse tan semejante como se quiera a la entropía de la fuente, pero nunca menor que ese valor).

Por tanto, para el código A del ejemplo anterior la eficiencia es 100%, mientras que para el código B la eficiencia es 75%.

Codificación eficiente

A continuación vamos a estudiar 2 códigos que se construyen de tal manera que tienden a asignar códigos más cortos a los símbolos más probables y con ellos logran códigos más eficientes.

Se definió la eficiencia de un código como

□ = H(s)/L

que se obtiene del Primer Teorema de Shannon: los mensajes de un cierto alfabeto pueden codificarse de tal forma que la cantidad promedio de símbolos necesarios para representar un mensaje de la fuente puede hacerse tan semejante como se quiera a la entropía de la fuente, pero nunca menor que ese valor.

En la que la longitud promedio L de las palabras de código se calcula a partir de las probabilidades de ocurrencia del carácter o mensaje:

$$L = \sum_{i=1}^{q} P(si)\, li \qquad \text{símbolos/mensaje}$$

donde q: cantidad de símbolos del alfabeto fuente (a codificar)
P(si): probabilidad de ocurrencia del símbolo (mensaje)
li : longitud de la palabra de código

 = H(s)/L log r
 = H(s)/L

Análisis de las vías para aumentar la eficiencia de un código.
1.- El único parámetro que puede variarse para mejorar la eficiencia de la codificación es la longitud promedio de las palabras del código.
2.- Para aumentar la eficiencia es necesario disminuir la longitud al máximo posible.

La forma más evidente de disminuir la longitud promedio de las palabras del código es asociando a los mensajes más probables, las menores longitudes de palabras del código y viceversa.

A continuación son examinados 2 algoritmos de codificación eficiente que reúnen las características mencionadas.

3.5 Código de Shannon-Fano

Sean n símbolos si (i=1,2,...,n) de probabilidades Pi distribuidas en orden decreciente. El método procede de la forma siguiente:

1º. Se toman los caracteres de más alta probabilidad y se obtienen dos subconjuntos, de tal modo que las sumas de las probabilidades de cada grupo sean, en la medida de lo posible, iguales.
2º. A todos los símbolos de la mitad superior se les agrega como primer símbolo del alfabeto de código el 1 (o el cero) y a los inferiores el símbolo 0 (o el 1).
3º. Los dos subconjuntos se dividen a su vez en dos subgrupos con iguales probabilidades sumarias.
Se repite desde el primer paso.

En general: El código se construye de la forma siguiente: Los mensajes se colocan en una tabla en orden creciente de probabilidades. Después, estos se dividen en r grupos (r es la base del código elegido para codificar) de forma tal que la suma de las probabilidades, en cada grupo, sea casi igual. A cada mensaje de estos grupos se le asocia como primer símbolo de la palabra codificada uno de los r símbolos del código Cada uno de los grupos se le asocia como primer símbolo de la palabra codificada, uno de los r símbolos del código. Cada uno de los grupos obtenidos es dividido a su vez en r subgrupos con probabilidades lo más parecidas posibles.
El proceso continua hasta que en cada grupo queden r mensajes.

Ejemplo
Obtener el código de Shannon-Fano para las siguientes letras del alfabeto s ={s1,s2,...,s8}

si	P(si)	Palabras código
s1	0,22	11
s2	0,20	101
s3	0,16	100
s4	0,16	01
s5	0,10	001
s6	0,10	0001
s7	0,04	00001
s8	0,02	00000

El código obtenido así es eficiente, ya que a los símbolos de probabilidades más altas le corresponden las palabras-códigos más cortas.
El cálculo de la longitud promedio de las palabras del código así obtenido es L=2,84 símbolos/mensaje, la entropía H(s)=2,76 Shannon/símbolo y la eficiencia del código es □=97,2 %.

Ejemplo #2
Realizar la codificación de Shannon-Fano efectuando la división inicial de tal manera que un subconjunto esté formado por los símbolos s1 y s2.

si	P(si)	Palabras código
s1	0,22	11
s2	0,20	10
s3	0,16	011
s4	0,16	010
s5	0,10	001
s6	0,10	0001
s7	0,04	00001
s8	0,02	00000

Se obtendrá una codificación diferente con una longitud de código promedio L=2,8 símbolos/mensaje, una entropía H(s)=2,76 Shannon/símbolo y una eficiencia □= 98,5%.

El código obtenido así es eficiente, ya que a los mensajes de probabilidades más

alta le corresponden las palabras – códigos más cortas.

3.6 Código de Huffman

El método de Shannon-Fano no siempre conduce a la construcción unívoca del código, ya que al dividir en subgrupos es posible elevar la probabilidad, tanto del subgrupo superior como del inferior. Así la tabla de los ejemplos anteriores puede haberse dividido de otras formas de manera que el código obtenido puede no resultar el mejor.

El algoritmo de codificación eficiente de Huffman garantiza una única solución en la síntesis del código, con la menor cantidad promedio de símbolos por mensajes para una distribución de probabilidades dada de los mensajes de la fuente.

Este es el código considerado como de mayor eficiencia. El resto de los códigos pueden igualrlo en determinado caso pero nunca superarlo. Las afirmaciones anteriores se argumentan como sigue:

- No hay dos mensajes con iguales secuencias.
- Cumple la condición de los prefijos.
- Los símbolos más probables son codificados con palabras de códigos de menor longitud l1 <= l2 <= l3.....<=lm
- Los dos mensajes menos probables son codificados con palabras de igual longitud lm = lm-1.

El método de Huffman consiste en lo siguiente:

1. Los mensajes del alfabeto se colocan en la columna principal se escriben en orden de disminución de las probabilidades.
2. Los 2 últimos mensajes se unen en uno auxiliar al que se le atribuye la probabilidad suma.
3. Las probabilidades de los mensajes que no han participado en la unión, así como la probabilidad suma obtenida, se ordenan otra vez en forma descendente en una columna adicional.
4. Los pasos 1,2 y 3 se repiten hasta obtener un único mensaje auxiliar con probabilidad igual a la unidad.

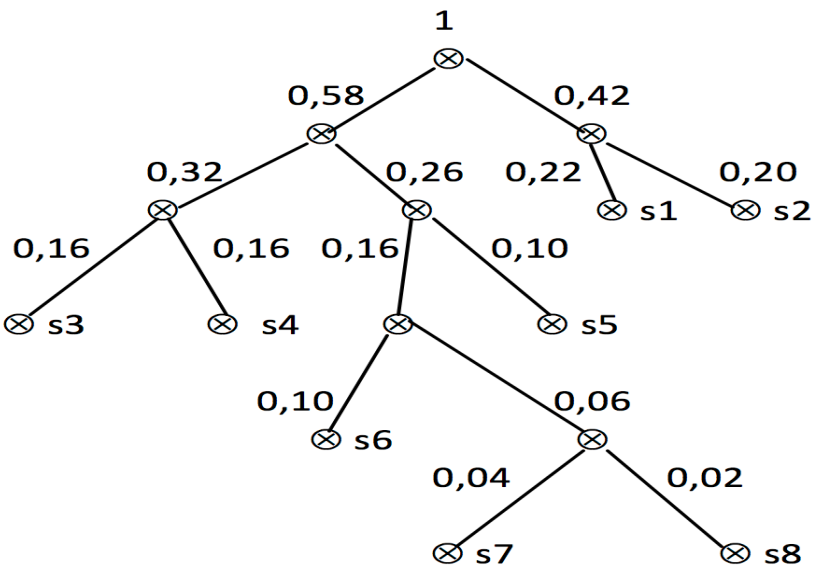
Luego es necesario obtener la codificación de cada símbolo original

Se construye un árbol de codificación. A partir del punto correspondiente a la probabilidad 1, se trazan 2 ramas otorgando a la de mayor probabilidad el símbolo

1 y a la de menor el símbolo 0. Se revierten las uniones en el orden inverso al que se formaron. Esta ramificación continúa hasta alcanzar la probabilidad de cada mensaje.

Ejemplo
Codificar por el método de Huffman la fuente siguiente:

si	P(si)	1	2	3	4	5	6	7
s1	0,22	0,22	0,22	0,26	0,32	0,42	0,58	1
s2	0,20	0,20	0,20	0,22	0,26	0,32	0,42	
s3	0.16	0,16	0,16	0,20	0,22	0,26		
s4	0,16	0,16	0,16	0,16	0,20			
s5	0,10	0,10	0,16	0,16				
s6	0,10	0,10	0,10					
s7	0,04	0,06						
s8	0,02							



De aquí:

Si	Código
s1	01
s2	00
s3	111
s4	110
s5	100
s6	1011
s7	10101
s8	10100

Calcular la eficiencia del código obtenido

L= 2(0,22)+2(0,20)+3(0,16)+3(0,16)+3(0,10)+4(0,10)+5(0,04)+5(0,02)= 2,8

 = 2,76/2,8 = 0,985

Inconvenientes de la codificación de Huffman

Las probabilidades de la fuente pueden no ser conocidas a priori. (Solución: empleo de mensajes patrones -> determinar estadísticas.
La naturaleza de la fuente puede significar un conjunto muy grande de símbolos (lenta codificación y requiere gran memoria su implementación)

Resumen de la visto hasta ahora

$$I(E)=\log_2 1/P(E)$$

$$H(s) = -\sum_{i=1}^{n} P(si) \log_2 P(si) \quad \text{Shannon/símbolo}$$

$$L = \sum_{i=1}^{q} P(si)\, li \quad \text{símbolos/mensaje}$$

$$\eta = H(s)/L$$

En las clases anteriores se ha realizado la caracterización desde el punto de vista

de la información emitida, de diferentes tipos de fuentes de mensajes, o sea, fue tratado el problema de la generación de la información.

Sin embargo nuestra labor no va a estar tan asociada al proceso de generación de la información como al proceso de transmisión de la información. Por lo que no es suficiente que una información sea generada sino que es necesario que esta llegue hasta el destinatario de la forma mas confiable posible.

El principio de la codificación eficiente se basa en garantizar un código de longitud variable que haga corresponder a los mensajes más probables longitudes de palabras de código menores y a los menos probables asignarles longitudes de código mayores. En este marco los códigos de Shannon-Fano y Huffman generan códigos eficientes y constituyen, de hecho, métodos de compresión de datos.

Este principio de eficiencia y los códigos que de él se obtienen tiene importantes desventajas en el proceso de comunicación:

- Se genera una dificultad práctica el tratar con códigos de longitud variable.
- El hecho de codificar bloques relativamente largo (para mensajes menos probables) trae como consecuencia la necesidad de acumular símbolos a la entrada del codificador.
- Las perturbaciones pueden crear situaciones específicas tales como la alteración de la longitud de un código que provocaría la incorrecta decodificación de una secuencia ulterior de símbolos (traza del error).

Los códigos eficientes son adecuados para canales sin ruido y en la práctica es más común la utilización de códigos de longitud fija.

La transmisión de datos es una actividad en donde los procesos no siempre tienen un carácter ideal. Los medios de transmisión influyen introduciendo uno de los procesos más indeseables del proceso de transmisión: el ruido. Se conoce como el ruido al conjunto de todas las interferencias que deterioran la calidad de una transmisión, generando errores que la dificulten.

La proporción de errores en un canal de transmisión de Datos se define como la relación entre el número de bits recibidos erróneos y el número de bits transmitidos.

Proporción de errores: No de bits recibidos / No de bits transmitidos

La proporción de errores que resulta aceptable en un sistema de transmisión de datos, depende de la naturaleza de la las aplicaciones que con los datos transmitidos y recibidos se realice.

Normalmente proporciones de errores mayores de 10-6 no resultan aceptables.

Partiendo de que en todo proceso de transmisión y recepción de datos existe la posibilidad de que aparezcan errores, es importante dotar a los dispositivos involucrados para detectar la presencia de errores y corregirlos.

Distancia de código

Ahora se tratarán códigos de longitud fija y se analizará la posibilidad de que ellos detecten si la información enviada es correcta o no y, en algunos casos, su capacidad de corregir errores presentes. Esto abarca una clase de códigos conocidos como detectores y/o correctores de errores.

Las acciones que pueden realizar estos códigos se deben a que ellos contienen información redundante. La redundancia se refleja en un conjunto de símbolos adicionales que portan información de control sobre las características del bloque. En algunos casos es posible determinar fácilmente cuáles son las posiciones ocupadas por los símbolos informativos y los símbolos de control, y en otro caso no es posible directamente.
En la entrada del dispositivo codificador aparece una secuencia k de símbolos de información. A la salida de este dispositivo le corresponde una secuencia de n símbolos donde: n>k

De todos modos en general se puede decir que en una palabra de código de longitud n existen k símbolos de información y (n-k) símbolos de control:

Información	**control**
k	**(n-k)**

En total pueden existir 2k diferentes secuencias a la entrada y 2n diferentes secuencias de salida.
De las 2n secuencias de salida solo 2k secuencias corresponden a las de entrada estas secuencias se denominan palabras del código permisibles.

Las restantes 2n-2k posibles secuencias de salida no son utilizadas para la transmisión se denominan palabras del código prohibidas o no permisibles.

Se denomina grado (multiplicidad) del error a la cantidad de símbolos deformados en una combinación codificada. Así la detección o corrección de un error de grado 1,2,3,... es la detección o corrección de un error cuando se alteran 1,2,3,... símbolos.

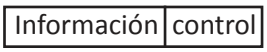
Es de señalar que, si el **alfabeto código es el binario**, entonces existirán:

2n palabras de código posibles
2k palabras permitidas
2n - 2k palabras prohibidas o no válidas

Ejemplo
Sea una fuente de 4 símbolos s1,s2,s3,s4 codificadas como sigue, utilizando un bit de paridad para el control:

Símbolo	código (sin bits de control)	código (con bits de control)	códigos posibles (con 3 Bits)
s1	00	000	000 *
s2	01	011	001 •
s3	10	101	010 •
s4	11	110	011 *
			100 •
			101 *

* Combinación permitida	bit de paridad	110 *
• Combinación prohibida	(redundancia)	111 •



Existen **n** = 3 bits totales, **k** = 2 bits de información y **n-k** = 1 bit de control. Es decir, 23 = 8 combinaciones posibles, 22 = 4 combinaciones permitidas y 23 - 22 = 4 combinaciones no legales.

Si se produce el cambio de un símbolo (un bit) de una combinación codificada el receptor lo detecta. Un cambio simultáneo de dos símbolos, sin embargo no sería detectado en recepción. Por tanto, ese código es detector de errores de grado 1.

Ejemplo

Sea ahora una fuente de 2 símbolos s1, s2 codificadas con 3 bits y utilizando 2 bits de control como sigue:

Símbolo	código (sin bits de control)	código (con bits de control)	códigos posibles (con 3 Bits)
s1	**0**	**000**	**000 ***
s2	**1**	**111**	**001**

010 => 000
=> 0

100

Combinaciones posibles: 23 = 8 **011**
Combinaciones permitidas: 21 = 2 **110**
Combinaciones prohibidas: 23 - 21= 6 **011** **=> 111 => 1**
111 *

Obsérvese que si se produce un error de grado 1 se detecta y se corrige, si se produce un error de grado 2 se detecta.

Como se supone, será necesario elaborar códigos para alcanzar determinada capacidad detectora o correctora y eso está relacionado con la composición de la palabra de código. Antes de formalizar esta cuestión se definirán algunos términos.

Distancia de código

Es el grado de diferencia entre dos palabras codificadas diferentes cualesquiera, expresada por el número de signos que difieren entre una y otra combinación.

Por ejemplo, si las palabras codificadas si las palabras codificadas son 011 y 101, la distancia de código será 2. Si las palabras de código son 000 y 111, la distancia de código será 3.

Para obtener la distancia del código entre 2 palabras codificadas de un código binario, es suficiente contar la cantidad de 1 en la suma módulo 2 de estas palabras.
1001111101
1100001010
0101110111
d=7

Distancia de codificación mínima
La distancia mínima encontrada entre todos los pares de palabras codificadas se llama distancia de codificación mínima (dmin).

Para el primer ejemplo dmin= 2 y para el segundo ejemplo dmin= 3.

Relación entre la capacidad detectora y correctora y la distancia mínima
Se concluyó en el primer ejemplo que el código definido (dmin= 2) puede detectar errores simples y que en el código del segundo ejemplo (dmin= 3) puede detectar errores dobles y corregir errores simples. En general se demuestra que para que un código pueda detectar errores de grado r se requiere que

$$\mathbf{dmin} \geq \mathbf{r + 1}$$

Análogamente, un código podrá corregir errores de grado s si se cumple

$$\mathbf{dmin} \geq \mathbf{2s + 1}$$

Ejemplos
Puede suponerse que la de codificación después de la recepción se realiza de forma tal que la palabra codificada recibida se identifique con aquella palabra permisible que se encuentra con ella, a la menor distancia de código.

Código con d=1 todas las palabras del código son permisible.
n=3. Las palabras permisibles forman el conjunto.
000, 001, 010, 011, 101, 110, 111.
Cualquier error simple transforma una palabra en otra permisible. Este es el caso de un código que no pose capacidad detector de error.

Código con d=2, n=3.
Palabras permisibles: 000, 011,101, 110
Palabras prohibidas: 001, 010, 100, 111
Es un código detector de errores simples o de grado 1.
Ahora se analizarán algunos esquemas de detección de errores tomando como fuente de mensajes un código estándar y de los más empleados: el **código ASCII.**

Indicadores de calidad de un código corrector

Uno de las características fundamentales del código corrector de errores es la redundancia del código, que indica el grada de prolongación de la palabra del código para el logro de un determinada capacidad correctora. Si por cada n símbolos de la secuencia de salida y n-k de chequeo, entonces la redundancia relativa al código puede ser expresada por cualquier de las relaciones siguientes:

Rn=(n-k)/n
Rk=(n-k)/k

Los códigos que garantizan una capacidad de corrección dada, con la mínima redundancia posible se denominan óptimos.
En relación con la determinación de los códigos óptimos debe evaluarse, por ejemplo, la mayor cantidad posible de Q palabras permisibles, que sea capaz de corregir errores mutuamente independientes de multiplicidad s.
Esto equivale a buscar la cantidad de combinaciones con distancia mínima entre ellos de dc=2s+1.

La cantidad total de diferentes errores corregibles para cada una de las palabras permisibles.

Otro indicador no menos importante de calidad de un código corrector, es la complejidad de la realización técnica de los procesos de codificación y descodificación.
Si la información debe transmitirse por un línea de comunicación de acción lenta (no confiable, costosa) y la complejidad de los dispositivos codificador y decodificador (realizados con elementos muy confiables y de gran velocidad) no juega un papel fundamental. El factor decisivo, en este caso, es aumentar la eficiencia en la utilización de la línea de comunicación y por esto es deseable utilizar códigos correctores con mínima redundancia.

Si el código corrector debe ser utilizado en un sistema, realizado con elementos de confiabilidad y velocidad iguales o cercanas a la confiabilidad y velocidad de los equipos codificador y decodificador, entonces el criterio de calidad del código corrector es la confiabilidad del sistema en su conjunto, o sea, tomando en cuenta las posibles distorsiones o interrupciones en los equipos de codificación y descodificación. En este caso frecuentemente son más racionales los códigos de gran redundancia, pero que sean sensibles a la codificación.

Los diferentes métodos para generar información redundante con vistas al control de errores se diferencian entre sí en la cantidad y ubicación de bits de redundancia y en la complejidad del algoritmo de codificación.

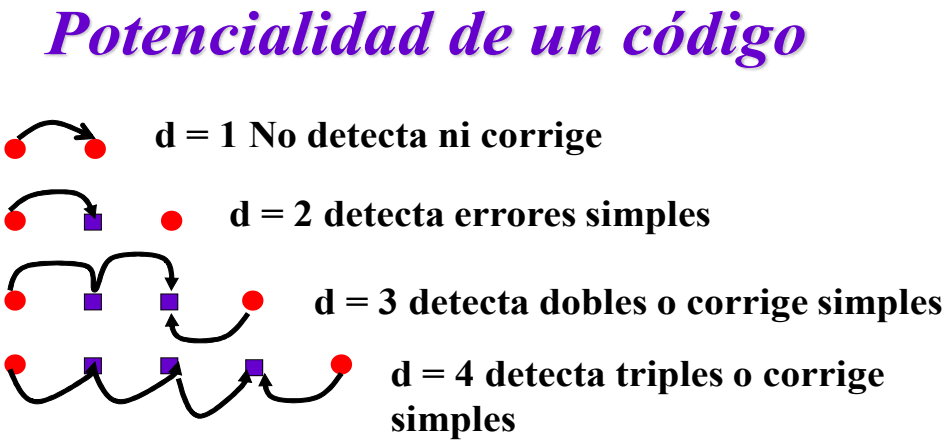


Figura 10: Potencialidad de un código.

Interpretación geométrica de los códigos de bloques correctores

Si se tiene de n dimensiones donde cada palabra es representada por un punto.

Cualquier palabra codificada binaria de n posiciones puede ser interpretada como el vértice de un cubo unitario-dimensional, o sea, de un cubo con longitud de arista unitaria.

En general el cubo unitario n-dimensional tiene 2n vértices, que es igual a la mayor cantidad posible de palabras del código.
Este modelo brinda una interpretación geométrica sencilla a la distancia entre diferentes palabras del código que corresponden a la menor cantidad de aristas del cubo unitario por las que es necesario pasar para ir de una palabra a otra.

Espacio de codificación

Espacio de n dimensiones donde cada palabra es representada por un punto

- *Distancia entre palabras*: -cantidad de segmentos quebrados entre dos palabras

- *Dist. entre pal.*= cantidad dígitos diferentes

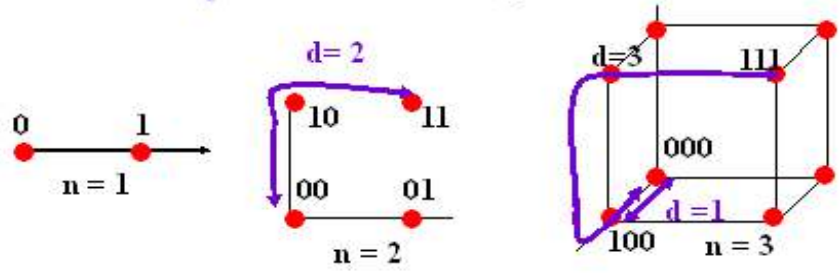


Figura 11: Espacio de codificación.

Vamos a ver una técnica para conocer qué cantidad de bits de control será necesario añadir en un código para lograr determinado grado de detección y corrección de errores.

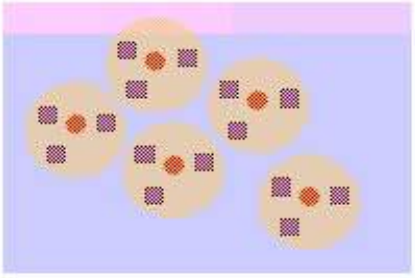
Límite de Hamming

Expresión que nos brinda la relación entre la cantidad total de dígitos que conforman el código, la cantidad de dígitos que representan información y la cantidad de dígitos de control.

k dígitos de información
b dígitos de chequeo
n cantidad total de dígitos del código
De manera que se podrán tener 2k palabras en el código.

A partir de la interpretación geométrica de los códigos de bloques correctores, para la corrección de r errores. Cada palabra del código permitida tiene asociada una esfera de radio r

Límite de Hamming



Relación entre n, k y b para la corrección de r errores.

2^k palabras de código

2^n palabras totales

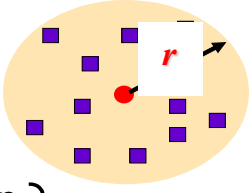
2^k esferas de radio r con centro en " ● "

Figura 12: Límite de Hamming.

Cantidad de palabras en cada esfera.

Límite de Hamming

Esfera de radio r con centro en palabra de código



Cantidad de palabras en cada esfera :

1: palabra de código

n: palabras sep. a distancia 1

$\binom{n}{2}$ **: palabras a dist. 2** ... $\binom{n}{r}$ **: palabras a dist. r**

palabras en una esfera $= 1 + n + \binom{n}{2} + \ldots + \binom{n}{r}$

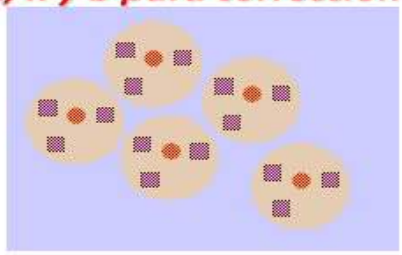
Figura 13: Potencialidad de un código.

Si se tienen 2k esferas.

Límite de Hamming

Relación entre k, n y b para corrección de r errores

$n = k + b$



$$2n \geq 2k \left[1 + n + \binom{n}{2} + \binom{n}{3} + \ldots + \binom{n}{r}\right]$$

$$2k <= \frac{2n}{\left[1 + n + \binom{n}{2} + \binom{n}{3} + \ldots + \binom{n}{r}\right]}$$

Figura 14: Límite de Hamming relación entre K,N B.

Caso particular de corrección de errores simples

$$r = 1$$

$$2^k \square \frac{2^n}{n+1}$$

$$2^b \; ^3 \quad n + 1$$

k	1	2	3	4	5	6	7	8	9
b	2	3	3	3	4	4	4	4	4
n	3	5	6	7	9	10	11	12	13

Si k ↑ b/n ↓

Figura 15: Caso particular de corrección de errores simples.

Descripción matricial de un código algebraico

Las reglas se expresan a través de un conjunto de ecuaciones cuya solución determinan los dígitos de chequeo.

Los bits de chequeo van a estar en función de la cantidad de bits de información y de un sistema de ecuaciones, obtenido a partir de HxT=0

Se tiene una matriz H, que es la matriz del código compuesta por n columnas y b filas donde:

n = Total de dígitos del código
b = dígitos de chequeo.

Características que tiene que tener la matriz H

- b filas x n columnas
- ninguna columna completa de 0s.
- todas las columnas son diferentes

Características deseables de la matriz H

- **El contenido de cada columna es su posición de la matriz**

Para la obtención del sistema de ecuaciones es necesario multiplicar dicha matriz por un vector de transmisión T, que representa la palabra a transmitir.

- Características del vector T
- Considerar que en la palabra codificada los bits de chequeo ocupan posiciones de potencias de 2.
- El resto de los bits son los de información.

Para formar el sistema de ecuaciones.
HxT=0

Cualidades del sistema:

- Cada bit de chequeo Ci, los bits aparece en una sola ecuación.

Para la detección de errores en lado receptor se comprueba HxR

Si:

- HxR = 0 ok no hay error detectado
- HxR<> 0 hay error detectado
- HxR =H(T+E) = HxE = S S= matriz síndrome

Veamos la posibilidad, en el ejemplo anterior, de un error simple (en una posición)

Códigos Algebraicos

La regla se expresa a través de un conjunto de ecuaciones cuya solución determina los dígitos de chequeo.

Representada matricialmente:

$$H_{bxn} = \begin{bmatrix} h_{11} & h_{12} \ldots\ldots & h_{1n} \\ h_{21} & h_{22} \ldots\ldots & h_{2n} \\ & \ldots\ldots & \\ h_{b1} & h_{b2} \ldots\ldots & h_{bn} \end{bmatrix} \qquad T_{nx1} = \begin{bmatrix} x_1 \\ x_2 \\ .. \\ x_m \\ c_1 \\ \ldots \\ c_b \end{bmatrix}$$

$$H \bullet T = 0$$

Códigos algebraicos para la corrección de errores simples

Condiciones de H

b filas y n columnas

Ninguna columna entera de ceros

Todas las columnas diferentes

Los dígitos de chequeo en las posiciones de las columnas que contienen un único "1".

Códigos Algebraicos

R : Palabra recibida, en el receptor se determina :

$$H \bullet R = 0 \ ?$$

$H \bullet R = 0$ No hay error detectado

$H \bullet R \ ^1\ 0$ Hay error y este puede detectarse o corregirse según la redundancia del código

$$R = T + E ; \qquad H \bullet R = H \bullet (T + E) = H \bullet E = S$$

S: *Matriz Síndrome*, distintivo de la posición del error en códigos correctores

Figura 16: Código algebraico.

3.7 Códigos cíclicos

El concepto de código cíclico es considerado uno de los mayores avances en la historia de la teoría de la codificación. Este método es solo aplicable en la transmisión por bloque.

Para la aplicación de este método el bloque de datos está compuesto por k bits y es representado por un polinomio de grado k – 1.

Si el bloque de datos está compuesto por una sucesión de bits de tipo:

Ak-1, Ak-2...., A1, A0 donde Ai puede ser 0 o 1, entonces el polinomio correspondiente es:

M(x) = Ak-1 Xk-1 + Ak-2 Xk-2 + … + A1X + A0

Este es entonces manipulado por otro polinomio P(x) denominado polinomio generador del código cíclico, que debe ser de grado menor que M(x) y que contiene al menos dos términos.

La selección de P(x) viene dada por el patrón específico de errores que se quiere detectar. Asúmase que ya P(x) ha sido seleccionado. La operación de codificación consiste en los siguientes pasos:

1. Multiplicar M(x) por Xr donde r es el grado de P(x).
2. Dividir, módulo 2 M(x) por Xr entre P(x). El resto estará representado por R(x).
3. El mensaje a transmitir representado por T(X) será:

T(x) = (M(x) * Xr) XOR R(x)

Luego en el receptor para comprobar que está correcto lo transmitido se divide T(x) entre P(x) y si no existe error el resto debe dar cero. También se emplea el mecanismo de dividir el polinomio generado por los bits de información en el receptor entre P(x) dando el resto R(x) que debe ser igual al que viene en la palabra codificada formada por los coeficientes de M(x)+R(x).

CAPÍTULO IV. COMPRESIÓN DE DATOS

4.1 Introducción

El auge que han tenido las redes de ordenadores hace que cada vez más usuarios pidan más prestaciones a la red sobre la que están conectados. Prestaciones que, como siempre, están por encima de las posibilidades reales. Cuando se habla de posibilidades se tiene en cuenta principalmente a la velocidad de transferencia de datos. Este es el principal handicap al que se enfrentan todas las redes.

En este entorno, para conseguir mayores prestaciones de velocidad, los implementadores de los programas deben recurrir a técnicas que les permitan superar de alguna manera las deficiencias físicas de la red y de los equipos de conexión.

Una de las técnicas más importante en este sentido es la compresión de datos. La compresión de datos es beneficiosa en el sentido de que el proceso de compresión-transmisión-descompresión es más rápido que el proceso de transmisión sin compresión.

Pero no sólo es para la transmisión para lo que se usa la compresión. También para el almacenamiento masivo. La necesidad de almacenamiento también crece por encima de las posibilidades del crecimiento de los discos duros o memoria. Basta pensar, por ejemplo, en los grandes servidores de vídeo en demanda con cientos o miles de películas, ocupando cada una varios Gigabytes.

El principio de la compresión de datos es simple. Prácticamente todas las formas de datos (texto, imágenes, video, audio) contienen elementos redundantes. Los datos pueden comprimirse eliminado esas redundancias.

Desde luego, que cuando se pretenda recuperar los datos desde su almacenamiento o al ser recibidos en un enlace de comunicaciones debe ser posible expandir los datos. Con este objetivo, cuando un dato se comprime, se debe sustituir algún tipo de código por los datos eliminados de modo que el receptor, basado es ese código, reconstruya los datos originales.

4.2 Tipos de compresión

Es muy difícil clasificar los distintos tipos de compresión de datos que existen, debido a que, por un lado existen muchos algoritmos: LZ77, LZ78, LZW, Huffman, aritméticos, fractales, MPEG, JPEG, etc.

Además, hay muchas aplicaciones que utilizan la compresión con distintas expectativas: compresión de datos, vídeo y voz, compresión en tiempo real, etc. A su vez, cada uno de estos usos requiere unas características de velocidad,

reversibilidad (que el algoritmo pueda ser aplicado de forma reversible para obtener los datos originales), pérdida mínima de información (en el caso de los algoritmos con pérdida de información, etc.).

Debido a esto, se ha elegido una clasificación muy general tomando como característica de división la reversibilidad del algoritmo. Así, los algoritmos de compresión se pueden dividir en dos tipos:

- **Compresores sin pérdidas:** en el sentido de que guarda absolutamente toda la información original (es reversible). Se utilizan para la compresión de datos, en los que no se puede dar pérdida de información.
- **Compresores con pérdidas.** La compresión hace que se pierda información de la fuente original. Sin embargo, esta pérdida es insignificante en comparación con la ganancia en compresión. Se utiliza sobre todo en imágenes y sonido, donde se puede "engañar" a los sentidos, donde una pérdida de calidad apenas es percibida (pero ocasiona un ratio de compresión mucho mayor).

Compresión sin pérdidas

Ahora se describirán algunos tipos de compresión sin pérdidas más comunes. Se comenzará exponiendo en líneas generales cuáles son los tipos de compresión sin pérdidas más conocidas y utilizadas, exponiendo algunos ejemplos de algoritmos reales que apliquen cada tipo de compresión.

4.3 Tipos de compresión sin pérdidas

Hay básicamente dos tipos de compresores / algoritmos de compresión sin pérdidas:

- Compresores estadísticos.
- Compresores basados en diccionario.

Compresores estadísticos

Habiendo usado una medida de información basada en las probabilidades es natural la presencia de algoritmos que utilicen las propiedades estadísticas de la fuente de información para mejorar la codificación (el conjunto de símbolos de salida asociados a cada mensaje emitido por la fuente) de los mensajes de la fuente. Estos son los compresores estadísticos.

Este tipo de compresores parte de:

- una fuente de información de n mensajes

- las probabilidades de aparición de cada mensaje de la fuente (que pueden ser extraídas a priori de forma experimental o pueden ser dadas y fijas)
- un alfabeto de salida que consta de una serie de símbolos (por ejemplo, el alfabeto binario consta de los símbolos 0 y 1).

y su objetivo es asignar una codificación para los mensajes de la fuente.

Se sabe que una codificación es una función que a cada mensaje de la fuente asigna una cadena de símbolos del alfabeto de salida. Esta codificación debe ser tal que explote la redundancia en la información dada por la fuente para producir compresión. (Si utilizamos un alfabeto binario, la esperanza matemática de la longitud de las cadenas de símbolos de la codificación, tomando como probabilidades las de los mensajes a los que representan, se debe acercar a la cota teórica de Shannon. Si es igual, la compresión es perfecta: hay una máxima eficiencia).

Son los compresores que utilizan la información de las probabilidades de los mensajes de la fuente para construir una codificación.

Por ejemplo, supóngase que las fuentes de información son ficheros ASCII de 8 bits y que cada carácter es un mensaje de la fuente (256 mensajes). El conocer las probabilidades de cada mensaje implica que el compresor debe realizar una primera pasada por el fichero para recuperar las frecuencias (o probabilidades, ya que son inversas) de cada mensaje (byte ASCII).

Esto puede ser un problema, sobre todo para dispositivos de cinta de una sola pasada. Sin embargo, se puede argumentar que la codificación se puede realizar sobre bloques ó utilizar algoritmos adaptativos.

Entre los compresores estadísticos podemos encontrar varios tipos:

- Compresores del **tipo Huffman)ó Shannon-Fano** (ya analizado)
- Compresores **aritméticos**

La codificación de Huffman no alcanza el máximo posible de compresión a menos que las probabilidades involucradas sean potencias enteras negativas de 2 (2-m). La codificación aritmética logra alta compresión para casos en que las probabilidades de los mensajes no sean de la forma 2-m, pero próximas a ellas, a un mayor costo computacional.

Se basan en las probabilidades de la ocurrencia de los mensajes a la entrada. Sin embargo, la codificación se basa en la representación de un valor del intervalo [0;1]. Por ejemplo, si se desea codificar una entrada de 2 símbolos X y Y, con probabilidades P(X)=2/3 y P(Y)=1/3, al recibir una X se divide el intervalo [0;1] y se toma el intervalo [0;2/3]. En caso de recibirse una Y, se toma el intervalo [2/3;1].

En cada paso se divide por los 2/3 el subintervalo inferior o por 1/3 el subintervalo superior. Al final, el código emitido en binario en un número que cae en el intervalo y se elige el que utilice menos bits en su representación.

- Compresores **predictivos** Son totalmente adaptativos, ya que procuran predecir el siguiente mensaje de la entrada tomando como base el conocimiento de la entrada ya procesada hasta ese momento (probabilidades). Si el mensaje a la entrada coincide con el predicho, su codificación se podrá hacer con menos bits.
- **Codificación según la frecuencia de aparición de los caracteres del alfabeto en uso**. Se va registrando la frecuencia de aparición de cada carácter del alfabeto en uso en los datos a codificar y simultáneamente se va actualizando una tabla de codificación mediante la cual se asignan palabras de código más cortas a los caracteres más frecuentes. (Huffman).
- **Codificación según la frecuencia de aparición de cadenas de caracteres.** Se trata de mejorar la eficiencia en la compresión codificando, no un carácter aislado, sino una cadena de caracteres. Esto se apoya en que en textos comunes la aparición de un carácter depende de la secuencia que lo precedió. Por ejemplo en español es casi segura la aparición del carácter "u" después del carácter "q".

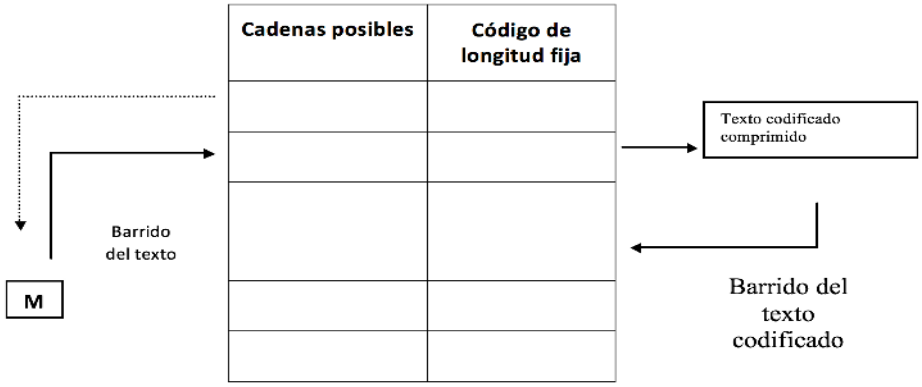
4.4 Compresores basados en diccionario

Los métodos adaptativos son muy susceptibles a la propagación de errores. Es necesario que los diccionarios que se usan en recepción sean actualizados a partir de la propia información codificada que se va recibiendo y que se mantenga en el mismo estado que en el transmisor.

Esto significa que si se altera un bit en el proceso de transmisión puede suceder, no sólo la alteración de un carácter, sino también la alteración de todo el mensaje.

Una alternativa en aplicaciones críticas, es el uso de diccionarios estáticos. Estos métodos parten de conocer la estructura estadística de los datos que se desean codificar.

Con este conocimiento se construye un diccionario que asocie palabras de código de longitud fija a las cadenas que tengan alta probabilidad de aparecer en cualquier secuencia de datos, del tipo que se desea codificar.



Un diccionario de este tipo puede usarse para codificar y decodificar. El aspecto delicado del método es el diseño del diccionario, ya que debe contener las cadenas necesarias para los textos que se van a comprimir. Este es un proceso bastante complejo.

4.4.1 Selección del tamaño del diccionario

Supóngase que la secuencia de datos contiene textos en español y que se quiere asociar a cada cadena una palabra de código de 8 bits. Esto limita a 256 el número de cadenas a representar en la tabla.

Todas las cadenas de longitud uno deben estar representados en principio en la tabla pues cualquier cadena de otra longitud que aparezca en el texto, y no esté codificada, debe codificarse en las subcadenas componentes.

Entonces se puede aceptar inicialmente que el alfabeto está compuesto por letras mayúsculas, minúsculas, signos ortográficos, los dígitos decimales, algunos símbolos especiales (+,-,etc.) y en total requiere unos 86 códigos.

Además de los caracteres anteriores hay que tener en cuenta que los textos se encuentran almacenados en ficheros construidos con un algún procesador de texto que agregan caracteres de control. También tener en cuenta si existen caracteres que se utilizan para la comunicación de datos. Todo este análisis puede conducir a la necesidad de codificar unos 100 caracteres. Quedan 156 posiciones por ocupar.

4.4.2 Selección de las cadenas para completar el diccionario

Las 156 posiciones restantes serían ocupadas por cadenas de largo ≥ 2 bytes.

¿Cuáles cadenas de este tipo incluir? Las más probables o las de mayor frecuencia de aparición, lo cual requiere un estudio estadístico. Algunas de las cadenas a

codificar pueden ser las siguientes:

_a **ir**

_c **un**

_al **_o**

_con **ac**

_de los

ión

_un

or

Ejemplo: Supóngase que ha sido construido el siguiente diccionario y que con él se necesita comprimir el texto.

"Comprimir una oración" 21 bytes

Cadenas posibles	Código de longitud fija
A	0
B	1
C	2
D	3
...	...
a	27
.....	...
p	43
r	45
...	...
_a	100
_c	101
_al	102
ión	103
_un	104
or	105
ir	106
un	107
om	108
_o	109
ac	110
...	...

C → 2

om → 108

p → 43

r → 45

i → 35

m → 39

ir → 106

_un → 104

a → 27

_o → 109

r → 45

ac → 110

ión → 103

El texto comprimido contendrá 13 bytes →39% de compresión.

En general, en los compresores basados en diccionario, la idea es más dinámica. Se trata de construir un diccionario tomando como referencia la entrada procesada hasta ese momento. El diccionario contiene las cadenas de mensajes (que pueden ser cadenas de caracteres ASCII o bytes).

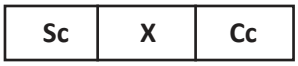
Estas cadenas están identificadas por un índice, de manera que índice y cadena son de correspondencia biunívoca. El resultado práctico a todo esto es que si en algún momento la entrada que se está procesando actualmente es una cadena que está presente en el diccionario, el compresor puede dar como salida el índice que identifica esa cadena en el diccionario.

Teniendo en cuenta que las cadenas pueden ser arbitrariamente largas, la producción del índice en lugar de la cadena representa un ahorro de información, y por lo tanto, compresión.

Incluidos dentro de este grupo están los algoritmos RLE, que puede ser considerado como poseyendo un diccionario de longitud 1 byte, el LZW y otros derivados de LZ78 y las variantes de LZ77. Se verán algunos de ellos de una manera simplificada.

4.5 Compresión RLE

RLE se usa para comprimir cualquier tipo de secuencia de datos repetidos. En la siguiente Figura se puede visualizar la técnica cuando se aplica a una secuencia de caracteres.



Sc: carácter especial que indica que sigue la compresión.

X: cualquier carácter de dato repetido.

CC.: número de veces que se repite el carácter comprimido.

El objetivo del algoritmo es reducir las cadenas de caracteres idénticos a una indicación del carácter y la longitud de repetición. Así el transmisor busca la secuencia de caracteres repetidos para reemplazarlos. En este caso, cualquier secuencia de caracteres repetidos puede ser eliminada y reemplazada por un código de 3 caracteres.

Ejemplo:

Cadena original de datos	Cadena de datos codificada
A******45.76	ASc*645.76
+++++++	Sc+7
parabbbbbesto	paraScb5esto

El código consiste de un carácter especial, que indica supresión, seguido por el carácter que va a suprimirse, y seguido por la cantidad de caracteres suprimidos. Así, cualquier secuencia de 4 o más caracteres idénticos pueden suprimirse con una reducción neta en el total de caracteres transmitidos.
La eficiencia de RLE depende del número de ocurrencia de caracteres repetidos en los datos a comprimirse y de la longitud promedio de la secuencia repetida. La medida estándar de la eficiencia de la compresión es la relación (ratio) de compresión, que es la relación de la longitud de los datos sin comprimir entre la longitud de los datos comprimidos.

Una variación al procedimiento ilustrado en el ejemplo es el usado en MNP,

que está implementado en algunos estándares de módems. MNP es tal vez la implementación más utilizada de codificación RLE . Con MNP cualquier cadena de 3 o más caracteres repetidos se representa en el formato

CXXX

donde C es la cantidad de caracteres que exceden de tres y X es el carácter. Por ejemplo:

Datos originales	Datos comprimidos
AAA	0AAA
BBBB	1BBB
CCCCC	2CCC

Se ha incluido al RLE dentro de los compresores basados en diccionario porque se puede considerar que utiliza un diccionario deslizante de tamaño 1 byte para predecir el siguiente carácter a la entrada, aunque así también podría ser clasificado de predictivo.

4.6 LZ77

En **LZ77** es donde realmente se concentra el interés actual. Estos algoritmos tienen tantas decisiones de diseño que cada uno implementado es distinto a los demás. Esto hace que no se puedan patentar o que se puedan saltar de una forma u otra las patentes actuales. De ahí el interés.

Se describirá este algoritmo mediante un ejemplo. Considérese la siguiente frase:

tiene 54 bytes=54*8=432 bits de longitud.

El algoritmo procesa este texto de izquierda a derecha. Inicialmente se le asigna a cada carácter un patrón de 9 bits que consiste en el binario 1, seguido de la representación ASCII extendido del carácter.

En la medida que el algoritmo avanza se buscan secuencias repetidas. Cuando se encuentra una repetición el algoritmo continúa explorando hasta que termina la repetición. Es decir, cada vez que ocurre una repetición, el algoritmo incluye tantos caracteres como sea posible.

La primera secuencia encontrada es EL **BLANCO PERR**. Esta secuencia se sustituye

por un apuntador a la secuencia anterior y la longitud de la secuencia. En este caso la secuencia EL **BLANCO PERR** ocurre 28 posiciones antes y la longitud de la secuencia es de 14 caracteres.

Supóngase que hay 2 opciones de codificación: una con un apuntador de 8 bits y una longitud de secuencia de 4 bits; o un apuntador de 12 bits y una longitud de secuencia de 6 bits.

Un encabezamiento de un bit en 0 indicará la presencia de una repetición codificada, a continuación otro bit indica cuál es la opción seleccionada (0 si es la primera opción y 1 si es la segunda). Por tanto, la segunda ocurrencia de EL BLANCO PERR se codifica como:

<00b><28d><14d> = 00 0001 1100 1110

Dónde:

b-→ Número Binario
d-→ Número Decimal
En resumen:

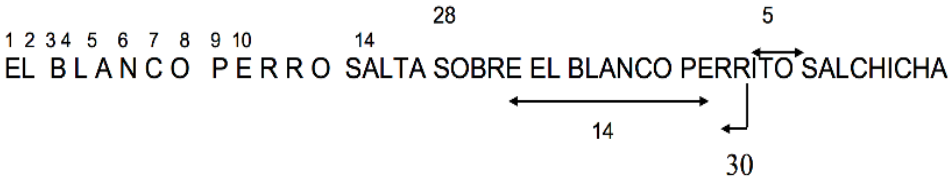
Para la primera Opción (0)

Bit de Comienzo de Repetición Codificada(0)	Bit de Opción de Codificación(0)	8 Bits para el apuntador	4 Bits para la longitud de la secuencia

Para la segunda Opción (1)

Bit de Comienzo de Repetición Codificada(0)	Bit de Opción de Codificación(1)	12 Bits para el apuntador	6 Bits para la longitud de la secuencia

Las partes restantes del mensaje comprimido son IT, la secuencia <00b><30d><5d>=00 0001 1110 0101 (que sustituye la secuencia 0_SAL) y la secuencia CHICHA.



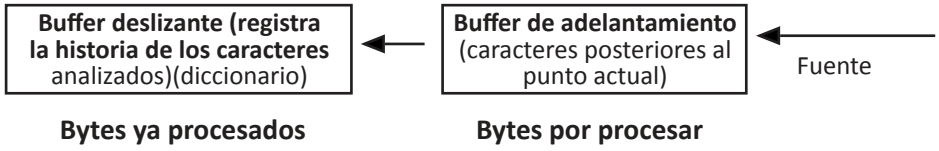
O sea, la secuencia a transmitir sería

EL BLANCO PERRO SALTA SOBRE 00b 28d 14dIT00b 30d 5dCHICHA

El mensaje comprimido consiste de 36 caracteres de 9 bits y 2 códigos, para un total de 36*9+2*14= 352 bits. La relación de compresión será entonces 432/352=1,23.

Las variantes de LZ77 también mantienen un registro de los últimos caracteres procesados de la entrada, igual que las LZ78. Pero al contrario que estas, no construyen un diccionario explícito. En cada momento, el algoritmo se encuentra procesando en un punto de la entrada, los n caracteres anteriores forman la "historia" del algoritmo ó "ventana", lo que equivale al "diccionario". Los caracteres posteriores al punto actual forman lo que se denomina el "lookahead buffer" ó buffer de adelantamiento. En cada paso, la cadena que comienza en el punto actual de la entrada se busca hacia atrás en la "historia". Si se encuentra una coincidencia que sea lo suficientemente larga como para tenerla en cuenta (después se comentará esta decisión), a la salida se sustituye la cadena coincidente por un par que indica el "desplazamiento hacia atrás" y la "longitud" de la coincidencia con la cadena hacia atrás. Como los pares (desplazamiento, longitud) ocupan menos que la cadena que coincidió, se obtiene compresión. Si no se encuentra una coincidencia, la salida es una copia literal de la entrada. Posteriormente se avanza la ventana (es decir, se avanza en la entrada) bien longitud si hubo coincidencia, bien un carácter si no la hubo.

El hecho de ir desplazando la ventana sobre la entrada hace que estos algoritmos sean llamados de "ventana deslizante". La siguiente figura ilustra el proceso del algoritmo:



El proceso inverso de descompresión debe salvar los últimos caracteres de su salida. Cuando se encuentra una cadena codificada, entonces usa los campos de apuntador y de longitud para reemplazar el código con los valores reales de la cadena de texto.

4.7 Compresión con pérdidas

La compresión sin pérdidas tiene como límite de compresión el tamaño de cualquier archivo la entropía del archivo. Con esa técnica es posible recuperar los datos originales bit por bit. Esta característica generalmente se considera fundamental en archivos de texto, bases de datos y archivos ejecutables.
Sin embargo, hay tipos de archivos donde no se requiere una recuperación

completa de la información original (voz, imagen, video). En estos casos se permite alguna cantidad de error en reconstruir la información original y se puede utilizar la compresión con pérdidas.

Así, uno de los compromisos en la compresión con pérdidas es la relación de compresión y la fidelidad, es decir, lograr una alta relación de compresión sin que la pérdida de bits de información impida la reconstrucción de los datos con una aproximación aceptable a la información original.

Aquí se describirá de manera general la técnica de compresión JPEG, pero antes se mostrarán, a nivel descriptivo la transformación discreta del coseno (DCT) que se utiliza en muchas aplicaciones, y la cuantización vectorial.

Codificación por fuente
En general, la codificación por fuente es un tipo de codificación que aprovecha las propiedades de los datos para lograr mayor compresión al costo de algunas pérdidas.

Uno de los ejemplos de codificación por fuente consiste en transformaciones. Gracias a la transformación de las señales de un dominio a otro, la compresión se puede lograr de una manera más fácil.

Por ejemplo, con la transformada de Fourier una función en el tiempo se representa como una lista de amplitudes. Con los valores exactos de todas las amplitudes se puede reconstruir la función original exactamente. Sin embargo, considerando sólo los valores de las primeras amplitudes es posible reconstruir la señal de forma que el que escucha no puede percibir si se ha perdido información. El hecho de transmitir sólo algunas amplitudes requiere menos bits que transmitir la forma de onda muestreada.

4.8 Transformación Discreta por Coseno (DCT)

Las transformaciones también se aplican a datos de imágenes de 2 dimensiones. Una transformación espacial de 2 dimensiones es la transformación discreta por cosenos (Feig y Winograd, 1992). Esta transformación tiene la propiedad de que, con imágenes sin discontinuidades pronunciadas, los términos más significativas (de mayor potencia del espectro) son los primeros.
La transformación DCT, en teoría, no tiene pérdidas, pero en la práctica el uso de números de punto flotante y funciones transcendentes, siempre introducen errores de redondeo.

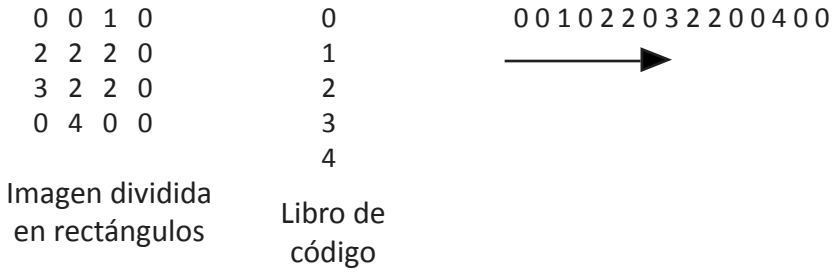
La aplicación DCT en 2 dimensiones en una componente clave, tanto en JPEG

como en MPEG. Genera una matriz numérica del tamaño de un bloque de imagen seleccionado.

La cuantización vectorial
La cuantización vectorial es otro caso de codificación por fuente. Teniendo la imagen dividida en rectángulos de tamaño fijo, se le asocia a cada uno de ellos un código y se construye otra tabla con los códigos, llamada libro de código. Cada rectángulo se transmite buscándolo en el libro de código, enviando simplemente su código. Si el libro de código se crea dinámicamente también se envía.

Nótese que si una pequeña cantidad de rectángulos predomina en toda la imagen, se logra mayor velocidad de transferencia.

En la figura se muestra un ejemplo de cuantización vectorial:



4.9 Estándar JPEG

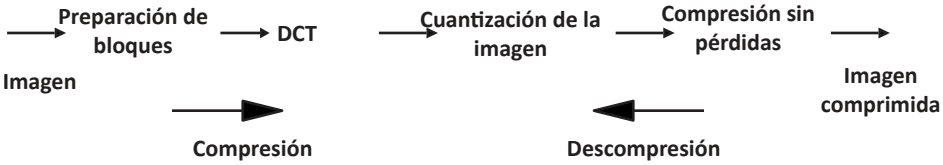
El estándar JPEG (Joint Photografic Experts Group, grupo conjunto de expertos en fotografía) está destinada para la compresión de imágenes fijas de tono continuo.

Para tener idea de la necesidad de comprimir imágenes, supóngase que una imagen digital a color tiene 640x480 pixels y que cada pixel se representa por 24 bits (8 bits por cada componente RGB). Esto requeriría 737 080 bytes. Para transmitir esta imagen sobre un canal de 64 Kbps se necesita más de 92 seg. Con JPEG, una buena calidad de imagen puede lograrse con sólo un bit por pixel, o sea, con una compresión de 24 a 1, lo cual reduce el requerimiento de memoria y a esa velocidad de transmisión se necesitarían 3,8 seg.
Es importante para multimedia porque el estándar multimedia para imágenes en movimiento, MPEG, es la codificación JPEG de cada cuadro por separado, adicionando características para la compresión entre cuadros y la detección de movimiento.

El JPEG tiene varios modos de trabajo, pero sólo se mencionará simplificadamente el modo basado en el DCT secuencial que combina de forma adecuada altas

relaciones de compresión con una buena calidad de la imagen recuperada.

La compresión y descompresión de la imagen tiene 4 etapas principales:



La preparación del bloque consiste de varios procesos con el objetivo final de obtener bloques de imágenes de 8x8 pixels.
Se hace uso de la luminancia y las crominancias y se calculan esos parámetros para cada pixel de la imagen total (Ej: imagen RGB de 640x480 con 24 bits/pixel). Se construyen matrices separadas para cada parámetro y se reducen las matrices de luminancias.

Por último se realizan una operación de normalización de los elementos de las matrices y cada matriz se divide en bloques de 8x8.

Para el caso de una imagen RGB de 640x480 se obtendrán 7 200 bloques de 8x8.
La segunda etapa de **JPEG** es aplicar una transformación DCT a cada uno de los bloques obtenidos en el paso anterior. La salida de cada DCT es una matriz de 8x8 coeficientes cuyos valores normalmente disminuyen a partir de la esquina superior izquierda de la matriz.

Por ejemplo:

Tabla 1. Matriz de coeficientes.

150	80	40	14	42	1	0	
92	75	36	10	6	1	0	0
52	38	26	8	7	4	0	0
12	8	6	4	2	1	0	0
4	3	2	0	0	0	0	0
2	2	1	1	0	0	0	0
1	1	0	0	0	0	0	0
0	0	0	0	0	0	0	0

En la etapa de cuantización se eliminan los coeficientes menos importantes. Esta transformación (con pérdidas) se hace dividiendo cada coeficiente de la matriz DCT entre un peso tomado de una tabla (tabla de cuantización).

Tabla 2. Cuantización.

Coeficientes DCT

150	80	20	4	1	0	0	0
92	75	18	3	1	0	0	0
26	19	13	2	1	0	0	0
3	2	2	1	0	0	0	0
1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

Coeficentes cuantizados

1	1	2	4	8	16	32	64
1	1	2	4	8	16	32	64
2	2	2	4	8	16	32	64
4	4	4	4	8	16	32	64
8	8	8	8	8	16	32	64
16	16	16	16	16	16	32	64
32	32	32	32	32	32	32	64
64	64	64	64	64	64	64	64

Los valores de la tabla de cuantización no son parte del estándar JPEG. Cada aplicación debe dar sus propios valores, permitiendo controlar el equilibrio pérdidas-compresión.

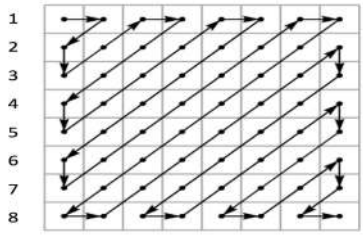
El cuarto paso es la compresión sin pérdidas que involucra 2 pasos preparatorios. Primero, el coeficiente cuantizado de la esquina superior izquierda de la matriz (DC) se reemplaza por el valor de su diferencia (**DIF)** respecto al elemento correspondiente del bloque previo (PREV)

DIF=DC-PREV

Como esos elementos representan los valores promedios de sus bloques respectivos, deben cambiar poco, por lo que al tomarse las diferencias la mayoría de ellos se reducen a valores pequeños. No se calculan las diferencias de los otros valores.

Segundo, los restantes valores de la matriz se ordenan en zig-zag, obteniéndose un arreglo lineal de 64 elementos.

Tabla 3. DIF=DC-PREV.



Como la mayoría de los coeficientes cuantizados, alejados de la esquina superior izquierda, son generalmente ceros, el ordenamiento tiene el efecto de concentrar muchos ceros al final de la lista. En el caso mostrado, la subcadena final puede reducirse a una sola diciendo que hay 38 ceros.

Estos dos pasos preparatorios dan una cadena de codificación por entropía, como por ejemplo, el código de Huffman. La descompresión sigue la secuencia de pasos en sentido inverso.

CAPÍTULO V. NOCIONES DE CRIPTOGRAFIA

5.1 Introducción

En el campo de la Telemática, la creciente tendencia a la interconexión de sistemas y a la integración de servicios, hace más fácil el acceso de un número cada vez mayor de individuos a mensajes intercambiados por los diferentes sistemas y aplicaciones.

En la actualidad miles de instituciones, negocios e individuos están interconectados. Es el fenómeno conocido como la explosión de Internet que ha alterado su composición en el sentido en el sentido de que ahora existe una gran diversidad de usuarios, con diversos grados de pericia, y diversos sistemas operativos. Se reporta que una década atrás, la mayoría de los servidores eran atendidos por personas con al menos conocimientos básicos de seguridad de redes y los ataques ocurrían pocas veces con relación a los posibles. En la actualidad en la comunidad de Internet es mayoritario el personal con débiles conocimientos acerca de la seguridad (el número de usuarios individuales crece), el volumen y valor de la información que transita ha aumentado, y junto a ello la actividad de los crackers.

Las actividades de los hackers y crackers han generado mucho interés público en los últimos tiempos, especialmente en E.U..

La noción de seguridad en sistemas de cómputo es de cierta amplitud. Cuando se habla de cuestiones de seguridad se deben considerar diferentes aspectos:

- **Seguridad física.** Se refiere a la salvaguarda de los soportes físicos de la información. En este nivel están las medidas contra incendios, sobrecargas eléctricas, prevención de sabotajes, salvas, etc.
- **Seguridad del canal de comunicaciones.** Pocas veces se consideran seguros, ya que en la mayoría de los casos escapan al control de los usuarios de la información. Es difícil asegurarse que no están intervenidos.
- **Seguridad de la información.** Se refiere a la preservación de la información frente a observadores no autorizados.
- **Problemas de autenticación.** Debido a la inseguridad típica de los canales de comunicación es necesario asegurarse de que la información que se recibe viene realmente de quien se supone.
- **Problemas de suplantación.** Sobre todo en el caso de redes cualquier usuario puede tener acceso al sistema desde fuera, por lo que se deben tener sistemas fiables que aseguren que los usuarios no estén siendo suplantados por intrusos.

En los más de 25 años de existencia la Internet no demostrado total seguridad para la información. En particular los sitios gubernamentales han sido los blancos principales de los crackers.
Como ejemplo se tiene un reporte de la Government Accounting Office (GAO) de E.U. relativo a la seguridad de las redes de defensa de la nación y que concluye que (1):

- La Defensa pudo haber sido atacada unas 250 000 veces el pasado año (1995).
- De acuerdo con oficiales de la Defensa, los agresores han obtenido, modificado y destruido datos y software. Han desconectado y destruido sistemas completos, obstruyendo el acceso a usuarios. Han sido afectadas funciones de defensa, incluyendo armas e investigaciones de supercomputadoras, logística, finanzas, gestión de personal, etc.
- Information Security: Computer Attacks at Department of Defense Pose Increasing Risks (Chapter Report, 05/22/96, GAO/AIMD-96-84); Chapter 0:3.2, Paragraph 1.

El mismo reporte reveló que, a pesar del cuarto de millón de ataques ocurridos, sólo 1 de 500 fueron reportados realmente.

5.2 La seguridad de datos en los negocios

En el área de los negocios comenzaron a usar criptografía hace más de 20 años, como una vía de añadir seguridad y cualquier negocio que tuviera una computadora empleaba algún de tipo de contraseña o algún control de acceso a la máquina o a ciertos archivos.

Poco a poco, en la medida que la criptografía salió del dominio de organizaciones de seguridad se comenzó a utilizar, lo que algunos llaman, criptografía fuerte como una herramienta más de los negocios.

Muchos prevén que en los medios comerciales de E.U. el uso (y tal vez desarrollo) de técnicas de cifrado se incrementará, atendiendo a las siguientes razones principales:

- Las computadoras han cambiado radicalmente con relación a 2 décadas anteriores. Ya no están centralizadas, ni en lugares cerrados, ni los enlaces externos son escasos, ni éstos exclusivamente por líneas dedicadas.
- Los mensajes y archivos electrónicos ahora se mueven a través de redes inseguras.
- La aparición de los sistemas de clave pública (1977) permite el envío de clave de cifrado del receptor a través de redes inseguras, sin comprometer

las privacidad del mensaje ulterior, resolviéndose unos de los problemas prácticos de los sistemas de cifrado.

5.3 Conceptos básicos de Criptografía

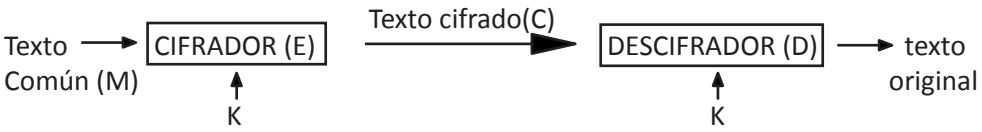
Criptografía es una palabra que proviene del griego y significa “arte de escribir con clave secreta o de un modo enigmático”. Actualmente la criptografía es un grupo de técnicas que tratan sobre la protección de la información.

El cifrado es la transformación de datos o mensajes en alguna forma no legible y su objetivo es mantener la información oculta para aquellos a los cuales no está dirigido. Un mensaje cifrado se le llama **texto cifrado.**

Entre los campos con los que se relaciona están la Teoría de la Información, la Teoría de los Números y la Complejidad Computacional.

La ciencia de generar los mensajes cifrado es la criptografía y la de descifrar los mensajes se conoce como criptoanálisis. Ambas forman la llamada criptología.

A los mensajes originales también se les llama texto claro o común. El descifrado es el proceso inverso del cifrado. Esquemáticamente el proceso cifrado-descifrado es



Criptosistema

Por tanto, un criptosistema se puede definir como una quíntupla (M,C,K,E,D), donde:

- M es el conjunto de todos los mensajes sin cifrar (texto claro).
- C es el conjunto de todos los posibles mensajes cifrados (criptogramas).
- K es el conjunto de claves que se pueden emplear en el criptosistema.
- E es el conjunto de transformaciones de cifrado que se aplica a cada elemento de M para obtener un elemento de C. Existe una función diferente de la familia Ek para cada valor posible de la clave k.
- D es el conjunto de transformaciones de descifrado.

Todo **criptosistema** cumple la siguiente condición:

$$Dk(Ek(M))=M$$

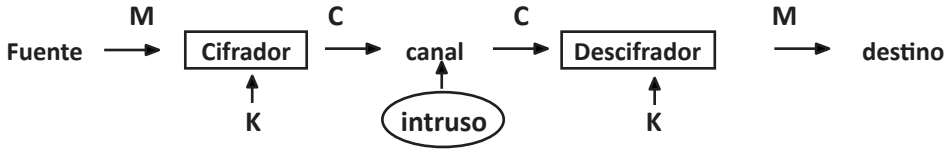
El cifrado y descifrado requiere la utilización de alguna información secreta que, usualmente, se refiere a una clave. Dependiendo del mecanismo de cifrado utilizado,

la misma clave podría usarse para cifrar y descifrar, mientras que otros algoritmos emplean claves diferentes para cada proceso.

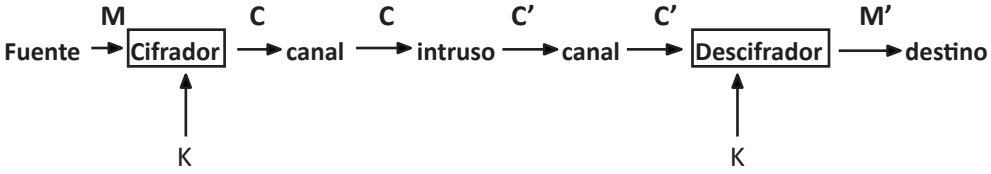
El cifrado es una operación de se puede realizar en diferentes partes de la arquitectura de un sistema de transmisión de datos. Sin embargo, en el modelo que se muestra se omiten todas las operaciones que no están relacionadas con el cifrado. Así, se supone que la fuente emite mensajes M (texto claro o legible), que estos mensajes se transforman por el cifrador en mensajes C (texto cifrado), y que de este modo llegan al receptor.

La característica que debe reunir un sistema de seguridad depende del entorno donde éste se va a aplicar y de los posibles peligros a los que se expondrá la transmisión y almacenamiento de datos. Los riesgos, a los que aquí se hace referencia, proceden de la presencia de usuarios no autorizados (intrusos), el cual puede actuar de 2 modos distintos: sólo leyendo la información o modificándola.

En el primer caso se habla de intruso pasivo



En el segundo caso se habla de un intruso activo



Los requisitos de un sistema de cifrado dependerán de si se supone o no la presencia de intrusos activos o no.

En aquellos casos en los que sólo se considera la existencia de intrusos pasivos, los posibles riesgos son:

- Análisis de los datos: si dispone de información suficiente para realizar el criptoanálisis del cifrado C, puede tener acceso al contenido de M.
- Análisis de tráfico: haciendo un análisis estadístico, puede obtener cierta información, como por ejemplo presencia o ausencia de determinado usuario, frecuencia de uso de determinados servicios, etc.

Si, además, se tiene en cuenta la posible existencia de intrusos activos, se pueden considerar los siguientes riesgos:

- **Manipulación de los datos:** si el intruso es capaz de descifrar los datos, también podría cambiarlos, volverlos a cifrar y retransmitirlos a su destino.
- **Suplantación de usuarios**: preparar mensajes falsos y enviarlos haciéndose pasar por un determinado usuario. También es posible suplantarlo para acceder a servicios que no le corresponden.
- **Reactuación:** grabar un mensaje cifrado para repetirlo después haciendo creer al receptor que dicho mensaje se ha enviado varias veces. El mensaje podría contener la orden de realizar una acción, con lo que esa acción se repetirá varias veces.

Existen otros riesgos que no provienen de intrusos, sino de usuarios autorizados, como es la negación por parte de un usuario de ser el originador de un determinado mensaje (Repudio).

Los anteriores riesgos de seguridad pueden ser minimizados o eliminados con una o varias alternativas que ofrece la criptografía y de la codificación:

- **Confidencialidad.** Es el servicio básico de un sistema de seguridad. Proporciona secreto a los datos y protege contra el análisis y manipulación de los datos. El mecanismo básico para este servicio es el cifrado de datos. A veces se emplea también tráfico de relleno para evitar el análisis de tráfico por un intruso.
- **Integridad.** Proporciona la seguridad de que los datos no han sido alterados durante la comunicación o almacenamiento. Protege contra la manipulación de los datos. El mecanismo básico utilizado para este servicio es el cifrado de los datos, comprimidos en su totalidad o no.
- **Autenticación.** Asegura que un usuario, al momento de utilizar el sistema, es el que se anuncia como tal. Este mecanismo protege contra la suplantación de usuarios y puede ser el intercambio de nombre y contraseña entre emisor y receptor, o la firma digital. La firma digital consiste en que el usuario firmante cifre determinada cadena de modo que ningún otro usuario lo pueda suplantar. Para comprobar la firma, ésta puede ser descifrada por cualquier usuario con la información correspondiente al usuario firmante. Si la firma contiene la fecha y la hora de la transmisión, se obtiene protección contra la reactuación.
- **Control de acceso.** Previene contra el uso no autorizado de recursos, como puede ser el acceso a una torre de disco compartido. Protege contra la manipulación de datos. Los mecanismo utilizados pueden ser los mismos que para la autenticación.
- **No repudio.** Asegura la integridad y la identidad del originador de los

datos, de modo que el mensaje recibido constituya una prueba de ambas cosas. Protege contra el repudio. El mecanismo utilizado es la firma digital, donde la cadena que se firma se obtiene del mensaje enviado, a partir de una función determinada.

La criptografía moderna toma como base problemas que son difíciles de resolver, es decir, estudia las técnicas y aplicaciones que dependen de la existencia de problemas difíciles.

5.4 Generalidades sobre los algoritmos criptográficos

El texto común M puede ser una secuencia de bits, un fichero texto, una imagen, voz digitalizada, etc. Con relación a la computadora, **M** es simplemente un dato binario. Además, M puede intentarse transmitir o almacenar y, en cualquier caso, es un mensaje a ser cifrado.

Si se denota por **C** el texto cifrado, se puede considerar que el proceso de cifrado es una función que opera sobre **M** y da C.

E (M) = C **E: cifrador**

En el proceso inverso, la función de descifrado D opera sobre C para producir M

D (C) = M **D: descifrador**

De manera general se cumple la siguiente identidad

D (E (M)) = M

Los algoritmos criptográficos son las funciones **E** y **D.**

Si la seguridad de un algoritmo se basa en mantener en secreto la forma en que trabaja el algoritmo se les llama algoritmo restringido. Ellos tienen interés histórico, pero se consideran como regla inadecuados con relación a los estándares actuales, ya que un grupo grande o cambiante de usuarios no podrá usarlos al tener que cambiar de algoritmo cada vez que un usuario abandone el grupo. Además, si alguien accidentalmente revela el secreto, cada uno debe cambiar su algoritmo.

Otra razón es que los algoritmos restringidos no permiten un control de calidad. Cada grupo debe escribir su propio algoritmo e implementarlo y, si ninguno del grupo es buen criptógrafo, no se podría saber si tienen un algoritmo seguro. A pesar de eso los algoritmos restringidos son uy populares es aplicaciones de baja seguridad.

La criptografía moderna resuelve este problema con una clave K. La clave podría ser una seleccionada entre un gran número de valores. El rango de posibles valores de la clave se llama espacio de claves.

Si, tanto el cifrador como el descifrador usan esa clave, las funciones se denotan como

Ek (M) = C **Dk (C) = M**

Algunos algoritmos, como se verá, usan diferentes claves para cifrar y descifrar. Es decir

Ek1 (M) = C Dk2 (C) = M

Toda la seguridad de estos algoritmos se basa en la(s) clave(s) y no en los detalles del algoritmo. Esto significa que el algoritmo puede publicarse y analizarse, y es claro que los productos que usen tales algoritmos pueden producirse masivamente.

Hay 2 tipos generales de algoritmos basados en clave: los simétricos y los de clave pública.

5.4.1 Algoritmos simétricos

Son algoritmos donde la clave de cifrado y descifrado es la misma y se llaman también de clave secreta o de clave simple. Requieren que el transmisor y el receptor acuerden la clave antes de comunicarse con seguridad. La seguridad aquí descansa en la clave.

Los algoritmos simétricos pueden dividirse en 2 categorías:

- **Cifradores de secuencias (stream ciphers):** operan sobre el texto común a nivel de un bit o byte a la vez.
- **Cifradores de bloques (block ciphers):** operan sobre el texto común a nivel de grupos de bits o bytes.

5.4.2 Algoritmos de clave pública

Llamados también asimétricos, se diseñan de tal manera que la clave usada para cifrar sea diferente de la clave usada para descifrar. Se llaman de clave pública porque la clave de cifrado puede ser pública. Cualquier persona puede usar la clave de cifrado para cifrar un mensaje, pero sólo la persona con la correspondiente clave de descifrado puede descifrarlo (clave privada).

El cifrado, usando la clave pública k1 , se denota por

$$Ek1\ (M) = C$$

y el descifrado

$$Ek2\ (C) = M$$

5.5 Seguridad de un sistema criptográfico

Para conocer el grado de seguridad que proporciona cada algoritmo conviene conocer la dificultad que encontraría un supuesto intruso que, sin conocer la clave, quisiera realizar la operación inversa al cifrado. Para ello, el intruso utilizaría una serie de técnicas que constituyen lo que se conoce como criptoanálisis.

El **critpoanálisis** consiste en descifrar un mensaje sin conocer la clave, o bien obteniendo la clave a partir de uno a más criptogramas. No se considera criptoanálisis el descubrimiento de un algoritmo secreto de cifrado. Se supone que los algoritmos son siempre conocidos.

Los intentos **criptoanalíticos** se llaman ataques. El presupuesto fundamental en criptoanálisis, enunciado por D. A. Kerchhoffs en el siglo XIX, es que el secreto descansa completamente en la clave. Es decir, se supone que se tienen los detalles del algoritmo criptográfico y su implementación. En la realidad no siempre el criptoanalista tiene toda esa información, pero es una buena suposición.

En general el **critpoanálisis** se ejecuta estudiando grandes cantidades de pares mensajes-criptograma generados por la misma clave.

Las condiciones en las que se realiza el ataque puede ser, en orden decreciente de dificultad, las siguientes:

1. **Ataque de texto cifrado**
   El criptoanalista tiene varios mensajes que han sido cifrados utilizando el mismo algoritmo. La tarea del criptoanalista es recuperar el texto común de tantos mensajes como pueda, o mejor aún, deducir la clave usada para cifrar, con el objetivo de descifrar otros mensajes con la misma clave.

2. **Ataque de texto claro conocido**
   Se tiene acceso no sólo el texto cifrado de varios mensajes, sino también el texto claro de esos mensajes. La tarea consiste en deducir la(s) clave(s).

3. **Ataque de texto claro seleccionado**
   No solo se tiene acceso el texto cifrado y al texto común asociado para varios mensajes, sino también escoge el texto común a ser cifrado que podría aportar

información acerca de la clave. La tarea es deducir la(s) clave(s) usada(s) para cifrar o un algoritmo para descrifrar cualquier nuevo mensaje cifrado con esa(s) clave(s).

4. **Análisis exhaustivo (fuerza bruta)**
   Se intenta descifrar el texto cifrado con todas las claves del espacio de claves, una a una, y examinar el texto obtenido. Cuando ese texto tiene sentido (texto válido), se considera que se ha encontrado la clave. Este método encontrará inevitablemente la solución, suponiendo que se disponga de la capacidad de cálculo y tiempo suficiente y si en todo el espacio de claves sólo hay una que da lugar a un texto con sentido.
5. **Análisis estadístico**
   Algunos métodos pueden tener ciertas propiedades estadísticas que relacionan el texto cifrado con el legible y la clave. Estas propiedades estadísticas se conocen como regularidades del método y permiten un ataque mucho más eficiente que el análisis exhaustivo.

Existen por lo menos otros 3 tipos de ataques.

Existen varios tipos o grados de seguridad:

- **Seguridad incondicional.** Es el caso en que se puede demostrar teóricamente que no existe ninguna posibilidad de romper el sistema, ni aun pudiéndose realizar todos los tipos de ataques y contando con una capacidad de cálculo ilimitada. Esta característica se enuncia diciendo que el texto cifrado no aporta información sobre el texto común.
  Se puede probar que en este caso la entropía o incertidumbre de la clave, conocido el texto cifrado, es al menos la incertidumbre del texto claro.

$$H(K/C) \geq H(M)$$

Es decir, existen al menos tantas claves como mensajes posibles y no hay modo de saber cuál es la que se utilizó.

- **Seguridad teórica.** Cuando no hay diferentes claves, con sus textos claros asociados, que produzcan el mismo texto cifrado. Si así fuera no se podría saber cuál de ellos es el mensaje enviado. Por tanto, si no se podrá llegar a un resultado concluyente.

$$H(K/C) >> 0$$

- **Seguridad computacional.** La criptografía tiene más relación con los criptosistemas que no son computacionalmente factibles de romper. Si el sistema no es teóricamente seguro, pero los medios necesarios para

realizar el criptoanálisis son tan elevados que lo hacen inviable, se puede decir que el sistema es seguro prácticamente.

La complejidad de un ataque se puede medir por diferentes maneras:

- **Complejidad de los datos.** La cantidad de datos que se necesita conocer para realizar el ataque.
- **Complejidad del procesamiento:** el tiempo que se necesita para realizar el ataque (factor de trabajo).
- **Requerimientos de almacenamiento**. La cantidad de memoria necesaria para efectuar el ataque.

Por ejemplo, si se requieren 2^{128} operaciones para romper el algoritmo, se supone que se tiene suficiente velocidad de cálculo para realizar 1 millón de operaciones por segundo y se puede colocar 1 millón de procesadores para la tarea, entonces se puede afirmar que se requerirían 1019 años para recuperar la clave haciendo un análisis exhaustivo. Eso es 109 veces la edad del universo (1 año $\approx 2^{25}$ segs).

5.6 Generalidades sobre los algoritmos DES, IDEA y AES

Se caracterizarán los algoritmos simétricos más importantes.

DES

Es un algoritmo de cifrado normalizado, bastante extendido en la actualidad. Después de ser evaluado públicamente, DES se adoptó como estándar federal en 1976 y publicado en 1977 oficialmente. Otros detalles y posibilidades se publicaron en los años siguientes.

DES se fundamenta en la aplicación sucesiva de varios métodos tradicionales. A pesar de la debilidad de estos métodos por separado, su uso conjunto puede dar lugar a un sistema suficientemente complejo para su uso en un gran número de aplicaciones. Los métodos clásicos o tradicionales utilizados en DES son de 2 tipos: sustituciones (sustitución de los símbolos del alfabeto del texto común por otro símbolo del mismo alfabeto, según cierto patrón) y permutaciones (reordenamiento de los símbolos del texto claro).

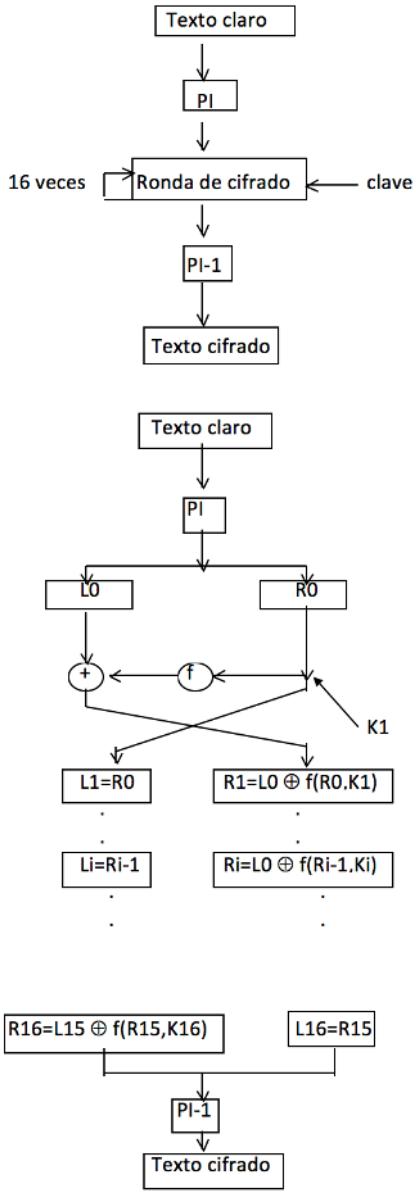


Figura 17: Descripción del algoritmo.

DES opera sobre un bloque de 64 bits. Inicialmente se realiza una permutación inicial que transpone los bits del bloque de entrada de acuerdo con una tabla y el resultado se almacena en 2 registros de 32 bits cada uno. (L0 y R0).
Entonces se realizan 16 rondas con las mismas operaciones, llamadas función de cifrado f.

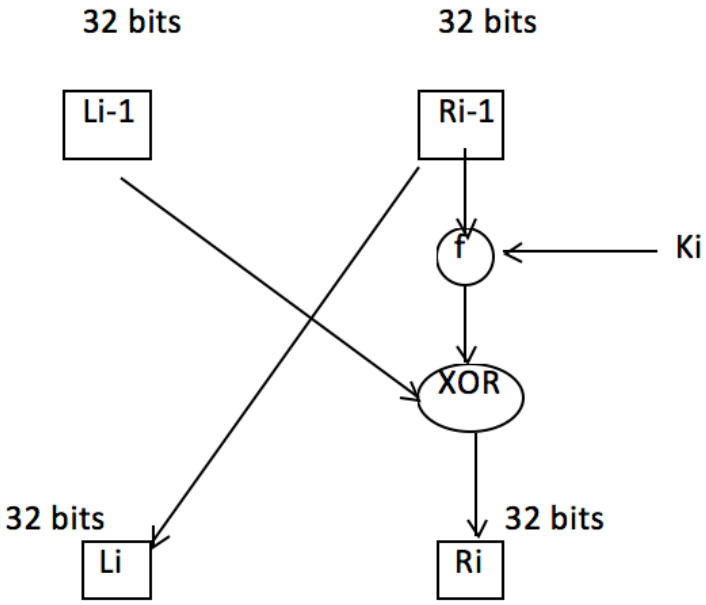
Finalmente se unen las dos mitades y se realiza una permutación final, inversa a la inicial.

En cada ronda los bits de la clave se desplazan, después de una permutación para eliminar los bits de paridad y se almacenan en 2 registros de 28 bits. Entonces se seleccionan 48 bits de ella mediante una permutación de compresión.

Todas las rondas se repiten 16 veces

La permutación final es la inversa de la inicial, después de agrupar las 2 mitades resultantes. Esto es el cifrado.

El algoritmo DES realiza procesos para eliminar la mayoría de las regularidades. El resultado es que es muy difícil encontrar correlación entre el texto legible y el texto cifrado, ni entre el texto cifrado y la clave. Se ha comprobado mediante mediciones que cuando cambia un bit en M, después de 5 o 6 iteraciones ya se han cambiado aproximadamente la mayoría de las regularidades del texto.



Operaciones en cada ronda de cifrado.

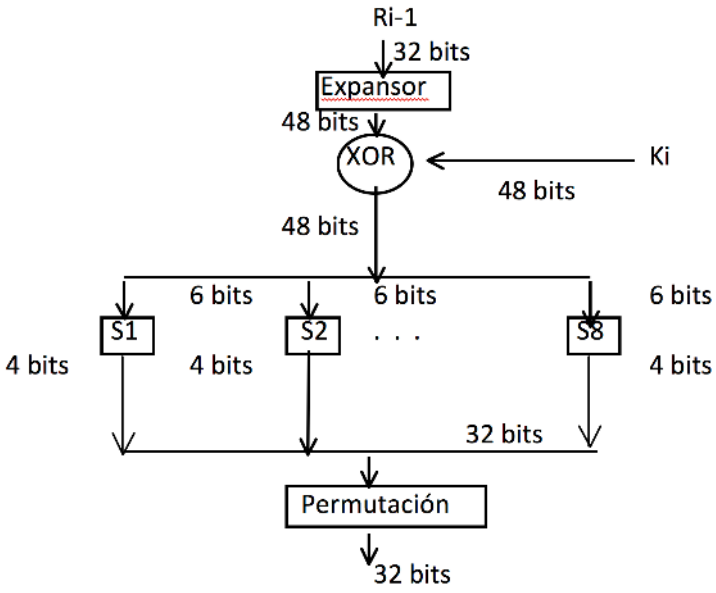


Figura 18: Función de cifrado f.

Descifrado en DES

Las diferentes operaciones fueron escogidas para lograr la propiedad de que el mismo algoritmo trabaja para el cifrado y para el descifrado. La única diferencia es que las claves deben emplearse en orden inverso. Es decir, si las claves de cifrado para cada ronda son K1, K2,…, K16, entonces las claves de descifrado son K16, K15,…, K1.

Seguridad del DES

En el año 1998 se demostró que un ataque de fuerza bruta era viable debido a la insuficiente longitud de la clave.

Otro aspecto a señalar es la forma en que la clave inicial se modifica para obtener las subclaves para cada ronda del algoritmo y hay valores de llave que conducen a una secuencia inadecuada de Ki. Existen 4 claves, llamadas claves débiles, que tiene la propiedad de que cifrando un texto dos veces con una de ellas se obtiene de nuevo el texto original, es decir, cumplen que

$$Ek(Ek(M)=M$$

Adicionalmente existen 6 parejas de claves, llamadas semidébiles, que cifran texto común a idénticos textos cifrados. En otras palabras, cumplen la propiedad de que se puede cifrar con una de ellas y descifrar con la otra.

De otro lado, si se complementan el texto común y la clave (se cambian 1s por 0s y viceversa), el resultado es el texto cifrado complementado:

$$C = T(M,K) \Rightarrow C = T(M,K)$$

Esto significa que en un ataque de texto claro seleccionado solamente se tienen que probar la mitad de las posibles claves: 255.

Básicamente se dice que el problema real del DES no radica en su diseño, sino en la longitud de su clave (56 bits), lo que hace que el ataque más prometedor sea el exhaustivo, probando con todas las claves del conjunto de claves posibles. El constante aumento de la velocidad de los circuitos integrados y la posibilidad de realizar la búsqueda con procesamiento paralelo pueden llegar a poner en duda la seguridad del DES.

IDEA

IDEA (International Data Encryption Algorithm, 1992) es la segunda versión de un cifrador de bloques y es considerado como uno de los mejores algoritmos de bloques disponibles al público en la actualidad. Se considera resistente a muchos ataques, entre ellos el criptoanálisis diferencial. Prácticamente no presente claves débiles.

Algunas de sus características se enumeran a continuación:

- Es un cifrador de bloques que opera con 64 bits de texto común.

Los 64 bits del bloque de datos se dividen en 4 sub-bloques de 16 bits que son las entradas a las rondas de cifrado.

- Utiliza una clave de 128 bits de longitud y 8 rondas de cifrado.

Con la clave se generan 52 subclaves de 16 bits, 6 por cada ronda de cifrado y 4 para la transformación de salida.

- La filosofía de diseño es la mezcla de 3 operaciones incompatibles que pueden ser implementadas en hardware y en software:
  1. **XOR**
  2. **Adición módulo 216**
  3. **Multiplicación módulo 216+1**

- Al utilizar una clave de 128 bits, un ataque de fuerza bruta requeriría 2128 (1038) pruebas para recuperar la clave.

El diseño de un chip que puede probar 109 claves por segundo, y disponiendo de 109 chips de ese tipo trabajando simultáneamente, requeriría 1013 años. Un array de 1024 de esos chips permitiría encontrar

la clave en un día, pero no hay suficientes átomos de silicio en el universo para construir esa máquina.

- Se considera inmune al criptoanálisis diferencial(E. Biham y A. Shamir, 1990).

Este tipo de ataque fue utilizado para atacar al DES y es básicamente de texto común seleccionado. Descansa en el análisis de la evolución de las diferencias, a través de las rondas de cifrado, entre dos textos comunes afines que son cifrados con la misma clave. Un cuidadoso análisis de los datos disponibles permite asignar probabilidades a cada una de las posibles claves y, eventualmente, identificar la clave más probable como la correcta. Ha sido utilizado contra muchos sistemas de cifrado con diferentes grados de éxito.

- Las implementaciones en software de IDEA son alrededor de 2 veces más rápidas que las similares de DES.

AES

El 2 de Octubre de 2000, el NIST hizo público el algoritmo ganador de la convocatoria AES (estándar de cifrado avanzado), que sustituirá al actual DES (estándar de cifrado de datos) hasta bien entrado el siglo 21. RIJNDAEL es un algoritmo belga, venciendo a algoritmos norteamericanos y a criptólogos de reputada fama mundial.

AES ha sido un concurso abierto a cualquier participante mundial que cumpliese con una serie de requisitos, como velocidad tanto en software como en hardware, reducida ocupación de memoria, implementación eficiente tanto en procesadores modernos como en CPUs de 8 bits (utilizadas, por ejemplo, en tarjetas inteligentes), etc. El ganador debe, por otra parte, liberar cualquier tipo de patente en el algoritmo, para potenciar así el desarrollo de soluciones tecnológicas de buena calidad.

El tamaño de clave debía ser de, al menos, 128, 192 y 256 bits (debe admitir los tres), y el tamaño de bloque de cifrado (se trata de un sistema de cifrado en bloque, no un sistema de cifrado en flujo) debía ser de 128 bits. Esto se compara muy favorablemente con el DES, de 56 bits de clave y 64 bits de tamaño de bloque.

Los motivos para seleccionar a “RIJNDAEL”, según el web del NIST son su buena combinación de seguridad, velocidad, eficiencia (en memoria y puertas lógicas), sencillez y flexibilidad.

En **AES** el resultado intermedio del cifrado se denomina Estado, que puede representarse como un array rectangular de bytes de cuatro filas. La transformación que tiene lugar en cada vuelta de cifrado está compuesta a su vez de cuatro transformaciones diferentes:

- Sustitución de bytes no lineal, operando independientemente sobre cada uno de los bytes del Estado.
- Desplazamiento de las filas del Estado cíclicamente con offsets (desplazamientos) diferentes.
- Mezcla de columnas, que se realiza multiplicando las columnas del Estado módulo x4+1, consideradas como polinomios en GF(28), por un polinomio fijo c(x).
- Adición de la clave de vuelta, en la que se aplica al Estado por medio de un simple XOR. La clave de cada vuelta se deriva de la clave de cifrado mediante el esquema de clave.

El esquema de clave consiste en dos operaciones: expansión de clave y selección de clave de vuelta de cifrado.

El proceso de cifrado consta de tres pasos: una adición inicial de la clave de vuelta, n-1 vueltas de cifrado y una vuelta final.

Todas las aplicaciones y comunicaciones, especialmente bancarias, que hacían uso del DES, serán actualizadas paulatinamente a Rijndael, de manera que antes de la primavera de 2001 se espera que de DES no quede rastro más que en los libros de texto. Como el algoritmo es público y no necesita ser licenciado, cualquier otro país distinto de los EEUU podrá adoptarlo igualmente como estándar en sus propias aplicaciones si así lo desea.

RSA (Rivest, Shamir y Adelman)

Poco después del algoritmo de mochila de Merkle, fue reportado un algoritmo de cifrado de clave pública y firma digitales (1978), llamado RSA.

La fortaleza del algoritmo descansa en la seguridad computacional que representa la dificultad de factorizar números grandes. Las claves públicas y privadas son funciones de una pareja de números primos grandes (de 100 a 200 dígitos o incluso mayor). La recuperación del texto claro a partir de la clave pública y del texto cifrado, se conjetura, es equivalente a factorizar el producto de 2 números primos.

La factorización de números primos requiere dividir un entero en factores que sean números primos y cualquier entero tiene una factorización prima única. La multiplicación de 2 números primos es fácil, pero la factorización es un problema difícil y, aún cuando en rigor esto no ha sido probado, los investigadores consideran remota la posibilidad de encontrar un método fácil de lograrlo.

En los últimos 15 años la solución de ese problema se ha facilitado algo debido al desarrollo de hardware más potente y el desarrollo de mejores algoritmos de factorización. Este es un campo muy activo de investigación entre matemáticos y especialistas de la computación. No obstante, el RSA siempre tiene la posibilidad de aumentar el tamaño de número y alejarle al intruso la posibilidad de explotar mejor hardware y algoritmos.

En resumen, el cifrado RSA es así:

Clave pública:

n: producto de 2 números primos p y q (p y q permanecen en secreto)
e: primo relativo a (p-1)(q-1)

Clave privada:

$d = e\text{-}1 \bmod ((p-1)(q-1))$ $\quad e.d \geq 1 \bmod [(p-1)(q-1)]$

Los números e y n son la clave pública y el número d es la clave privada. Los números p y q pueden descartarse pero no revelarse.

Cifrado

$\mathbf{c} = me \bmod \mathbf{n}$

El mensaje cifrado C será construido de bloques ci de tamaños similares a mi.

Descifrado

$\mathbf{m} = cd \bmod \mathbf{n}$

Ejemplo:

Sean p=47 y q=71, números primos

Entonces

$$\mathbf{n = p.q = 3337}$$

La clave de cifrado e, no debe tener factores en común con el resultado

$$\mathbf{(p-1)(q-1) = (46)*(70) = 3220}$$

Escogiendo al azar e=79, que es primo relativo a 3220

d=e-1 mod [(p-1)(q-1)] = 79-1 mod 3220

79 d = k(3220) + 1

El valor de d pudiera ser calculado por tanteo, dando valores a k, hasta k=25. También utilizando el algoritmo de Euclides extendido se llega a que

d= 1019

Los número n= 3337 y e = 79 son públicos y el valor de d= 1019 es secreto. Los valores p y q se descartan.

Para cifrar el mensaje:

M = 6882326879666683

se descompone en pequeños bloques de 3 dígitos

m1= 688	**m4= 966**
m2= 232	**m5= 668**
m3= 687	**m6= 003**

El primer bloque se cifra como

c1= m1e mod n = 68879 mod 3337 = 1570

Realizando la misma operación en los siguientes bloques se genera el mensaje cifrado

C = 1570 2756 2091 2276 2423 158

Para descifrar el mensaje se requiere efectuar la misma operación, utilizando la clave de descifrado d= 1019

m1= c1d mod n = 15791019 mod 3337 = 688

y así para el resto del mensaje.

Hay varios chips que realizan el cifrado RSA. El alrededor de 1000 veces más lento que el DES, y en software unas 100 veces más lento que el DES (Tecnología de 1995). Se supone que esas cifras se modifiquen ligeramente con los cambios de tecnología.

Seguridad del RSA

Depende totalmente del problema de la factorización de números grandes. Aunque no ha sido probado matemáticamente que se necesita factorizar **n** para calcular mi de ci y e. Existe la hipótesis de que una forma diferente de criptoanálisis puede descubrirse para el RSA. No obstante, la factorización de **n** en la forma más clara de ataque.

5.7 Firmas digitales

Además de asegurar la privacidad, la criptografía se utiliza para verificar autenticidad. Los sistemas de clave pública ofrecen una solución elegante y simple para crear firmas digitales. La noción de funciones unidireccionales, en que se basan los sistemas de clave pública, tienen la propiedad.

D(E(M)) =M

Si la función usada por estos criptosistemas también tiene la propiedad

E(D(M)) = M

se dice que es una permutación unidireccional.

El criptosistema RSA tiene estos requerimientos. El usuario A podría cifrar un mensaje (la firma) usando su clave privada, que es la de descifrado. Sea la notación para esta operación la siguiente

DA (s)

Cualquiera que reciba DA puede descifrarlo usando la clave pública de A.

EA(DA(s)) = s

Como DA se conoce sólo por A, el receptor sabe que A es el autor.

Se sabe que en RSA, un usuario tiene una clave pública (e,n) y una clave privada d. Entonces el usuario cifra la firma con d :

CS = Sd mod n

Cualquiera puede verificar esa firma usando la clave pública

CSe mod n = Sde mod n = S

Si lo que se quiere es enviar un mensaje privado que puede autenticarse por alguien (sea A que envía a B), entonces:

a) Se cifra DA(M) con la clave pública de B, o sea

EB (DA(M))

b) B usa su clave privada para descifrar

DB (EB(DA(M))) = DA(M)

c) DA(M) puede guardarse como prueba de autenticidad y entonces descifrar DA(M) usando la clave pública de A

EA (DA(M)) = M

Así, se lograría privacidad y autenticidad.

CAPÍTULO VI. LOS PROTOCOLOS DE COMUNICACIÓN

6.1 Introducción

Cuando se pretende comunicar un sistema informático con otro a través de una red de comunicaciones, es necesario que exista un conjunto de elementos físicos y lógicos que permitan dicha comunicación. Debemos entender por comunicaci6n no la conexión entre los equipos, sino todo el conjunto de elementos que permiten el entendimiento entre ambos, con independencia de sus características individuales. En las primeras computadoras, las conexiones existentes entre dispositivos se limitaban a 1os periféricos propios de1 sistema. A medida que la informática fue tomando auge y comenzaron a utilizarse las conexiones a larga distancia; éstas en principio también se limitaban a equipos de un mismo sistema e incluso de un mismo fabricante. En la actualidad la tecnología permite la conexión de equipos (computadoras, terminales, computadoras personales, etc.) de distinta arquitectura, distinta naturaleza y por supuesto de distintos fabricantes.

Con estas posibilidades se han propiciado profundas mutaciones o cambios en la forma de desarrollar la informática en las empresas apareciendo conceptos tales como sistemas heterogéneos, sistemas distribuidos y sistemas abiertos.

Si ya desde el principio de la teleinformática era necesaria la presencia de un conjunto de elementos físicos y 1ógicos para hacer posible la transmisión de 1os datos, con mayor motivo es necesaria en el estado actual donde el movimiento de informaciones entre equipos a distancia es pare consustancial con la forma de utilizar la informática.

6.2 Concepto de protocolo

El estado actual de la conectividad entre equipos de tan distinta naturaleza hace necesario el estudio de 1os elementos que coordinan las conexiones y transmisiones, por niveles bien definidos y separados de tal forma que el conjunto de todos ellos engloba todos 1os aspectos que pueden presentarse.

Un protocolo es un conjunto de normas que permiten el intercambio de información entre dos dispositivos o elementos de un mismo nivel.

No sólo permiten la comunicación sino que articulan métodos y procesos para la detección y corrección de errores.

Los niveles en una comunicación

Para establecer una comunicación entre dos sistemas como ya hemos dicho es necesario considerar un conjunto de elementos físicos y lógicos capaces de conseguir un total entendimiento entre ambos. Ante esto surge la necesidad de estructurar de algún modo este conjunto de elementos. Una de las mejores formas

de hacerlo es la modularización consistente en dividir el conjunto en subconjuntos más fáciles de entender y de manejar. Estos subconjuntos pueden ser desarrollados de forma independiente, pudiendo ser sustituidos por otros cuando las condiciones varían debido a nuevos avances tecnológicos o simplemente por ser cambios de tipo de aplicación.
Tomaremos a modo de ejemplo la comunicación entre dos personas utilizando lenguaje oral. Podemos distinguir tres niveles de comunicación:

- **Nivel de razonamiento**. Trata de la comprensión del mensaje o la idea transmitida de una persona a la otra.
- **Nivel de lenguaje**. Trata las reglas sintácticas y simétricas que deben ser utilizadas para transmitir las ideas.
- **Nivel de transmisión.** Se refiere al medio físico utilizado para la transferencia de las palabras de una persona a otra.

Estos niveles son independientes entre sí, de tal forma que el primer nivel comprende el conocimiento de las cosas, mientras que el Segundo podríamos decir que se dedica al estudio de la lengua utilizada (castellano, inglés, etc.) y el último, a la forma de transmitir las palabras (oral, escrita, etc.).

En el caso de la comunicación entre entidades de un sistema Teleinformático que pueden ser el propio sistema central, un Terminal o cualquier unidad de almacenamiento de datos, se trata de proporcionar un transporte fiable teniendo en cuenta que son muchos 1os factores que intervienen. Estos factores son 1os siguientes:

- **El lenguaje utilizado**. Se compone del código en que se presentan 1os datos y en algunos cases de funciones de traducción a otros códigos.
- **Normas para el diálogo.** Se refiere a las normas que se han de establecer para controlar el flujo de datos, turnos de intervención y turnos de espera.
- **Control de la transmisión de los datos.** Comprende todos 1os aspectos relativos a la comunicación entre 1os sistemas, en cuanto a conexión y movimiento de 1os datos.

Toda esta problemática general de un sistema Teleinformático se debe tratar de forma estructurada dividiendo sus funciones en niveles que de forma individual sean fácilmente controlables y en conjunto resuelvan satisfactoriamente las necesidades de comunicación.

Las redes que existen actualmente para la comunicación de datos se organizan en un conjunto de capas o niveles cuyo objetivo es el de simplificar su estudio y desarrollo. Cada capa o nivel se desarrolla sobre la anterior, de tal forma que, recibe una serie de servicios de ella sin conocer 1os detalles de c6mo se realizan dichos servicios.

El número de capas o niveles puede variar de una red a otra. Entendiéndose que todas las funciones que deba realizar la red estarán incluidas en alguna de sus capas.

En la realización de una transmisión de datos entre dos equipos, cada nivel de una maquina conversa con su equivalente en la otra. El conjunto de reglas que regulan la comunicación entre dos equipos a un determinado nivel H. se denomina protocolo de nivel H.

Entre cada dos niveles adyacentes existe una interfaz que define los servicios que cada nivel inferior ofrece al nivel superior. Estos servicios que ofrece un nivel al siguiente dependen de1 numero de niveles existentes en la red (menos niveles = mas servicios).

En general, al conjunto de niveles con sus servicios y protocolos existentes en una red se le denomina **Arquitectura de la Red.**

En una red teleinformática lo que realmente pretendemos es comunicar aplicaciones que se ejecutan en distintos sistemas. Por ejemplo enviar un archive de un sistema a otro, transmitir un mensaje de un usuario a otro. O distribuir funciones entre una computadora personal y una computadora central. El nivel superior de una arquitectura estructurada proporciona 1os servicios necesarios para la comunicación entre aplicaciones y se denomina por ello, **Nivel de Aplicación.**

Cuando se intercambian datos entre aplicaciones, es necesario presentarlas con un determinado formato, por ejemplo, el formato de un documento, la estructura de un archive o el formato de salida de una impresora. Para descargar a 1os usuarios y a 1os programadores de aplicaciones de la tediosa tarea de programar 1os formatos de presentación de 1os datos, se utiliza un nivel denominado Nivel de Presentación. El nivel de aplicación le dice al nivel de presentación mediante unos ciertos parámetros como se desea el formato de 1os datos y el nivel de presentación se encarga de proporcionar este servicio.

Adicionalmente, en la comunicación entre dos sistemas es necesario un elemento moderador capaz de coordinar y controlar el intercambio de 1os datos. Controla la integridad y el flujo de 1os datos en ambos sentidos. Este nivel es el denominado **Nivel de Sesión.**

La misión de identificar el sistema al que se dirigen 1os datos que desea transmitir otro sistema es cometido de1 denominado Nivel de Transporte es decir, este nivel establece el camino 1ógico de los datos (de quién a quien se dirigen 1os datos).
Definido el camino lógico de los datos se hace necesario el establecimiento de un camino real o ruta de datos que permitan la transmisión de los mismos a través de 1os nodos de la red. El establecimiento de esta ruta se realiza en el denominado **Nivel de Red.**

El camino real puede estar compuesto de elementos de distinta naturaleza. Parte del camino puede ser una línea telefónica, otra parte puede ser un radio enlace. etc. El siguiente nivel, denominado Nivel de enlace, se ocupa de que los mensajes lleguen de un nodo a otro de la red controlando que los datos se transmitan correcta

y eficazmente por el enlace así como la posible aparición de errores y estableciendo el correspondiente proceso de recuperación.

Por último el acceso al medio físico por el que se va a establecer una comunicación corresponde al nivel denominado **Nivel Físico**.

Como resumen de estos niveles que se establecen en una comunicación podemos decir que 1os niveles proporcionan respuesta a las siguientes cuestiones:
¿Qué: se desea hacer?.....Aplicación
¿Cómo me entenderá el otro proceso?.....Presentación
¿Con quién y cómo se establece la comunicación?.....Sesión
¿Dónde está el otro proceso?.....Transporte
¿Por qué ruta se llega allí?.....Red
¿Cómo ir a través de esa ruta?.....Enlace
¿Cómo se puede conectar al medio físico?.....Físico

6.3 Modelo de referencia OSI

El Modelo de Arquitectura de Interconexión de Sistemas Abiertos (OSI) publicado en 1983 por la Organización Internacional de Estándares (ISO), tiene como objetivo la normalización en las redes teleinformáticas abiertas, es decir, aquellas en las que se pueden interconectar terminales y equipos de distintas organizaciones y naturalezas.
Las diferentes funciones de que consta la arquitectura OS1 han sido estructuradas en 1os siete niveles, siendo las funciones asignadas a cada uno de ellos complementarias.

Transmisión de datos en OSI
La transmisión de datos desde un equipo emisor a otro equipo receptor a través de la arquitectura OS1 se realiza de la siguiente forma: el equipo emisor entrega 1os datos que desea transmitir al nivel de aplicación donde los datos reciben una cabecera y son entregados al nivel de presentación donde, de igual forma, se añade una nueva cabecera entregándose el resultado al nivel de sesi6n. Por su parte, el nivel de sesión añade otra cabecera entregando el resultado al nivel de transporte, repitiéndose este proceso hasta que los datos llegan al nivel físico, desde donde son realmente trasmitidos al equipo receptor a través del medio físico. Al llegar los datos al equipo receptor se van retirando las cabeceras en los distintos niveles de forma inversa a como se pusieron.

6.4 Niveles del modelo de referencia OSI

Nivel de aplicación
Es el nivel superior en el modelo OSI y trata de coordinar y controlar las funciones a realizar por los programas de usuarios permitiéndoles el acceso al entorno 0SI. Los procesos de aplicaciones se comunican entre sí por medio de las entidades de aplicación a las que están asociadas controladas por protocolos de aplicación

utilizando servicios de presentación (de su nivel inferior inmediato).
Se pueden distinguir tres tipos de procesos de aplicación:

- **Procesos del propio sistema**. Son los que ejecutan funciones para controlar y supervisar operaciones de los sistemas conectados a la red de comunicación.
- **Procesos de gestión de las aplicaciones.** Son los encargados de controlar y supervisar las operaciones de los procesos de aplicación.
- **Procesos de aplicación de usuario.** Son los que procesan la información real para 1os usuarios finales.

Nivel de presentación
El nivel de presentación es el encargado de la transferencia de los datos contenida en 1os protocolos de aplicación. En este nivel intervienen los aspectos sintácticos de la información. o lo que es igual la forma o código en que se presentan los datos. A través de este nivel, los procesos de aplicación adquieren independencia de la representación de los datos incluyéndose en su entorno las posibles transformaciones de códigos.
Los servicios que proporciona el nivel de presentación serán, por tanto la transformación de la sintaxis de 1os datos y la sintaxis de la presentación de 1os mismos.

Nivel de sesión
Este nivel proporciona los medios necesarios para la cooperación entre entidades de presentación para organizar, sincronizar y gestionar el intercambio de los datos. Para ello, el nivel de sesión proporciona 1os servicios para establecer una según cada vez que se desea establecer una comunicación entre entidades de presentación.
Los servicios proporcionados por el nivel de sesión son los siguientes:
Establecimiento de la conexión de sesión. Se conectan dos entidades de presentación a petición del usuario.

- **Liberación de la conexión de sesión**. Cuando la transferencia termina se procede a la desconexión.
- **Intercambio de datos**. Es el servicio que permite la transferencia de datos en un sentido, en otro o en 1os dos.
- **Sincronización y mantenimiento de la sesión.** Para proporcionar un intercambio ordenado de datos de las entidades de presentación, el nivel de sesión realiza la función de sincronización de1 dialogo.

Nivel de transporte
El nivel de transporte trata de proporcionar una comunicación de datos transparente entre las entidades de sesión liberándolas de los detalles en que han sido transferidos dichos datos.
Es el nivel que controla la comunicación entre dos nodos de una red, con independencia de 1os elementos de comunicación, es decir, actúa coma puente entre el nivel de sesión y el nivel de red, escondiendo los detalles de la red.
La función de este nivel es aceptar los datos del nivel de sesión, fragmentándolos en

unidades más pequeñas en case necesario y pasarlos al nivel de red, asegurando que todos los fragmentos llegan correctamente a su destino. Todo ello lo consigue con funciones de direccionamiento, multiplexación, el establecimiento de la conexión y desconexión y, por último, la transferencia de los datos y su control del flujo.
En ocasiones, el nivel de transporte ofrece además, servicios de detección y corrección de errores, para asegurar la integridad de 1os datos.

Nivel de red
El propósito de este nivel es proporcionar la comunicación existente entre 1os nodos de la red.
Las funciones de este nivel que se ofrecen al nivel de transporte comienzan con la conexión y desconexión de las redes, sincronizan y controlan el flujo de las transferencias y detectan la presencia de errores en la transmisión, recuperándolos en case necesario.

Nivel de enlace de datos
Es el nivel encargado de la transmisión de los datos sobre un canal de comunicaciones donde se detectan y corrigen los errores que puedan ocurrir en el nivel físico.
Los protocolos del nivel de enlace definen reglas para establecer y liberar un enlace de datos controlar la correcta transferencia de información y recuperar anomalías.
Las funciones del nivel de enlace comienzan con el establecimiento y liberación de conexiones de enlace de datos, sincronización y control del flujo de datos detección y recuperación de errores y gestión del propio nivel.

Nivel físico
El nivel físico es el encargado de definir las señales y características físicas y eléctricas para la conexión entre 1os equipos informáticos (unidades centrales y periféricos) y las unidades de comunicación.
Las funciones básicas que realiza este nivel son las propias conexiones Físicas la identificación de los circuitos de datos, el secuenciamiento de los mismos y la propia gestión de1 nivel físico.

CAPÍTULO VII. EL NIVEL DE ENLACE DE DATOS

7.1 Introducción

Podemos definir el concepto de enlace de datos coma el conjunto de dos equipos terminales de datos más los elementos que configuran la red de transmisión, que permiten el intercambio de información entre ambos.
La parte lógica de control de la comunicación es la que corresponde al nivel de enlace de datos. Está compuesta por el conjunto de procedimientos para el establecimiento, mantenimiento y desconexión de circuitos para él envió de bloques de información, controla la correcta transferencia de los datos y articula los métodos necesarios para la detección y corrección de errores.
La unidad de transferencia de información en el nivel de enlace es la trama, de manera que la información que se desea transmitir se trocea en unidades (tramas) que se envían secuencialmente En algunos protocolos se denomina bloque a cada una de estas unidades de información.

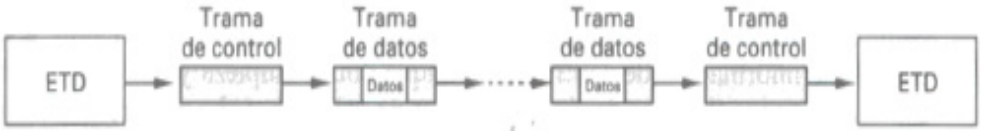


Figura 19: Trama de datos ETD.

7.2 Protocolos de enlace de datos

Un protocolo de enlace de datos es un conjunto de normas o formatos de los datos que define el modo en que se deben comunicar dos o más estaciones.
Para la comunicación de información entre las estaciones de enlace de datos se puede considerar un conjunto de funciones comunes a todo protocolo de enlace si bien cada protocolo realiza cada función a su estilo.

7.3 Funciones de un protocolo a nivel de enlace de datos

Las funciones que han de considerar los distintos protocolos de enlace de datos son las siguientes:

Iniciación. (En muchas ocasiones se emplea el anglicismo inicialización.) Esta función conlleva el envío de tramas de control entre las estaciones dc enlace para descubrir la disponibilidad de ambas para transmitir o recibir la información.

Identificación. Se compone del conjunto de procesos necesarios para identificar a la estación destino o a la que da origen a la información que se desea transmitir. Frecuentemente se realiza mediante el envío de tramas de identificación entre las estaciones para que puedan reconocerse mutuamente. El procedimiento de identificación se utiliza en enlaces conmutados o enlaces punto a punto.

Terminación. Se refiere a los procesos que determinan que todos los datos han sido bien recibidos y, por tanto, se produce la desconexión del enlace dejando libres los recursos ocupados en ambas estaciones.

Sincronización. El envío de información se produce normalmente a nivel de conjuntos de 8 bits o algún múltiplo de 8, por lo que es necesario establecer una serie de procesos cuya misión sea la de sincronizar el envío y recepción de dichos octetos de información.

Segmentación y bloqueo. En ocasiones aparece la necesidad de transmitir información compuesta por mensajes 0 excesivamente largos 0 excesivamente cortos. Para adaptar estos tamaños a un formato que esté en consonancia con las características del enlace para la eliminación de posibles errores en la transmisión así como optimizar la utilización del enlace, se dividen los mensajes largos en varías tramas (segmentación) 0 se unen los mensajes cortos en una trama (bloqueo).

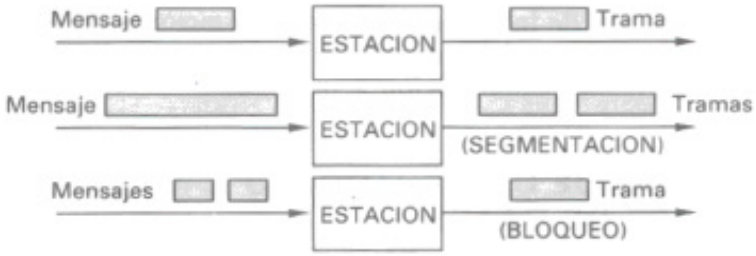


Figura 20: Segmentación y Bloque.

Sincronización de trama. Se trata de cómo definir y diferenciar una trama del conjunto de informaci8n que se transmite. El proceso consiste en unir a la trama una información de control que indique dónde empieza y d0nde termina.
Existen tres modalidades de sincronización de las tramas. La primera consiste en la utilización de caracteres para indicar el principio y el fin de la trama.
La segunda se compone de un carácter de principio de la trama y un contador que indica el número de caracteres de la misma. La tercera modalidad utiliza un guión para indicar el principio y el final de una trama.

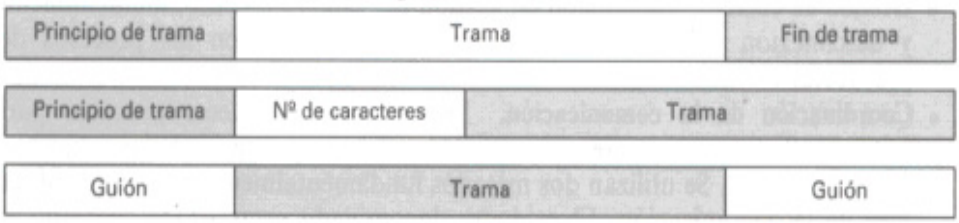


Figura 21: Trama de datos ETD.

Transparencia. La función de transparencia trata de eliminar la mala interpretación que pueda darse si se desea transmitir un conjunto de bits o caracteres similares a

los de algún elemento de control del protocolo. Existirán elementos en este nivel para la eliminación de estos posibles conflictos. Se necesita la transparencia, por ejemplo, cuando se transfieren archivos binarios o volcados de memoria puesto que en estos casos puede aparecer cualquier combinación de bits.

Control de errores. La función de control de errores debe asegurar que ante la aparición de un error en la transmisión de una trama esta se detectara y posteriormente se recuperará de algún modo. Las técnicas más utilizadas para la recuperación de errores son los denominados controles directo de errores petición automática de retransmisión y la de control del eco que se utiliza para transmisiones asíncronas.

Control de flujo. Esta función trata de regular el ritmo de envío de tramas desde el transmisor al receptor. Las distintas técnicas utilizadas para el control de flujo aseguran que cuando el transmisor envía información el receptor tiene recursos para recibirlas. En general se utilizan diversas técnicas para el control de flujo. En primer lugar la técnica de parada y espera, que consiste en él envió de una trama, produciéndose una parada por parte del transmisor hasta que el receptor le envía una señal de estar dispuesto a recibir la siguiente. La segunda técnica se denomina parada y arranque y consiste en que el transmisor envía tramas hasta que el receptor le envía una señal para que pare un determinado tiempo; a continuación el receptor vuelve a indicar, mediante una señal, al transmisor que puede seguir enviando tramas. Por último la técnica de ventana deslizante, en la que el receptor autoriza en cada instante al transmisor al envío de un determinado número de tramas: el receptor renovará las autorizaciones según tenga disponibilidad para la recepción de las mismas. Esta última técnica es la más frecuentemente utilizada en los protocolos modernos.

Recuperación de anomalías. Esta función incluye el control de todas las situaciones imprevistas que puedan presentarse a lo largo de una transmisión. Estos procesos de control de situaciones anómalas estiman plazos de espera para establecer reintentos de normalización de la situación, así como limitar a un determinado valor el número de reintentos.

Gestión de enlace. La función de gestión del establecimiento, mantenimiento y desconexión del propio enlace también se incluye en los procesos del protocolo.

Coordinación de la comunicación. Es necesario establecer procesos para evitar conflictos en el establecimiento de los enlaces por parte de las estaciones que los solicitan. Se utilizan dos métodos fundamentalmente para la coordinación de la comunicación. El primero, denominado centralizado, consiste en la asignación de una estación como principal, y toma la responsabilidad del intercambio de información que se produzca enviando información al resto de estaciones y sondeándolas cada cierto tiempo para recibir información de ellas; en este caso una estación secundaria no puede transmitir hasta que la principal se la solicite. Es un método utilizado en los enlaces multipunto. El segundo método es el denominado de contienda, que se utiliza en los enlaces punto a punto, en los que cada estación

puede solicitar transmisión en cualquier momento, existiendo procedimientos para solucionar situaciones de colisión de peticiones.

7.4 Fases de un protocolo a nivel de enlace de datos

Para la realización de una transmisión de información entre dos estaciones a través del enlace de datos es necesario pasar por las siguientes fases:

Conexión física del circuito. Constituye esta fase todo el conjunto de operaciones necesarias para el establecimiento del circuito físico que une las dos estaciones que desean comunicarse.

Conexión lógica del circuito. Esta fase se compone del conjunto de procesos necesarios para la preparación de la transmisión de la información a través del circuito físico ya establecido, comprobándose que ambas estaciones se encuentran en disposición de empezar.

Transferencia de los datos. Esta fase está compuesta por los procesos necesarios para que el conjunto de tramas de información llegue desde la estación emisora a la receptora sin ningún tipo de problemas y sin la aparición de errores.
Terminación de la transferencia de datos. Esta fase comprende los procedimientos de liberación de los elementos que han intervenido en la transmisión de los datos.

Desconexión del circuito físico. Constituye esta fase todo el conjunto de operaciones necesarias para la desconexión del circuito físico entre las dos estaciones.

La unidad de datos en este nivel, como ya hemos dicho, se denomina trama, existiendo dos tipos diferentes:

Trama de datos. Es la unidad que se transmite. Contiene un elemento de información acompañado de una cabecera y una cola de control.

Trama de control. Son las unidades de información que se envían las estaciones para el control y coordinación del establecimiento y terminación del enlace de datos.

7.5 Protocolos orientados a carácter

Un protocolo orientado a carácter es aquél en el que los mensajes se componen de un conjunto de caracteres de un determinado código. Cada carácter, tanto de información como de control, tiene una significación específica y no ambigua. Estos fueron los primeros protocolos que se utilizaron y aún siguen estando en vigor.
A nivel de enlace, en los comienzos de la teleinformática, aparecieron una gran variedad de protocolos incompatibles entre sí debido a la falta de normalización de estos primeros años. En 1970 aparece la norma ISO 174.50 que regulaba la conexión entre equipos informáticos a través de redes de transmisión públicas. Entre los protocolos orientados a carácter más extendidos se encuentran los desarrollados por empresas privadas, como el BSC, desarrollado por IBM y adoptado por varios

fabricantes, y el SLC desarrollado para empresas de transporte aéreo.
Los protocolos orientados a carácter emplean un determinado código para la transmisión de la información en el que utilizan ciertos caracteres para establecer control en la comunicación. El envío de tramas de información va acompañado de determinados caracteres de control.

Estos caracteres de control se clasifican según su función, en tres categorías:
Delimitadores de bloques.
SYN. Se utiliza para mantener el sincronismo en la transmisión de los caracteres.
SOH. Este carácter indica el principio de un mensaje compuesto por una secuencia de caracteres.
STX. Este carácter indica el comienzo de un bloque de información dentro de un mensaje.
ETX. Este carácter indica el lina1 de un bloque y el final del mensaje.
ETB. Este carácter indica el final de un bloque al que le siguen otros bloques.
Controladores del diálogo entre las estaciones.
EOT. Este carácter indica que la transmisión ha terminado y puede ser liberado el enlace.
ENQ. Este carácter indica que se desea respuesta de la estación contraria o que se desea establecer el enlace.
ACK. Este carácter indica que se ha recibido bien un bloque de información.
NAK. Este carácter indica que se ha recibido mal un bloque de información.
Para realizar transmisión transparente.
DLE. Este carácter cambia el significado de los caracteres de control que le siguen para que las estaciones puedan enviarse información coincidente con los propios caracteres de control.
El dialogo que se establece entre dos estaciones que desean transmitir una determinada información comienza con el establecimiento del enlace. Para ello en el caso de enlaces punto a punto, la estación transmisora envía a la receptora la seña1 ENQ; ésta, si se encuentra lista para recibir, devuelve la seña1 ACK con lo que se establecería el enlace para la transmisión. Si la estación receptora no estuviese preparada, enviaría la señal NAK, con lo que se produciría un período de espera para repetir nuevamente el intento del establecimiento del enlace.
Si se trata de enlaces multipunto la estación primaria establecerá el enlace por sondeo o selección según se trate de recibir o enviar información a una estación secundaria. En el caso de recepción por sondeo la estación secundaria. Cuya dirección coincide con la de sondeo si desea enviar información, contestará con un bloque que comenzará por STX o SOH, y si no desea transmitir contestara con EOT. En el caso de transmisión de la estación primaria, la estación secundaria seleccionada contestara con ACK o NAK según se encuentre o no dispuesta a recibir los datos.
Establecido el enlace, tiene lugar la transferencia de la información entre las estaciones bajo el control de la estación que estableció el enlace. La estación receptora, por cada trama de datos que recibe, responde a la estación transmisora con ACK para validar la trama o NAK para rechazarla. El lina1 de la transmisión estará indicado por la señal ETX o ETB.
Por último, se libera el enlace por envío de la señal EOT desde la estación transmisora a la receptora.

7.6 Protocolos orientados a bit

Son protocolos más modernos que los anteriores y su aparición se debe a las dificultades que en determinados casos presentan los protocolos orientados a carácter. Las grandes ventajas de estos protocolos son: Independencia del código utilizado. Se trata de transmitir conjuntos de bits que en principio pueden configurar información en cualquier código.

Gran eficiencia en la transmisión. La relación existente entre los bits de información y los bits de control es muy alta.

Gran fiabilidad en las transmisiones. Se dispone de métodos de control para la detección y recuperación de errores con gran eficacia.

CAPÍTULO VIII. INTRODUCCIÓN A LA ARQUITECTURA TCP/IP

8.1 ¿Qué es INTERNET?

Internet es la mayor red de redes de computadoras interconectadas entre sí, que utiliza un grupo común de protocolos de comunicación (TCP/IP)

8.1.1 Historia de Internet

- ARPANET (Advanced Research Projects Agency Network)
- TCP/IP (Transmission Control Protocol/Internet Protocol)
- NSFNET (National Science Foundation Network)
- Computadoras de oficina
- Mejoramiento Tecnológico de la red
- Herramientas Avanzadas de Internet

Historia de Internet

1969 - Se crea ARPANET como una red experimental por parte de la DARPA (Defense Advanced Research

Project Agency) del Departamento de Defensa (DoD) de EU para investigar en el campo de la conmutación de paquetes.

ARPANET inicialmente conectó entre sí a 4 Universidades y permitió a los científicos intercambiar información y recursos a grandes distancias.

1970's Se desarrollan las primeras herramientas de red, entre las que se incluyen:

1972 - La NCSA (National Center for Supercomputing Applications) desarrolla la aplicación Telnet.

1973 - Se introduce el FTP, el cual estandariza las transferencias de ficheros entre computadoras.

Se crea, en 1979, el primer comité para coordinar y guiar el diseño de los protocolos de Internet, la Junta de Control y Configuración de Internet (ICCB).

1982: Aparece la primera computadora personal. Muchas de ellas ya venían equipadas con un sistema operativo denominado Berkeley Unix, el cual incluía el software necesario para conectarse a la red.

Además brindaba un interfaz de programación sobre TCP/IP.

1983: TCP/IP se convierte en el único conjunto de protocolos usados por ARPANET.

1984: ARPANET se divide en 2 redes (MILNET y ARPANET) para separar la red militar de la red no militar, pero DoD mantiene su presencia en ARPANET.

1985-86: La NSF (National Science Foundation) conecta entre sí 6 grandes centros de computadoras a 56 Kbps. A esta red se le llamo la NSFNET y se une a ARPANET.

1986: Con el objetivo de expandir el acceso, la NSF comienza el desarrollo de redes regionales conectadas al backbone. Estaredes incluían las principales Universidades de los EU. y algunas instituciones científicas.

1987: La NSF le cede el derecho a la Merit Networks Inc para operar y administrar el futuro desarrollo del backbone de la NSFNET. MERIT en colaboración con IBM, MCI y las Corporaciones de Telecomunicaciones en los EU. Comienzan las investigaciones para desarrollar tecnologías mas rápidas de comunicación.

1989: Al backbone de la NSFNET se le eleva la velocidad hasta el rango de "T1" (1,544 Mbits/seg).

1990: ARPANET es disuelta. Se amplía la red más allá del dominio académico y científico.

1991: Se desarrolla la aplicación Gopher en la Universidad de Minnesota.

1992: Muchas compañías alrededor del mundo comenzaron a conectarse a la red, se generaliza el término Internet. Se crea la Sociedad Internet (ISOC).

1993: El CERN (Centro Europeo para las Investigaciones Nucleares) desarrolla el WWW. Se revoluciona totalmente la forma en que la información es organizada y presentada en Internet.

1993: El backbone de Internet es elevado a "T3" (45 Mbits/seg)

1993-1994: Se introducen en Internet los manejadores gráficos para WWW Netscape y Mosaic.

1995: Los sistemas operativos de redes generalizan el uso de TCP/IP.

1996: La velocidad del backbone asciende a un valor efectivo de 622 Mbits/seg.

1999: 56 218 000 hosts conectados a Internet. Comienza el aumento de velocidad

del backbone en los EU a 2.5 Gbits/seg.

8.2 Modelo OSI

Arquitectura de Redes

- El conjunto de funciones a cumplir para la transferencia exitosa de datos a través de una red se estructuran en capas.
- Las funciones más elementales en las capas inferiores.
- Cada capa le brinda servicio a su capa superior y recibe servicios de su capa inferior.

Nivel Físico

- Recibe la información de la capa de enlace y la entrega al medio de comunicación.
- Establece las relaciones con el medio de comunicación.
- Responsable de la transferencia de bits no estructurados.
- Establece niveles y formato de señal, conectores, características eléctricas, etc.

Nivel de Enlace

- Recibe paquetes del nivel de red y entrega tramas al nivel físico.
- Transforma el canal de transmisión en un enlace virtualmente libre de errores.
- Responsable de la transmisión de tramas entre dos máquinas directamente conectadas.
- Garantiza control de flujo y direccionamiento en comunicaciones multipuntos.

Nivel de Red

- Una comunicación puede implicar múltiples enlaces.
- Es responsable del enrutamiento en la red.
- El nivel de Red transporta paquetes a lo largo de la red.
- Controla la congestión y la contabilidad de la red.*

Nivel de Transporte

- Protocolo extremo- extremo.
- Asimila los mensajes de la capa de sesión, los divide en segmentos si fuera necesario y los entrega a la capa de red.
- Garantiza la integridad y orden de la información transmitida.
- Decide, en función de la aplicación que sea, los servicios que solicitara

de la capa de red.

Nivel de Sesión

- Permite el establecimiento de una sesión entre aplicaciones y controla la misma.
- Gestiona el control del dialogo.
- Ante una interrupción de la comunicación define quien la reanuda y a partir de donde se reanuda.

Nivel de Presentación

- Establece los aspectos de sintaxis.
- Si hay compactación de la información, es esta la capa encargada.
- Si se establece cifrado, la capa de aplicación la entrega a presentación para que ésta la cifre y la capa de presentación del receptor la descifra.
- Deberá resolver los problemas de diferencias en la codificación en los equipos finales.

Nivel de Aplicación

- Garantiza la adecuada comunicación entre las aplicaciones.
- Adecua la información a transferir a las condiciones de los equipos finales.

8.3 Arquitectura TCP/IP

Tecnologías usadas en Internet

- X.25
- Comunicación Satelital
- Frame Relay
- ATM
- Packet Radio
- Conexión por líneas conmutadas (SLIP, CSLIP,PPP)
- Ethernet
- Token Ring
- FDDI
- Protocolos inalámbricos

8.4 Retos a través del desarrollo de Internet

Necesidad de interoperabilidad: un conjunto de normas de comunicación (protocolos) y forma de interconexión de diferentes tecnologías.

Sistemas Abiertos: Permite la comunicación entre máquinas con diferentes arquitecturas de hardware y sistemas operativos diferentes. En el caso de Internet,además, las especificaciones no pertenecen a ningún fabricante, son del dominio público.

Solución: TCP/IP

¿Qué es TCP/IP?

- Conjunto o familia de protocolos desarrollados para permitir a computadoras cooperativas y heterogéneas compartir recursos a través de una red.
- Se diseñó teniendo en cuenta como elemento básico la existencia de muchas redes interconectadas por medio de routers o pasarelas (gateways).

¿Por qué tienen tanto éxito los protocolos TCP/ IP?

- Los protocolos TCP e IP son los más conocidos y de ahí el nombre generalizado.

El éxito de los protocolos TCP/IP radica en su capacidad de adaptarse a casi cualquier tecnología de comunicación subyacente.

8.5 Características distintivas de TCP/ IP

- Independencia de la tecnología de las redes soportes (abstracción del hardware).
- Interconexión universal: Sistema de direccionamiento que permite que cada estación conectada a la red posea una dirección diferente, usada para tomas las decisiones de enrutamiento.
- Acuses de recibo punto a punto y no en cada tramo del trayecto.
- Amplia gama de servicios estandarizados que soporta: Se basan en la existencia de las interfaces software existente en cualquier Sistema Operativo actual.
- Trabajo de estandarización ágil y normas libremente disponibles en la propia red en los RFC.

Ventajas de la abstracción del hardware

- Los programadores pueden trabajar a un nivel más alto no necesitan aprenderse o recordar detalles sobre una configuración de hardware en particular.
- Los programas no necesitan cambiarse cuando se reconfiguran las máquinas o las redes.

- Puede proporcionarse una comunicación directa entre un par arbitrario de máquinas.

¿Qué hay que entender bien?

- Internet no es un nuevo tipo de red física.
- Es un método de interconexión de redes físicas y un conjunto de convenciones que permiten que las computadoras conectadas en una red interactúen unas con otras.
- Permite construir sistemas de comunicación homogéneos que usen tecnología de hardware heterogénea

8.6 Estructura de capas

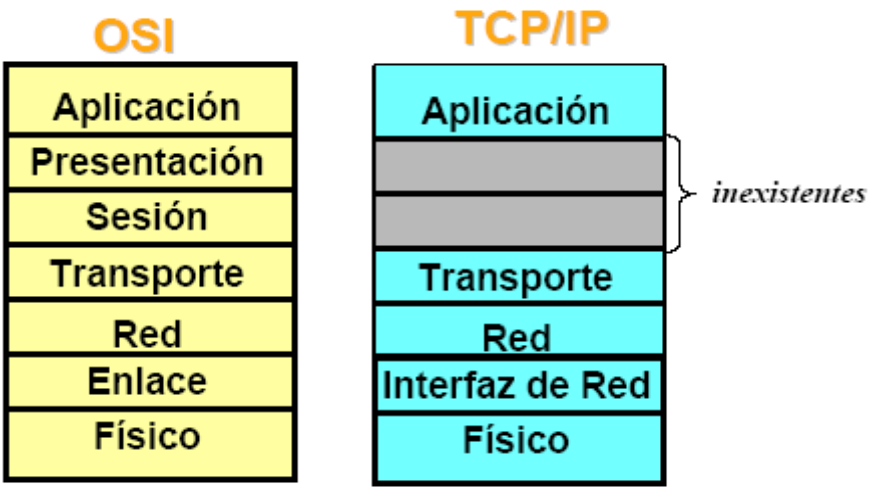


Figura 22: Estructura de capas OSI.

Capa de Interfaz de Red + Físico

- No define reglas para este nivel. En él se ubican todos los protocolos que permiten que los datagramas IP viajen por la red.
- Se encuentran desarrollados interfaces para todo tipo de red.
- Varían de host a host y de red a red.
- Se aceptan los estándares definidos por otras normativas.

Capa de Red

- Define una red de conmutación de paquetes no confiable y no orientada a conexión.

• Se acostumbra a denominar capa de Internet por su capacidad de interconectar redes independientemente de sus características físicas.

• Su función es entregar datagramas a la red para que éstos viajen a su destino final.

• Contiene el protocolo IP, que entrega a los niveles inferiores los datagramas para que sean transportados por la red.

• IP es responsable del esquema de direcciones.

Servicios de la Capa de Red

• Son la base de todos los servicios TCP/IP.

• Es una abstracción del mismo servicio que ofrece la mayoría de las redes de conmutación de paquetes.

• Enrutamiento de paquetes por separado, basándose en la información presente en cada paquete.

• No se garantiza una entrega confiable y en orden.

• Se introducen directamente en el hardware subyacente (alta eficiencia).

• Responsables de que los protocolos TCP/IP sean adaptables a un amplio rango de hardware.

Capa de Transporte

• Permite que procesos de igual nivel realicen una comunicación extremo-extremo.

• Definidos dos protocolos para el nivel de transporte:

• TCP (Transmission Control Protocol)

• UDP (User Datagram protocol)

Servicios de Transporte

• Transporte de flujo no confiable, no orientado a conexión: Brindado por el UDP como extensión simple de los servicios del IP.

• Empleado por aplicaciones que no requieren de control de flujo y de errores porque no los necesitan o porque desean implementarlos en los niveles superiores.

• Usado ampliamente en transferencias cortas y rápidas, en solicitudes respuestas

• Cliente-servidor y para servicios en que es más importante la llegada en tiempo que la llegada libre de errores.

Ej. Transmisión de audio y vídeo.

• Transporte de flujo confiable, orientado a conexión:

Brindado por el TCP. Permite que una aplicación en una computadora establezca una “conexión” con una aplicación en otra computadora para intercambiar un gran flujo de datos.

Realiza las funciones de fragmentación del flujo de datos en segmentos y viceversa, recuperación automática de los errores de transmisión, control de flujo: paquetes perdidos o duplicados, etc.

Capa de Aplicación

- Contiene todos los protocolos de nivel superior.
- Responsable de las aplicaciones que se ejecutan sobre la internet.
- Los primeros protocolos de aplicación fueron Telnet, FTP y SMTP. Han surgido algunos de gran impacto como el HTTP (WWW).

Servicios de la Capa de Aplicación

- Correo electrónico (SMTP)
- Conexión a otras máquinas (Telnet, rlogin)
- Transferencias de archivos (FTP)
- Grupos de Noticias (News - NNTP)
- Listas de Distribución (ListServ, Majordomo..)
- Servidor de Información orientado a menú (Gopher)
- Servidor de Información multimedia (World Wide Web Server - HTTP)
- Conversaciones Interactivas (IRC, ICQ)
- Transmisión de video y sonido

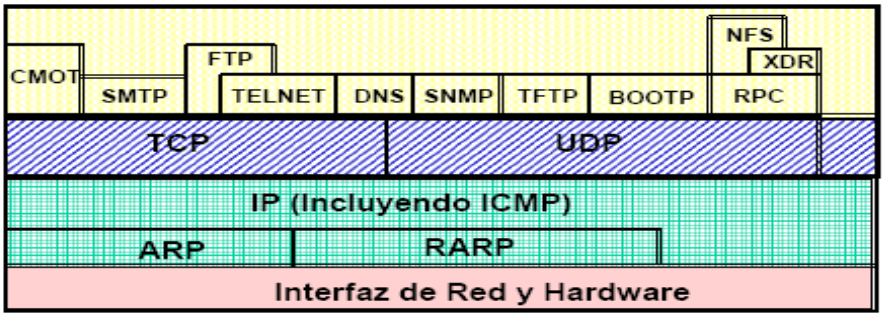


Figura 23: Arquitectura de niveles.

Arquitectura de Niveles

- ARP Address Resolution Protocol
- RARP Reverse Address Resolution Protocol

- IP Internet Protocol
- ICMP Internet Control Message Protocol
- TCP Transmission Control Protocol
- UDP User Datagram Protocol
- SMTP Simple Mail Transfer Protocol
- TELNET Remote Terminal Protocol
- FTP File Transfer Protocol
- DNS Domain Name System
- SNMP Simple Network Management Protocol
- TFTP Trivial File Transfer Protocol
- BOOTP BOOTstrap Protocol
- NFS Network File System
- XDR eXternal Data Representation
- RPC Remote Procedure Call

8.7 IP Protocolo Internet

Es uno de los dos protocolos más importantes de la familia de protocolos TCP/IP pues es el que se encarga de la entrega de paquetes (datagramas). Define un mecanismo de entrega sin conexión, no confiable y con el mejor esfuerzo.

Mecanismos de entrega

Sin conexión: Cada paquete se trata de forma independiente de los demás.

No confiable: La entrega no está garantizada. Los paquetes se pueden perder, duplicar, retrasar o entregar sin orden, pero esto no será detectado.

Con el mejor esfuerzo: Se hace un serio intento por entregar los paquetes, o sea no se descartan paquetes caprichosamente y la no confiabilidad sólo aparece cuando no existe otra opción y la red física falla.

Propósitos del protocolo IP

Define la unidad básica (formato de los datos) para la transferencia de datos por las redes TCP/IP: el datagrama IP.

Su software realiza la función de ruteo, seleccionando la ruta por la que los datos serán enviados.

Incluye un conjunto de reglas que dan forma a la idea de entrega de paquetes no confiable.

Reglas para el procesamiento de los datagramas Aspectos importantes

1. Forma en que los routers y host deben procesar los paquetes.

2. Cómo y cuándo se deben generar los mensajes de error.

3. Condiciones bajo las cuales los paquetes deben ser descartados.

Datagrama IP

Su estructura tiene mucho parecido al de una trama física.

Los datagramas, a diferencia de las tramas físicas que son reconocidas por hardware, son manejados por software y su tamaño dependerá del seleccionado por el protocolo.

Los datagramas viajan en las tramas físicas. **Relación Datagrama – Trama Física**

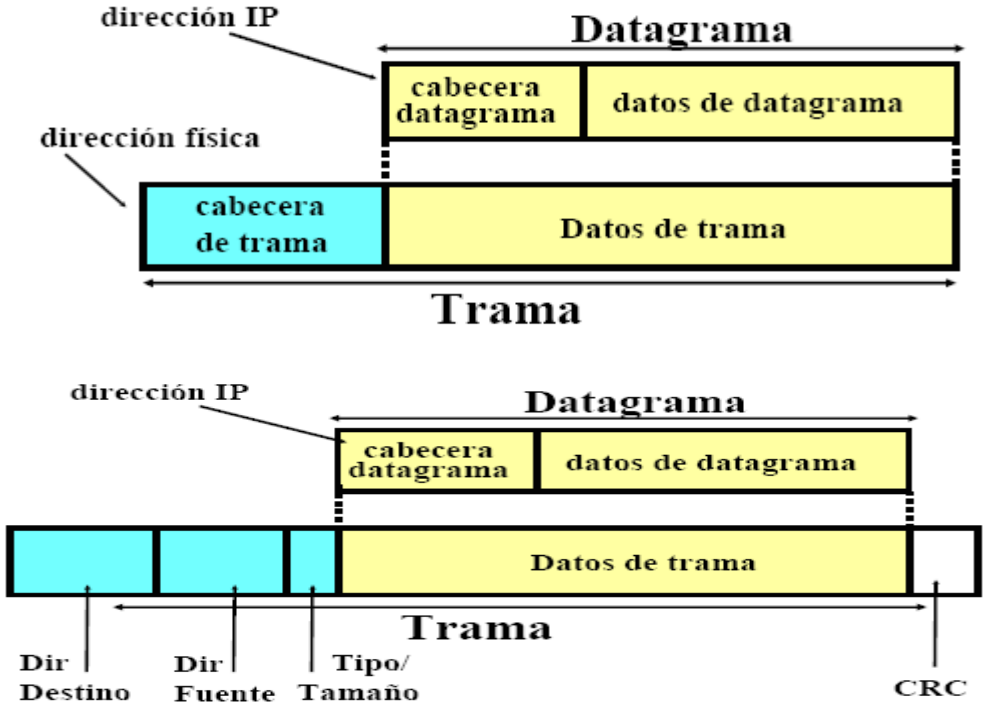


Figura 24: Encapsulación en Ethernet.

Encapsulación

Transportar cada datagrama dentro de una trama de red física.

Esto limita el tamaño del datagrama al tamaño que permite el campo de datos de la trama lo cual a su vez depende del hardware de la red en particular.

8.8 Fragmentación

Transferencia Máxima

Es el límite superior fijo para la cantidad de datos que se puede transmitir en una trama física y depende del hardware de la red.

La capacidad de adaptación de IP para diferentes redes físicas debe contemplar adaptarse al MTU d1e la red por la que viaje el datagrama.

Fragmentación

Es la técnica que se emplea para que el tamaño del datagrama sea independiente del campo de datos de la trama.

Si un datagrama viaja por diferentes redes físicas, que tienen diferentes MTU, es muy difícil seleccionar adecuadamente el tamaño del datagrama y por ello el origen lo predefine, y se fragmenta el datagrama cuando tiene que viajar por una red con MTU más pequeña.

La fragmentación por lo general se realiza en un router intermedio.

El tamaño de los fragmentos debe seleccionarse de forma tal que sea un múltiplo de 8.

Encabezado de cada fragmento reproduce la mayor parte del encabezado del datagrama original excepto por un bit del campo Banderas que indica que es un fragmento.

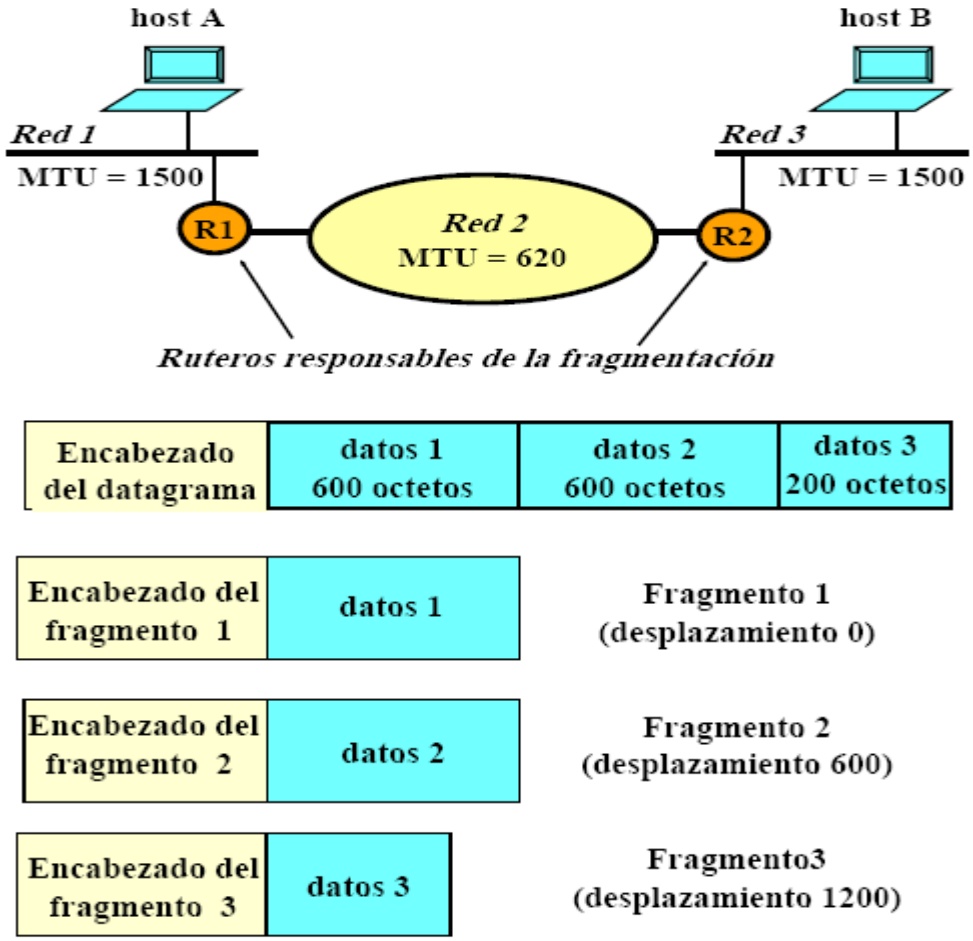


Figura 25: Ruteros responsables de la fragmentación.

Re ensamblado

• Es la unión de los fragmentos, en el destino final, antes de que sean procesados.

• Tanto la fragmentación como el reensamblado son procesos automáticos que se dan bajo el control del sistema operativo y no del usuario final.

• La máquina de recepción hace arrancar un temporizador de reensamblado cuando recibe el fragmento inicial, si el temporizador expira antes de la llegada de todos los fragmentos la máquina descarta los ya recibidos.

Desventajas

• Los fragmentos pequeños transitan así aunque pasen por una red con MTU mayor lo cual trae cierta ineficiencia.

• Si se pierde cualquier fragmento el datagrama no podrá reensamblarse aunque la mayoría haya llegado al destino.

• La probabilidad de perder un datagrama se incrementa con la fragmentación pues la pérdida de un solo fragmento provoca la pérdida del datagrama completo.

Formato del datagrama

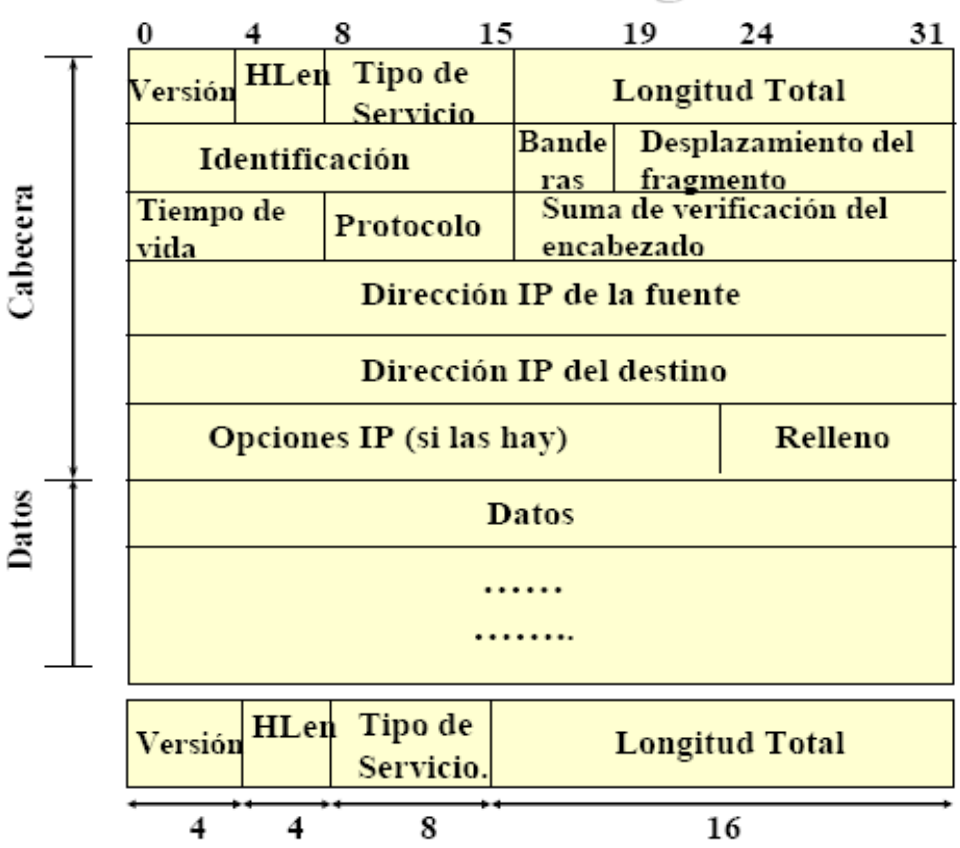
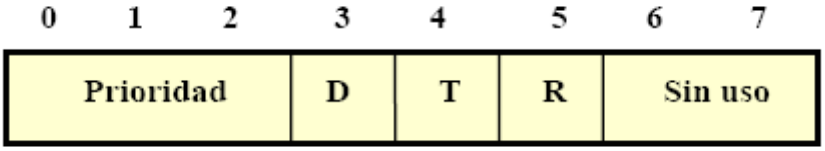


Figura 26: Formato del datagrama.

- Versión: Contiene la versión del protocolo IP que se utilizó para crear el datagrama.
- HLen: Proporciona la longitud del encabezado medida en palabras de 32 bits.
- Longitud Total: Proporciona la longitud total del datagrama IP incluyendo encabezado y datos. Como este campo tiene 16 bits, el tamaño total del datagrama puede ser 216 o 65,535 octetos.
- Tipo de Servicio (TOS): Especifica cómo debe manejarse el datagrama por el algoritmo de ruteo, al seleccionar una ruta entre varias hacia un destino, a

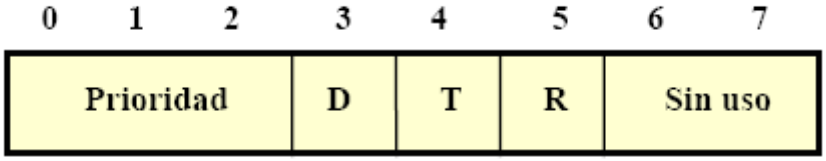
partir de conocer el hardware disponible en esas rutas.

- No se garantiza la realización del tipo de transporte solicitado pues puede no existir una ruta adecuada para ello.



Campo TOS

- Bits 0, 1 y 2: Especifican la prioridad del datagrama con valores extremos 0 (prioridad normal) y 7 (control, de red). Es desconocido por la mayoría de los sistemas.
- Bits D, T y R: Especifican el tipo de transporte que se desea para el datagrama.



- Indicación de qué aspecto es más importante a tener en cuenta en el caso que un router conociera más de una ruta para alcanzar un destino determinado.
- No es un requerimiento obligatorio.
- Solo uno puede estar "1".

Bit D: En "1" solicita procesamiento con retardos cortos.
Bit T: En "1" solicita alto desempeño.
Bit R: En "1" solicita alta confiabilidad.

- Los campos Identificación, Banderas y Desplazamiento del Fragmento controlan la fragmentación y el reensamblado de los datagramas.
- Identificación: Es un entero que identifica de forma unívoca a cada datagrama. Al copiarse en el encabezado de cada fragmento permite que se identifique qué fragmentos pertenecen a que datagramas.
- Desplazamiento del Fragmento: Especifica el desplazamiento que tienen los datos en cada fragmento, en unidades de 8 octetos comenzando con un desplazamiento igual a cero y que es muy útil para poder reensamblar en orden el datagrama pues estos por lo general no llegan en orden.

- Banderas: Posee 3 bits, uno no se usa y los dos restantes toman parte en el proceso de fragmentación y reensamblaje.
- Uno de esos bits se conoce como bit de no fragmentación pues cuando está en “1” especifica que el datagrama no debe fragmentarse.
- El otro bites conocido como bit de más fragmentos e indica si el fragmento contiene datos intermedios (“1”) o es el último de un datagrama (“0”).
- Tiempo de vida: Especifica el tiempo máximo que el datagrama puede permanecer en la red.
- Cada router o host que procese el datagrama debe decrementaren “1” este número.
- Cuando alcanza el valor 0 el routerdebe sacar de la red el datagrama y enviar un mensaje de error al que lo envió.
- Estimar este número es difícil pues entre otras cosas depende de las redes físicas por las que pase el datagrama.
- El objetivo es evitar que el datagrama esté en la red indefinidamente.
- Protocolo: Especifica que protocolo de alto nivel se usó para crear el mensaje que se está transportando en el área de datos.
- Define a que protocolo de nivel de transporte debe entregarse el mensaje en su destino final.
- La numeración asignada a cada protocolo y su definición se encuentra en la RFC 1700.
- Suma de Verificación del Encabezado: Asegura la integridad de los valores del encabezado y sólo se aplica a éste.

Ventajas: Suma de Verificación del Encabezado:

1. Como los routerstrabajan con el encabezado, les es más rápido procesar los datagramas ya que sólo tienen que calcular si existen errores en la cabecera.
2. Se permite que los protocolos de alto nivel seleccionen su propio esquema de verificación de los datos.

Desventajas: Suma de Verificación del Encabezado:

1. Los protocolos de alto nivel tienen que añadir su propia verificación de errores o estos no serán detectados.
2. Dirección IP de destino y de la fuente: Especifican la dirección IP de 32 bits de la fuente original y del destino final.
3. Datos: Muestra el comienzo del área de datos del datagrama y su longitud depende de lo que se esté enviando en el datagrama.

4. Opciones de IP: Posee longitud variable de acuerdo a la cantidad y tipos de opción que sean seleccionadas así como al tamaño del área de datos de las opciones.
5. No se requieren en todos los datagramas.
6. Se incluyeron para permitir opciones no contempladas en el diseño original, son usadas con frecuencia para monitorear y controlar la red.
7. Relleno (padding): Es un grupo de bits puesto a cero, necesarios para asegurar que la extensión, debido a las opciones del IP, sea un múltiplo exacto de 32 bits.
8. Su longitud depende de las opciones seleccionadas y puede ser cero.

Opciones IP
Opciones para los datagramas

- Cuando las opciones están presentes en un datagrama IP, aparecerán continuas sin separador entre ellas.
- Son parte integral del protocolo IP, lo cual hace que todos los estándares de implementación la deban incluir.
- Cada opción debe llevar un octeto que es el código de opción en el cual aparece la clase de opción y el número de la opción, así como un bit llamado copy que le indica a los routers, al fragmentar el datagrama, si la opción debe ser copiada en todos los fragmentos o solo en el primero.

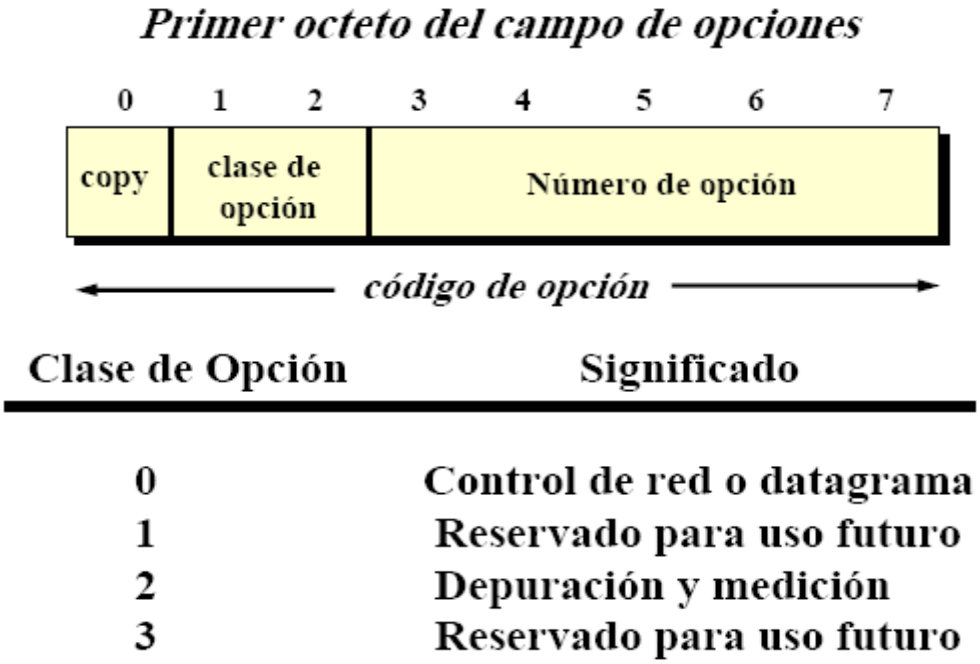
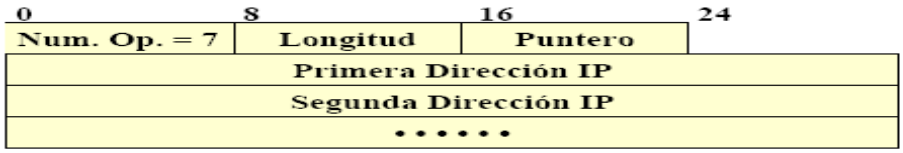


Figura 27: Clase de Opción.

Opción: Registro de ruta

- Clase de opción = 0; Número de opción = 7.
- La fuente indica que se cree una lista de direcciones IP que lleguen al destino junto al datagrama.
- Cada router que maneje el datagrama añade su propia dirección IP.
- Para usarla se requiere que la fuente y el destino estén de acuerdo pues si no el destino ignorará las rutas registradas.
- Esta opción solo se debe copiar dentro de uno de los fragmentos (Bitcopy= 0) pues en caso contrario llegarían al destino muchas listas diferentes de rutas y este no podría producir una lista completa de rutas al reensamblar los datagramas.
- Puede verse con: ping -R



Longitud:Longitud total del campo de Opción de Registro de ruta.

Figura 28: Registro de Ruta.

Formato de Opción Registro de Ruta

- **Puntero:** Especifica el desplazamiento dentro de la opción de la siguiente ranura disponible. Se inicia con 4 y se incrementa en 4 cada vez que un router inscribe su dirección en el campo de direcciones.
- Cada router va escribiendo su dirección siempre que el valor del puntero no sea mayor que el expresado en Longitud.
- Al llenarse el espacio reservado el router rutea sin inscribirse.

Opción de sello de hora

- Opción de sello de hora:Trabaja como la opción de registro de ruta.
- En una lista inicial vacía, cada router a lo largo del camino escribe sus datos: dirección IP(es opcional) y sello de hora (fecha y hora en que maneja el datagrama, expresada en msa partir de la medianoche).
- Cada entrada a la lista contiene 2 datos de 32 bits: la dirección del router que proporciona la entrada y un entero de sellos de hora de 32 bits.
- Oflow: campo de 4 bits, contiene contador de routers que podrían no

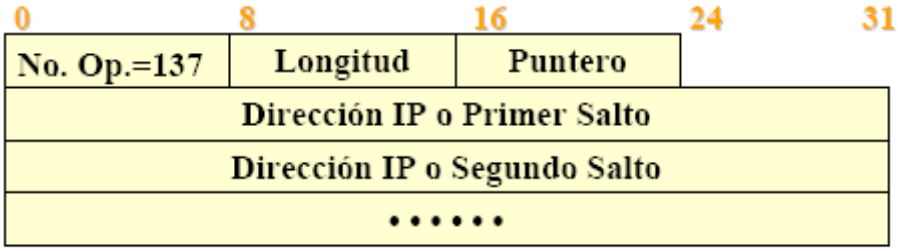
proporcionar su sello de hora por haberse declarado la Longitud demasiado pequeña.

- Bandera: 4 Bits que controlan el formato exacto de la opción permitiendo alternativas a la opción.

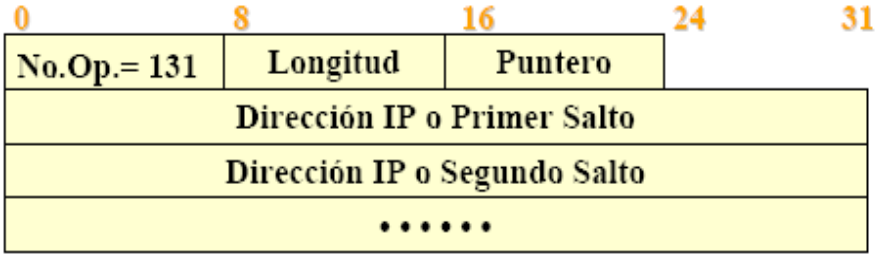
0 = Registro de sello de hora solamente, omitir dirección IP
1 = Anteponer a cada sello de hora la dirección IP del router.
3 = La dirección IP se especifica por el emisor, un router sólo registra un sello de hora si la próxima dirección IP en la lista coincide con su dirección IP.

Opción de Ruta Fuente Opción de Ruta Fuente

- Permite determinar una ruta en la red con lo cual los administradores pueden forzar a los datagramas a usar una ruta determinada, aun cuando los routers normalmente sigan otra ruta.
- Opción muy útil para probar el desempeño de las redes físicas.
- Existen dos formas de ruteo fuente: ruteo estricto de fuentey ruteo no estricto de fuente.
- Esta opción debe ser reproducida en todos los fragmentos pues en caso contrario, los fragmentos que no la tengan, no seguirán la ruta especificada. Bit copy= 1



- Ruteo Estricto de Fuente: Las direcciones especifican la ruta exacta que deben seguir los datagramas hasta su destino. Si el router no puede seguir la ruta indicada se produce un error



Ruteo no estricto de fuente: el datagrama debe seguir la secuencia de direcciones especificadas en la opción, pero permite múltiples saltos entre direcciones sucesivos.
Resumen

- IP proporciona un servicio de entrega de datagramas sin conexión, no confiable y con el mejor esfuerzo.
- IP define el formato de los datagramas.
- El datagrama IP posee cabecera y campo de datos.
- La cabecera contiene dirección fuente y destino (IP), control de fragmentación, prioridad y suma de verificación para la detección de errores.
- Además de los campos de longitud fija, cada datagrama puede contener un campo de opciones.
- Las opciones son utilizadas para monitorear y controlar la red de redes.
- Las opciones permiten especificar o registrar rutas o reunir sellos de hora conforme viajan los datagramas por la red.

8.9 Direccionamiento IP

Interconexión universal en Internet

- Un sistema de comunicaciones proporciona servicio universal de comunicaciones si permite que cualquier host se comunique con cualquier otro host.
- Los identificadores de host se clasifican en:
  - nombres: lo que un objeto es (preferidos por los humanos)
  - direcciones: donde está (usados por las computadoras)
  - rutas: como se llega a él (usados por los enrutadores)

Direcciones IP

- En Internet se emplea el término de direcciones IP.
- Cada host para conectarse a Internet y ser reconocido debe tener asociada una dirección IP la cual es un número binario de 32 bits.
- Esta dirección IP de 32 bits, es agrupada en octetos y se representa a través de 4 números decimales separados por puntos.

Ej: 169.158.144.1

Componentes de Direcciones IP

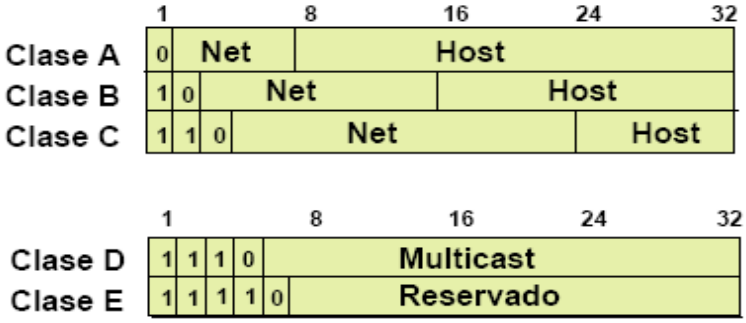
- Cada dirección es un par (netid, hostid), donde:

netid identifica una red

hostid identifica un host sobre esa red
Con esto, una dirección IP codifica a una red y a un host sobre esa red. No específica a una computadora en particular, sino una conexión a la red.

Componentes de Direcciones IP

- Las host multi-homed y los routers requieren de varias direcciones IP pues cada dirección corresponde con una conexión de la máquina a la red.
- Existen varias formas normalizadas de hacer esta división de direcciones. De acuerdo a esto se han establecido varias clases.



Direcciones clase A

- Clase A: Usada para redes grandes con alrededor de 224 (16777216) hosts.
- [1 - 127] . x . x . x
- En particular la dirección 127.0.0.0 queda reservada como dirección de loopback

Direcciones clase B

- 214 redes = 16,384 redes
- [128 - 191] . x . x . x
- Clase B: Usada para redes de dimensiones intermedias de hasta 216 (65,536) hosts.

Direcciones clase C

- 221 redes = 2’097,152 redes
- Clase C: Usada para redes de pocos hosts 28 (max 254) host.
- [192- 223] . x . x . x

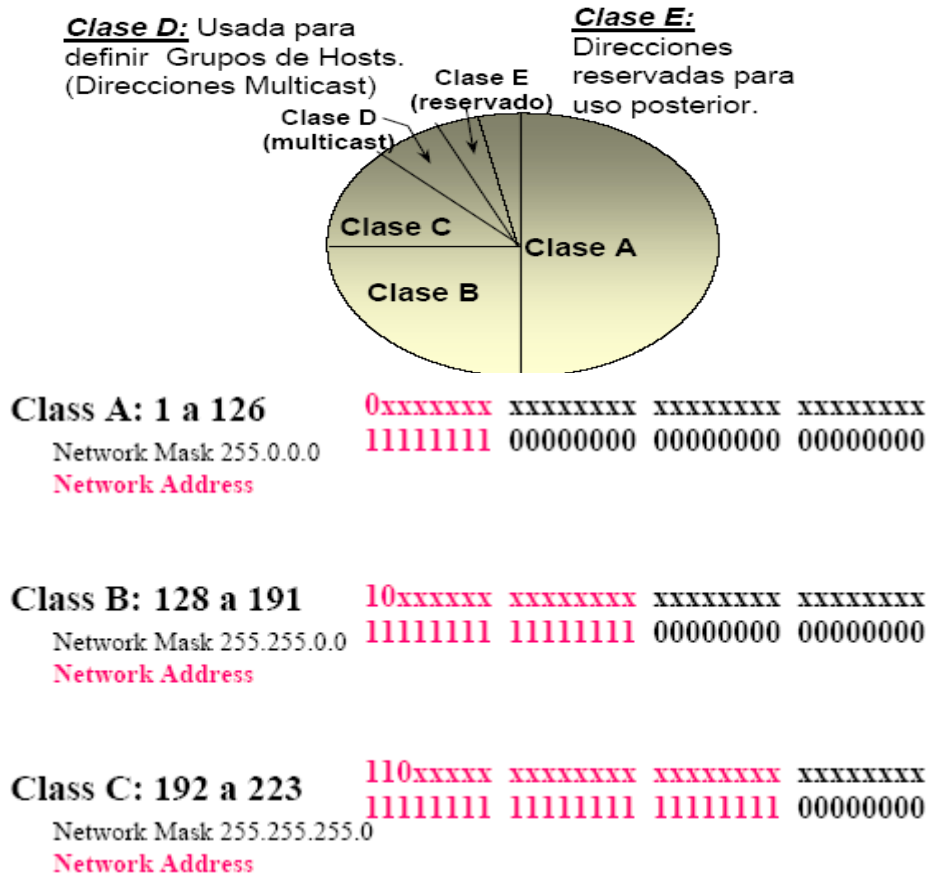


Figura 29: Direcciones clase C.

Identificación del host y la red en la dirección IP

- Los routers necesitan realizar una extracción eficiente del campo netid para lograr una velocidad alta de enrutamiento.
- No todas las redes cumplen con los requerimientos de cantidades de hosts que imponen el direccionamiento por clases: muchas veces se cuenta con subredes de menor número de máquinas o se pasan de la cantidad estipulada.

Direcciones de Host y de Red

- La existencia de las direcciones se debe a la necesidad de identificadores con vistas a poder enrutar los paquetes a través de la red.
- El enrutamiento es más eficiente si se realiza usando direcciones de redes y no direcciones de hosts.

- Las direcciones IP deben poder referirse tanto a hosts como redes:
  - una dirección IP cuyo campo hostid sea igual a cero se refiere a la red en cuestión.

Ejemplo: 168.158.0.0

Dirección de Loopback

- Dirección que se usa para la realización de pruebas del TCP/IP y para la comunicación de los procesos internos en una máquina. Los datos regresan a la computadora origen sin generar tráfico en ninguna red.
- La dirección de loopback es: 127.0.0.0 (Tipo A)
- Se especifica que:
  1. Un paquete enviado a esta dirección de red nunca debe aparecer en ninguna red.
  2. Un host o un router nunca deben difundir información de ruteo o de accesibilidad para el número de red 127, pues no es una dirección de red.

Direcciones de Difusión

Difusión Dirigida: Los campos hostid con todo en “1” se reservan para la difusión en la red, (dirección de difusión dirigida pues identifica la red objetivo), la cual sólo es posible si el hardware de la red lo permite.

Difusión limitada o Difusión de la Red Local: Proporciona una dirección de difusión para la LAN con independencia de la dirección IP asignada y consiste en 32 unos (todos “1”) y puede ser usada por un host como parte de su procedimiento de arranque antes de conocer su dirección IP o la dirección de su LAN.

Resumen de las Reglas Especiales de Direccionamiento

Todos “0”
Este host. (Sólo permitido en el arranque del sistema. Nunca es una dirección válida dedestino).

Todos “0” host
Un host en esta red (Sólo permitido en el arranque del sistema. Nunca es una dirección válida de destino).

Todos “1”
Difusión Limitada (en LAN). Todos los host de la LAN (Nunca es una dirección válida de origen).
Red todos 1

Difusión dirigida para red. (Nunca es una dirección válida de origen). La cantidad de "1" depende del tipo de red

127 cualquier (0.0.1)
Loopback (nunca debe aparecer en una red).

Problemas del direccionamiento
Desventajas del Direccionamiento de Internet

- Las direcciones se refieren a conexiones de red y no a host por lo que si esta se mueve de una red a otra, su dirección IP debe cambiar.
- Cuando una red tipo C crece y llega a más de 254 hosts, tiene que cambiar su dirección a una tipo B y de este tipo de direcciones quedan pocas.
- Como el ruteo usa la parte netid de la dirección IP, el camino hacia un host multi-homed depende de cual sea la dirección usada.

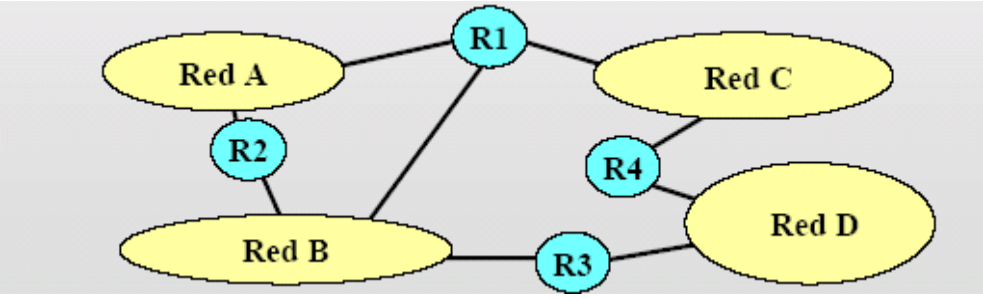
Problemas del direccionamiento en Internet (IPv4)

- Falta de flexibilidad en la asignación de rangos de direcciones.
- Agotamiento de direcciones clase B (corto plazo).
- Agotamiento del espacio total de direcciones IP a mediano plazo.

8.10 Enrutamiento de Datagramas IP

Conceptos

- Enrutamiento: proceso de selección de un camino por el que se mandarán los paquetes, pasando por varias redes físicas si fuera preciso.
- Enrutador: equipo que realiza la selección del camino. Una red de redes se compone de múltiples redes físicas interconectadas por enrutadores. Cada enrutador tiene conexiones hacia dos o más redes físicas.



Enrutamiento

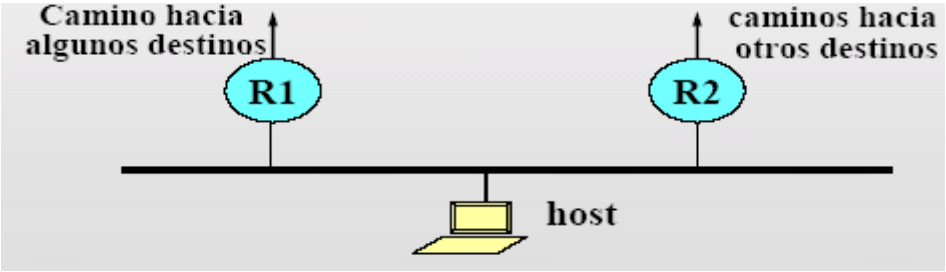
- Objetivo: Búsqueda de rutas en una red desde un punto origen a uno destino que satisfaga una serie de condiciones: rutas de mínimo costo, mínimo

retardo, o algún criterio administrativo.

- Algoritmo de enrutamiento: Parte del software del nivel de red responsable de decidir el camino a seguir por un paquete.
- Como la red es no orientada a conexión la decisión debe tomarse para cada datagrama.

Enrutamiento

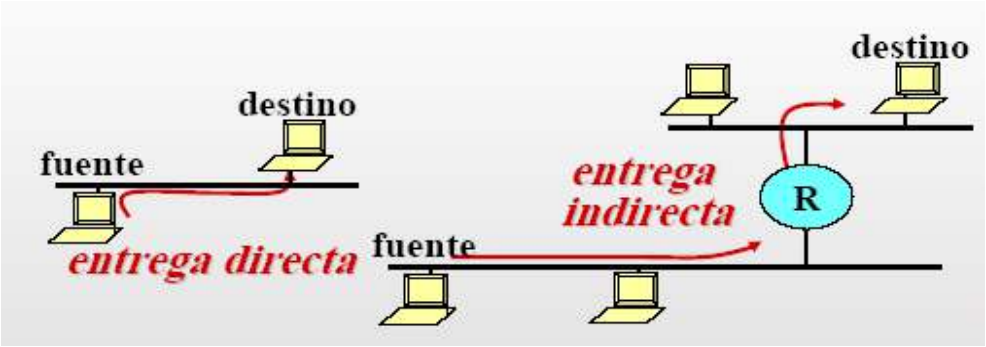
- En el proceso de enrutamiento participan tanto los hosts como los routers.
- Cuando un host genera un datagrama debe tomar una decisión de ruteo al decidir a quién va a enviar el datagrama aún en el caso que posea una única conexión de red.



Enrutamiento Directo e Indirecto

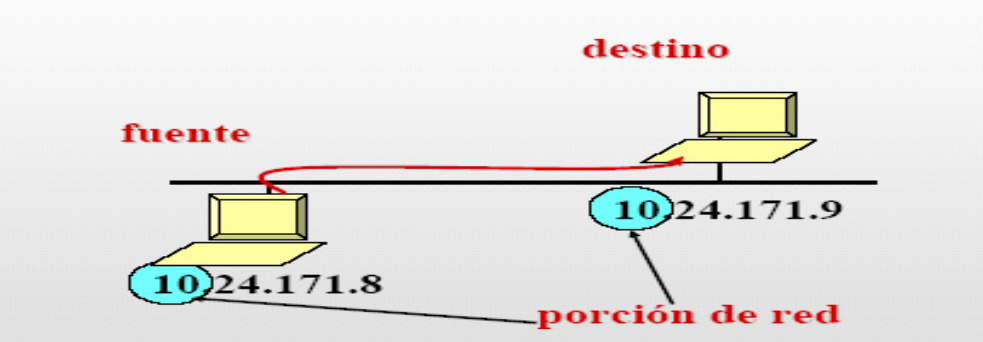
Entrega Directa e Indirecta

- En el enrutamiento pueden darse dos condiciones: Entrega directa y entrega indirecta.
- La entrega directa es la transmisión de un datagrama desde una máquina a través de una sola red física.
- La entrega indirecta ocurre cuando el destino no es una red conectada directamente, lo que obliga a la fuente pasar el datagrama a un ruteador para su entrega.



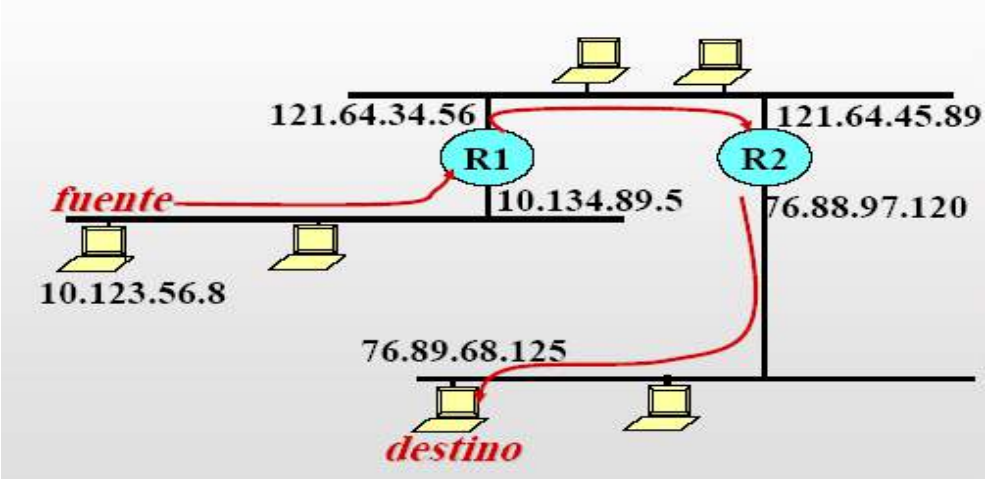
Entrega Directa

- El transmisor encapsula el datagrama dentro de una trama física, transforma la dirección IP de destino en una dirección física de hardware y envía la trama resultante directamente a su destino.
- ¿Cómo sabe el transmisor si el destino reside en una red directamente conectada?***
- La dirección IP tiene dos porciones, porción de red y porción de host.
- El enrutamiento se realiza con la porción de red.
- El transmisor extrae la porción de red de la dirección destino y la compara con la porción de red de su propia dirección, si son iguales ambas residen en la misma red y la entrega es directa.



Entrega Indirecta

- El transmisor identifica que el destino NO está en su propia red física y que debe decidir a cual router de su red entregar el paquete, atendiendo a su tabla de ruteo, para que éste lo encamine.
- El enrutadordebe entonces encaminar el datagrama a su destino.
- Entrega Indirecta Entrega Indirecta
- Una vez que la trama llega al enrutador extrae el datagrama encapsulado y el software IP selecciona el siguiente enrutadora lo largo del camino hacia el destino.
- Los enrutadoresen una red TCP/IP forman una estructura cooperativa por las que pasa el datagrama hasta llegar a uno que lo pueda entregar de forma directa.



Tablas de ruta Enrutamiento

- Para realizar sus funciones los routers mantienen tablas de enrutamiento.
- Las tablas pueden ser estáticas o dinámicas:
  - Las estáticas son introducidas por los administradores de las redes.
  - Las dinámicas son aprendidas gracias a los protocolos de enrutamiento.
- Las tablas se confeccionan utilizando como entradas las direcciones de RED y no de HOST.
- Los hosts también requieren tablas de enrutamiento para decidir a qué enrutador entregan la trama en caso de entrega indirecta.

Tablas de enrutamiento

- Es imposible que guarden información sobre cada posible host destino.
- Las tablas de enrutamiento sólo contienen direcciones de redpara decidir el salto siguiente.
- El tamaño de las tablas depende del número de redes interconectadas y no del número de hosts
- Las tablas se mantienen reducidas al contener sólo direcciones de red.
- El proceso de enrutamiento es más eficiente.
- Los detalles de hosts específicos quedan confinados al ambiente local en que operan.

Tipos de rutas

Tipos de Enrutamiento con salto al siguiente

- Una tabla de enrutamiento contiene pares (N, R), donde N es la dirección IP

de una red de destino y R la dirección IP del siguiente router hacia la red N.

- El enrutador R es conocido como el siguiente salto (nexthop).
- Enrutamiento con salto al siguiente
- Los enrutadores no conocen el camino completo hacia el destino.
- Cada registro de la tabla del enrutador apunta hacia un enrutador que se puede alcanzar de forma directa.

Enrutamiento

- Las entradas en las tablas indican la vía para llegar a las REDES y no entradas para cada HOST de las redes. (Excepcionalmente de manera estática se definen entradas para hosts).

Escoger rutas basado sólo en identificador de red determina:

- Todo el tráfico destinado a una red toma el mismo camino sin importar el retraso o la generación de salida de las redes físicas, aunque existan trayectorias alternativas.
- Sólo el enrutadorque entrega al destino final puede determinar si el host existe y si está en operación. Deben utilizarse mecanismos para indicar a la fuente posibles mensajes de errores la entrega.
- El enrutamiento del tráfico en sentido contrario se realiza de forma independiente, los datagramas de A a B pueden seguir caminos distintos a los de B a A.

Rutas por omisión

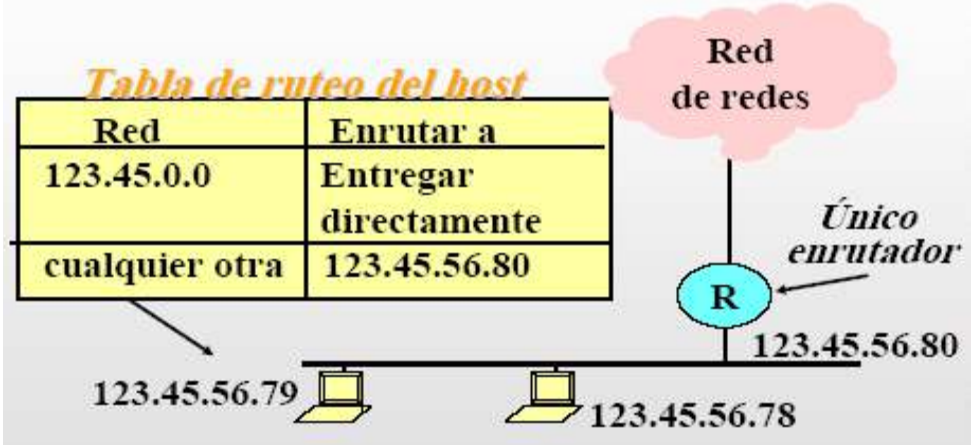
Rutas asignadas por omisión

- Para mantener reducidas las tablas de enrutamiento se emplea la técnica de “ocultar información”

- Una variante es asociar muchos registros a un enrutador asignado por omisión.
- El software IP busca primero en la tabla de enrutamiento para encontrar la red de destino.
- Si no aparece una ruta en la tabla las rutinas de enrutamiento envían el datagrama un enrutador asignado por omisión.
- El enrutador asignado por omisión es de gran ayuda cuando un sitio tiene pocas direcciones locales y sólo una conexión con el resto de la red de redes.
- Las rutas asignadas por omisión trabajan bien en hosts que se conectan a una sola red física y alcanzan sólo un enrutador.
- La decisión de enrutamiento del host consiste seleccionar una de dos alternativas :

  -entrega directa

  -entrega al router asignado por omisión.
- Aún si el sitio sólo contiene unas cuantas redes locales, el enrutamiento es sencillo. Consiste en pocas comprobaciones de las redes locales más un valor asignado por omisión para todos los demás destinos.



Rutas por host específico

- El enrutamiento está basado fundamentalmente en direcciones de redes y no de hosts individuales.
- No obstante, la mayor parte de las implementaciones de ruteo IP permiten que se especifiquen rutas por hosts.
- Esta variante permite mayor control sobre el uso de la red y hacer pruebas de depuración de tablas de enrutamiento.

Algoritmo de enrutamiento
Pasos

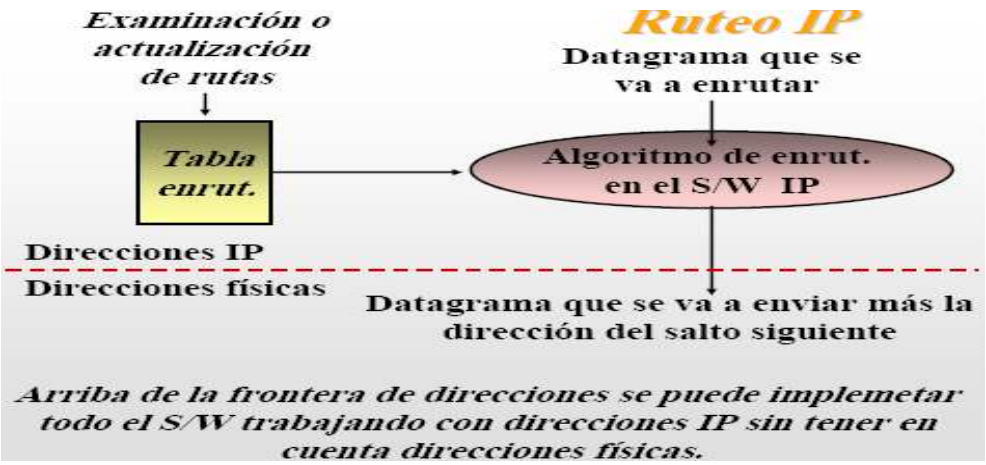
- Extraer la dirección IP de destino, D, del datagrama y computar el prefijo de red N.
- Si N corresponde a cualquier dirección de red directamente conectada entregar el datagrama a D. Esto equivale a la transformar D en dirección física, encapsular el datagrama y enviar la trama.
- Si lo anterior no se cumple ver si la tabla contiene una ruta con host específico para D, enviar el datagrama al salto siguiente especificado en la tabla.
- Si lo anterior no se cumple ver si la tabla contiene una ruta para la red N, enviar el datagrama al salto siguiente especificado en la tabla.
- Si lo anterior no se cumple ver si la tabla contiene una ruta asignada por omisión, enviar el datagrama al enrutador asignado por omisión especificado en la tabla.
- Si lo anterior no se cumple, declarar error de ruteo.

Enrutamiento con direcciones IP

- El enrutamiento IP no altera el datagrama original, sólo le disminuye el tiempo de vida y recalcula la suma de verificación.
- Las direcciones fuente y destino del datagrama permanecen inalterables.
- Una vez determinada la dirección IP del salto siguiente, se pasa esta dirección y el datagrama al software de interfaz de red, responsable de la red física.
- El software transforma la dirección IP en dirección física y conforma la trama.

¿Por qué las tablas de enrutamiento no almacenan directamente direcciones físicas?

- Primero: La tabla de enrutamiento proporciona una interfaz muy transparente para el S/W IP.
- Segundo: Para depurar problemas de enrutamiento los administradores de red necesitan examinar las tablas de enrutamiento y las direcciones IP se lo facilitan.
- El sentido del protocolo IP es construir una abstracción que oculte los detalles de las redes subyacentes.



Manejo de datagramas entrantes

- El datagrama entrante a un host es entregado al S/W IP para su procesamiento.
- Si la dirección IP de destino coincide con la IP del host o trae dirección de difusión, el datagrama es pasado al nivel superior correspondiente.
- Si la dirección IP del campo destino no corresponde con la del host, éste descarta el datagrama.
- Un host pudiera configurarse para que hiciera funciones de enrutamiento aunque no es una buena solución.
- Los hosts no deben re direccionarlos datagramas que por error le son remitidos.

Razones por las que un host no debe realizar funciones de debe realizar funciones de enrutamiento

- Cuando un host recibe un datagrama que es para otra máquina es porque existió problema con el enrutamiento; el problema no se detectará si el host en rutael datagrama.
- Los enrutadores pueden dar mensajes de error y tienen mecanismos para actualizar las tablas de rutas mientras que los hosts no poseen recursos para actualizar las tablas de ruteo.

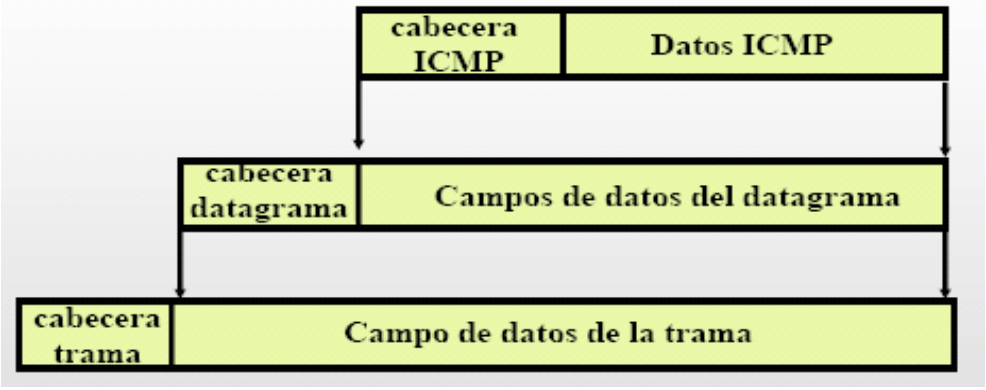
ICMP Protocolo de Mensajes de Control de Internet

ICMP

- Es un mecanismo de reporte de errores.
- Permite que los routers envíen mensajes de control y de error a otros routers y hosts.
- Lo necesita un enrutador para informar al origen cuando no puede enlutar o entregar un datagrama (máquina destino desconectada, contador de

tiempo expira...), o cuando detecta una condición anormal que afecta su capacidad de direccionamiento (congestión).

- Utilizado por la administración de las redes para detectar problemas.
- No especifica la acción a tomar frente a cada error.
- Reporta problemas a la fuente original aunque esta no tenga responsabilidad sobre el problema, ni control sobre la causa del mismo.
- Esto se debe a que el datagrama lo que lleva en su cabecera es la dirección de la fuente origen y del destino final.
- No se considera un protocolo de alto nivel sino parte obligatoria de IP.
- Los paquetes viajan por la red en la porción de datos de los datagramas IP.
- Pero su destino final, no es un software de aplicación, sino el software IP de la máquina destino.



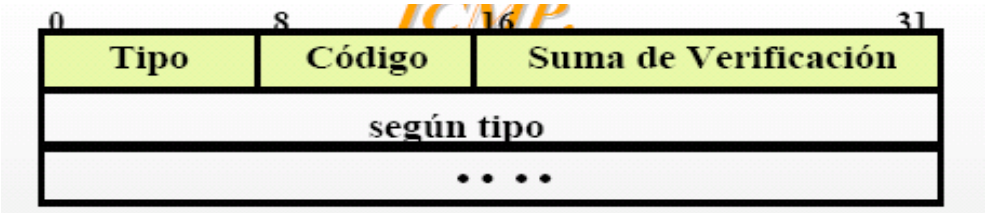
ICMP

- Los host pueden enviar mensajes ICMP a un enrutadoro a otro host.
- Los mensajes que viajan dentro de los datagramas IP, no poseen ni confiabilidad ni prioridad adicionales por lo que se pueden perder o descartar.
- No se generan mensajes por error en los propios mensajes ICMP. Esto se hace para evitar tener mensajes de error sobre mensajes de error que puedan provocar congestionamiento.

Tipos de Mensajes ICMP

- Respuesta de Eco.
- Destino inaccesible.
- Disminución de origen.
- Redireccionar (cambiar una ruta).
- Solicitud de Eco.
- Tiempo excedido para un datagrama.

- Problemas de parámetros en un datagrama.
- Solicitud de time-stamp.
- Respuesta de time-stamp.
- Solicitud de máscara de dirección.
- Respuesta de máscara de dirección.

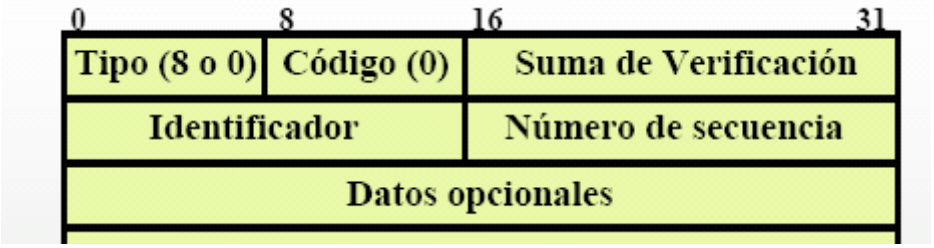


- Cada mensaje ICMP tiene su propio formato.
- Todos comienzan con los campos Tipo, Código y Suma de verificación.
- Tipo = Tipo de mensaje ICMP.
- Código= Proporciona más información sobre el tipo.
- Suma de verificación= Verifica el mensaje ICMP.
- El resto de los campos depende del tipo de mensaje.

Prueba de accesibilidad y estado de un destino

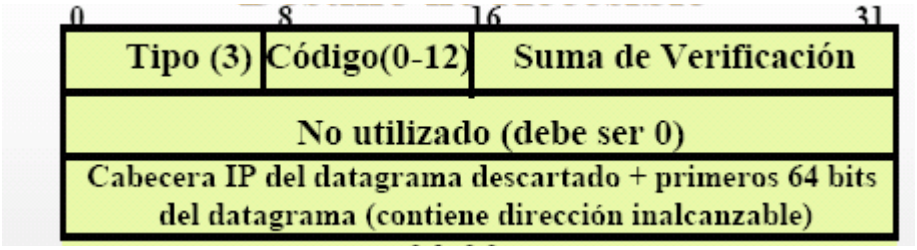
- Mensaje ICMP de solicitud y respuesta de eco.
- Utilizado para comprobar si el destino es alcanzable y si responde.
- Cualquier enrutador o host puede emitir un mensaje de este tipo a un destino específico.
- Cuando una máquina recibe un mensaje de solicitud de eco, responde con una respuesta de eco en donde va una copia de los datos enviados en la solicitud.
- Tanto la solicitud como la respuesta de eco viajan en un datagrama IP.
- La recepción exitosa de una respuesta verifica que el sistema de transporte esta funcionando bien pues:
    1. El software IP en la computadora origen debe enlutar el datagrama.
    2. Los enrutadores intermedios deben funcionar bien y tener rutas correctas.
    3. La máquina destino debe estar funcionando.
    4. Tanto el software IP como el ICMP deben estar funcionando.

Solicitud de Eco Y Respuesta de Eco



- El campo TIPO del formato permite especificar si es una solicitud o una respuesta.
- Identificador y Número de Secuencia son usados para responder a las solicitudes.
- Datos Opcionales tiene longitud variable y en él se incluyen datos que regresarán al transmisor en la respuesta de eco.

Mensaje ICMP: Reporte de Destino no Accesible



- Lo envía un enrutador, al origen, siempre que no pueda direccional o entregar un datagrama y antes de descartar el mismo.
- Código: Describe detalles del problema.
- A pesar de las opciones posibles, los enrutadores no pueden detectar la totalidad de todos los errores que ocurren.

Algunos ejemplos de códigos y su significado:
0==> Red inaccesible
1==>Anfitrión inaccesible
2==>Protocolo inaccesible
3==>Puerto inaccesible
4==> Necesario fragmentar y configuración NF
.....*****

Congestión de los Routers

- Los enrutadores se puedan saturar con el tráfico; se congestionan por :
- Computadora de alta velocidad generando más tráfico de lo que la red puede transferir.
- Muchas computadoras enviando datagramas al mismo tiempo a través de un solo enrutador.

- El descarte de cada datagrama por este motivo genera un mensaje ICMP de Disminución de Tasa de Origen para reportar congestión al origen. Otros generan el mensaje ante condiciones inminentes de descarte.
- Cuando la fuente recibe un mensaje de este tipo disminuye la velocidad de entrega de datagramas.

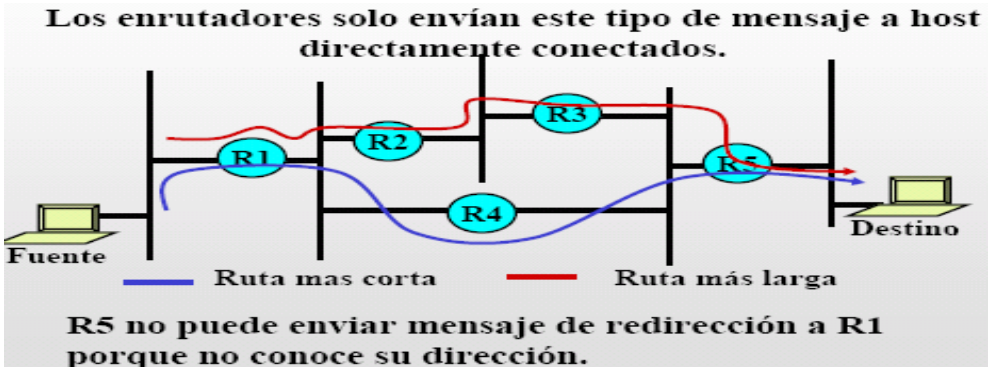
Formato ICMP de disminución de origen

Tipo (4)	Código (0)	Suma de Verificación
No utilizado (debe ser cero)		
cabecera de datagrama IP + primeros 64 bits datos		

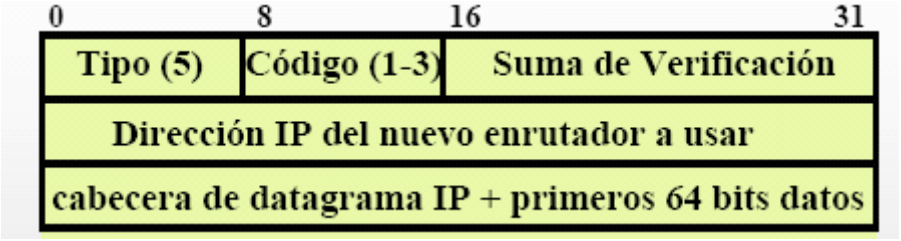
El mensaje ICMP refleja los datos del datagrama que activó la solicitud de disminución de origen.

Mensaje ICMP: Redirección o Redirección o cambio de ruta desde un router

- Se usa para que un enrutador, que detecte que un host conectado directamente a élestá usando una ruta incorrecta, (debido a cambios en la topología en la red, a reparaciones de hardware, etc.) le avise al mismo que existe una mejor ruta.
- Da la posibilidad de que los host comiencen a trabajar con información mínima de ruteo en sus tablas y las mantengan reducidas pues confían en que los enrutadores, que conocen las rutas correctas, les avisen si deben hacer algún cambio.
- Este mensaje ICMP es enviado por los enrutadores solo a los host directamente conectados y no a otros Reuters. A pesar de todo el enrutador en ruta el datagrama.



Formato de mensaje ICMP de Re direccionamiento



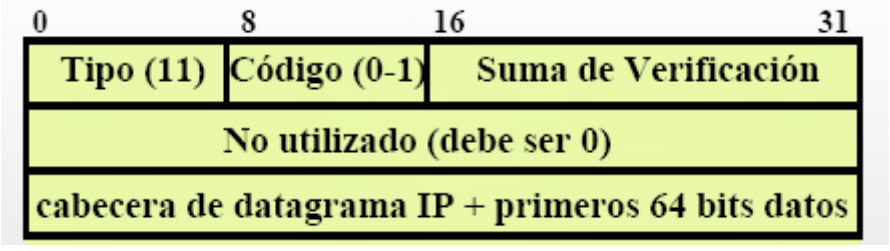
- Código: permite interpretar mejor la dirección del nuevo enrutador.
- Dirección IP del nuevo enrutadora usar: Modificación que el host debe hacer en su tabla de ruta.
- Cabecera de datagrama: Especifica el datagrama que activó este tipo de mensaje.

Código
1==> Redireccionardatagrama para el anfitrión.
2==> Redireccionardatagrama para la red y tipo de servicio.
3==> Redireccionardatagrama para el anfitrión y tipo de servicio.

Mensaje ICMP: Tiempo excedido para un datagrama

- Se envía siempre que un enrutadordescarta un datagrama, ya sea porque su conteo de saltos (tiempo de vida) llega a cero o porque expira el temporizador de espera para el reensamblado de un datagrama.
- En el formato del datagrama, el campo CÓDIGOindicará a que causa en específico se debe el descarte.
- Una de las causas que hace que el tiempo de vida de un datagrama llegue a su fin, es que por la existencia de errores en las tablas de enrutamiento, este entre en un ciclo de enrutamiento en el cual recorrerá continuamente un conjunto de enrutadoreshasta que su conteo de saltos llegue al final.

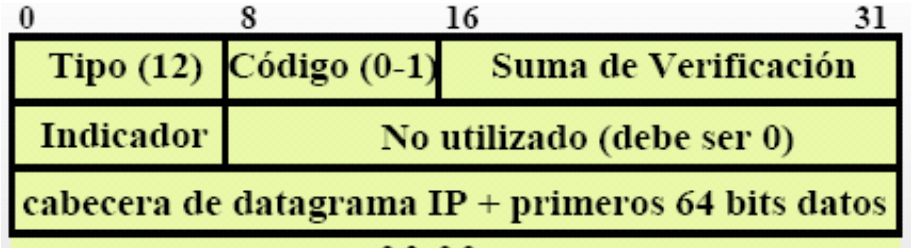
Formato de ICMP de tiempo excedido



- Código = 0: Tiempo de vida excedido
- Código = 1: Tiempo para el reensamblado de fragmentos excedido.

Mensaje ICMP: Problemas de parámetros en un datagrama.

- Se envía cuando un enrutadoro un host encuentran problemas no contemplados en el resto de los mensajes ICMP (Ej: datagrama con el encabezado incorrecto debido a qué los argumentos para una opción son incorrectos).
- Solo se envía cuando el problema es tan severo que se tiene que descartar el datagrama.
- En su formato posee el campo INDICADOR, que permite identificar el octeto del datagrama que causó el problema.
- El campo CÓDIGO en "1" indica que lo ocurrido es una falta de opción requerida y entonces no es necesario usar el campo POINTER.

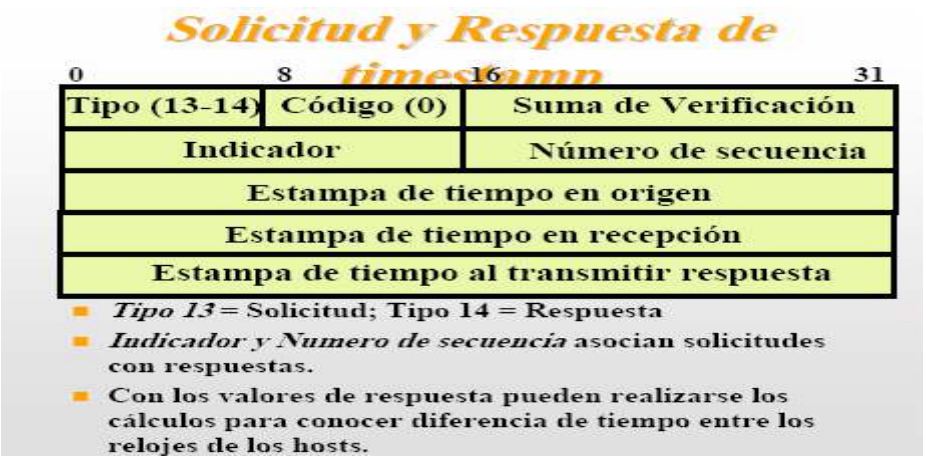


- Indicador: Puntero que indica el octeto del encabezado que origina el problema.
- Código = 0, octeto erróneo.
- Código = 1, Falta de opción requerida

Mensajes ICMP: Solicitud y Respuesta de estampa de tiempo

- Permite obtener la hora de otra máquina al enviarle la solicitud de estampa de tiempo y recibir de esta la respuesta de estampa de tiempo.
- Así dos máquinas pueden sincronizar sus relojes lo cual es importante en algunas aplicaciones (sistemas distribuidos).
- En su formato posee tres campos que permiten especificar la hora universal en milisegundos del:
  - -momento en que se envió el mensaje de solicitud.
  - -momento en que se recibe el mensaje de solicitud.
  - -momento en que se envía el mensaje de respuesta.

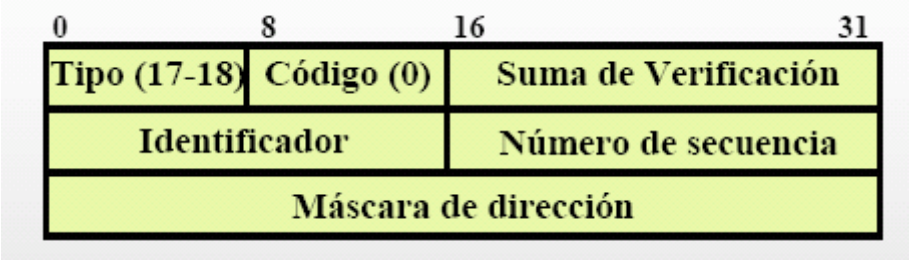
Solicitud y respuesta de time ICMP



Mensajes ICMP: Solicitud y Respuesta de máscara de red

- Cuando se utiliza direccionamiento de subred, una máscara de red permite a un host saber qué bits de su dirección IP identifican a la red física y cuales corresponden a los identificadores del host.
- Con estos mensajes ICMP una máquina le pregunta a un enrutadorpor su máscara y obtiene una respuesta del mismo.
- Si la máquina no conoce la dirección de un enrutadorque le de esta información, puede transmitir un mensaje por difusión.

Formato ICMP: Solicitud y Respuesta de mascara de red



- TIPO: indica si es una solicitud o una respuesta.
- IDENTIFICADOR y NÚMERO DE SECUENCIA: para que la máquina asocie las solicitudes con las respuestas.
- DIRECCIÓN DE MÁSCARA: usado en la respuesta, para colocar la máscara de la dirección de subred.

Funciones más importantes del ICMP

- Control de Flujo. (Disminución de los mensajes de la fuente).
- Detección de Destinos no alcanzables (mensajes de destinos no alcanzables).
- Redireccionamiento de rutas (mensajes de redireccionamiento ICMP).
- Chequeo de host remotos (solicitud de eco, respuesta de eco, Ej. Ping)

Bibliografía

Abad, R., Medina, M., Careaga, A. (1993). Fundamentos de las estructuras de datos relacionales. Grupo Noriega Editores, Limusa. México.

A. Silberschatz, H. F. Korth y S. Sudarshan, Fundamentos de Bases de Datos, España: MacGraw Hill, 2002.

Camuña, J., (2014). Lenguaje de definición y modificación de datos SQL [Versión electrónica]. (1ra. ed.). IC Editorial, Málaga.

C. J. Date, Introducción a los Sistemas de Bases de Datos, México: Pearson Educación, 2001

Connolly, T., Begg, C. (2005). Sistemas de bases de datos. Un enfoque práctico para diseño, implementación y gestión (4ta. ed.). Pearson, Madrid.

De Miguel, A., Martínez, P., Castro, E., Cavero, J., Cuadra, D., Iglesias, A., Nieto, C. (2005). Diseño de bases de datos, Problemas resueltos. Alfaomega Ra-Ma. México.

Elmasri, R., Navathe, S. (2007). Fundamentos de sistemas de bases de datos (5ta. Ed.). Pearson. Madrid.

Groff, J., Weinberg, P., (2003). SQL manual de referencia. McGraw-Hill. Madrid.

Ibañez, l. (2014). Gestión de bases de datos (2da. ed.). Ra-Ma. Madrid.

Jiménez, M. (2014). Bases de datos relacionales y modelado de datos, [Versión electrónica]. IC Editorial.

Koutchouk, M. (1992). SQL et DB2 le relationnel et sa practique (2da. ed.). Masson. Paris.

M. V. N. Cabello, Introducción a la Base de Datos Relacionales, Madrid: Visión Libros, 2010.

M. Y. Jimenez Capel, IFCT0310: Bases de datos relacionales y modelado de datos., Malaga: IC Editorial, 2015.

Piattini, M., Clavo, J., Cervera, J., Fernández, L., (1996) Análisis y diseño detallado de aplicaciones informáticas de gestión. Ra-Ma, Madrid.

Silberschatz, A., Korth, H., Sudarshan, S. (2014). Fundamentos de bases de datos (6ta. ed.). McGraw-Hill. Madrid.

V. F. Alarcón, Desarrollo de sistemas de información: una metodología basada en el modelado., Catalunya: Edicions UPC, 2010.

Ingeniería y Tecnología

3ciencias