Ethernet

(TDC, FCEN UBA. PROTOTIPO)

Introducción

Ethernet es un sistema de red de area local, basado en normas de la IEEE (o en variantes de las mismas). Tradicionalmente se centra en las normas 802.3, aunque también están involucradas otras de la serie 802.

Las capas de nivel de enlace de LAN de múltiple acceso están usualmente divididas en la subcapa MAC (Media Access Control) y LLC. Ethernet no es excepción a esta regla, aunque hay ciertos casos en donde la capa LLC es eliminada. Usualmente la capa MAC está asociada a un medio físico en particular, mientras que LLC es independiente de la misma (pudiendo aplicarse sobre otras tecnologías similares, como 802.5, Token Ring, o 802.11).

Medios físicos

Originalmente el medio físico de ethernet eran dos tipos de cable coaxil (coaxil grueso, 10-BASE-5 y coaxil fino, 10-BASE-2). Estos dos tipos de cable son capaces de transportar datos a 10Mbps a 500 y 185 metros respectivamente, parámetros de los que derivan los nombres. Luego se introdujo ethernet sobre UTP (10-BASE-T, 100-BASE-T, 1000-BASE-T) y fibra (10-BASE-F, 100-BASE-F, 1000-BASE-S). En el protocolo original se utiliza una codificación Manchester, lo que significa que en el medio hay 20 millones de cambios por segundo. En los estándares de mayor velocidad se sustituyó la codificación Manchester por 4B/5B u 8B/10B, lo que significa que el medio debe poder enviar una señal 20% más rápida que la velocidad de transferencia.
Puesto que no se espera que un segmento de cable transfiera una señal 802.3 a lo largo de 2500 metros, se pueden usar repetidores para incrementar el largo de la red. Un repetidor es un dispositivo eléctrico (de nivel físico) que toma y amplifica la señal recibida y la retransmite. Un amplificador que toma la señal desde una entrada y la amplifica hacia múltiples salidas es un HUB. El hub no es capaz de entender que esa señal es un frame, a diferencia de un switch, que es un dispositivo de nivel 2. Por ahora se supondrá que no hay switches involucrados.
Cuando el medio es compartido (como sucede con el cable coaxil o en ciertas condiciones con UTP), hay limitaciones de largo del cable. En total, el medio no puede medir más de 2500 metros, en 5 segmentos, con 4 repetidores (o hubs), con 3 de ellos poblados. Si se trata de una red Fast Ethernet, esta distancia pasa a ser 250 metros, pudiendo tener como máximo dos hubs.

Acceso al medio

La base de ethernet se remonta al año 1970 en Hawaii, donde se desarrolló un sistema de comunicaciones satelital llamado ALOHA. El sistema consistía basicamente en dejar que las estaciones transmitieran libremente, dejando que el satélite reflejara las señales, para luego detectar en tierra si dos estaciones habían transmitido simultaneamente (lo que destruye ambos mensajes). Este protocolo era extremadamente ineficiente, ya que las estaciones emisoras no tenían la capacidad de detectar si el medio estaba libre o no, por lo que era extremadamente factible que dos estaciones se interfirieran

mutuamente. En 1973 Bob Metcalfe diseñó un protocolo similar para la transmisión via cable que incluia detección de colisiones. Este protocolo es CSMA/CD.

Antes de entender la relación entre ALOHA y ethernet hay que tener en cuenta que un cable de 2500 metros las señales no se comportan como uno esperaría, sino que se parecen a ondas en un estanque. La reflexión de una señal puede interferir con la misma, asi como es posible que haya una condición de error en un segmento del cable sin que el otro extremo la haya detectado.

Esto se puede ver al transmitir un mensaje. Si dos estaciones intentan una transmisión simultánea, ambas pensarán que tienen el canal libre, para luego darse cuenta que no era así. Esto se llama una "colisión" y el protocolo CSMA/CD es el encargado de resolver este tipo de conflictos. Fijense que esto es un problema similar (pero más acotado) que el que sufría ALOHA.

CSMA/CD significa Carrier Sense, Multiple Access/Collision Detect. Esto significa que muchas estaciones comparten un mismo medio (MA), que al transmitir sensan el medio para ver si está siendo usado (CS) y que mientras transmite escucha el medio para ver si hay interferencia de otra estación (CD).

En el cable de 2500 metros entran 32 bytes (256 bits). Cuando una estación quiere transmitir, primero verifica que el canal esté libre. Una vez verificado, transmite, pero escucha durante un período para verificar que no suceda una colisión. Este período es 51,2 us, el doble del tiempo de propagación. Aunque este valor no parece ser muy intuitivo, es importante considerar que una colisión tambien tarda en propagarse. Imaginemos dos estaciones en puntas opuestas del medio. Si una estación detecta que el medio esta vacío y comienza a transmitir, la segunda estación no vera la señal hasta que hayan pasado 25,6 us. Si un instante epsilon antes de que llegue la señal esta otra estación decide transmitir, se producirá una colisión, que tardara 25,6 us en llegar a la otra punta del medio (para ser detectada por la estación original. Esto define una longitud mínima de 64 bytes. Si este límite inferior no se respetara, una estación podría terminar de enviar el frame y desconectarse antes de que la interferencia se propague hasta dicha estación. En ese caso, la estación considerará enviado el frame cuando en realidad no es así. Para evitar que una estación acapare el medio, el estandar define una longitud máxima de 1500 bytes.

El protocolo CSMA/CD no elimina las colisiones, estas son normales y ocurren todo el tiempo, sino que permite corregir el problema cuando una aparece. Para esto existe el algoritmo de exponential backoff. El algoritmo consiste en enviar el mensaje no inmediatamente, sino con un retraso elegido al azar. El número de retransmisiones y la espera máxima están acotadas para evitar que un emisor espere un tiempo excesivo (o eternamente), ya que en esos casos suele ser más conveniente reportar un error a retrasar las comunicaciones del nivel superior.

```
i=1
        #Numero de reintentos
        while ( i <= 16 )
                #Cota máxima de espera – Crecimiento exponencial, limitado a 2^10 - 1
                maximo = min(1023,2^i – 1)
                slot = random(0...maximo)
                status = transmitir(mensaje,slot)
                if (status = exito)
                        then terminar_backoff(exito)
                fi
                i = i + 1
        endwhile
```

```
terminar_backoff(fracaso)
```

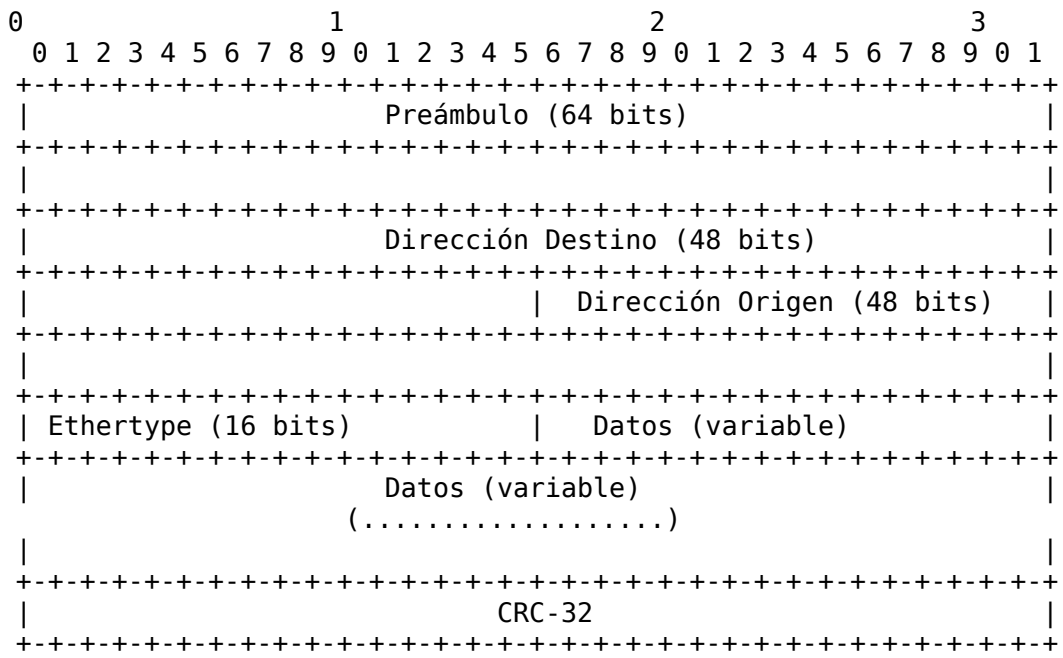
Este algoritmo introduce una variación aleatoria en los tiempos de transmisión, lo que (probabilisticamente) resuelve la colisión.

Cuando el canal se libera es necesario considerar que hará una estación que tiene datos a transmitir. Intuitivamente, el camino más lógico a seguir sería transmitir inmediatamente, pero si dos estaciones tienen datos para enviar, habrá una colisión inmediatamente despues. Otra forma de actuar sería postergar la transmisión basandose en el azar, para mejorar las posibilidades de que uno de los mensajes sea transmitido exitosamente. Es de particular interés crear intervalos de tiempo similares a los slots de exponential backoff y transmitir con una probabilidad p, o posponer el mensaje al próximo intervalo (y repetir este algoritmo) con una probabilidad (1-p). Este concepto se llama p-persistencia. Tener un valor pequeño de p mejora la performance cuando la red está cargada, en detrimento de la misma cuando la red tiene poca carga. Ethernet es 1-persistente, o sea siempre transmite y si es necesario colisiona.

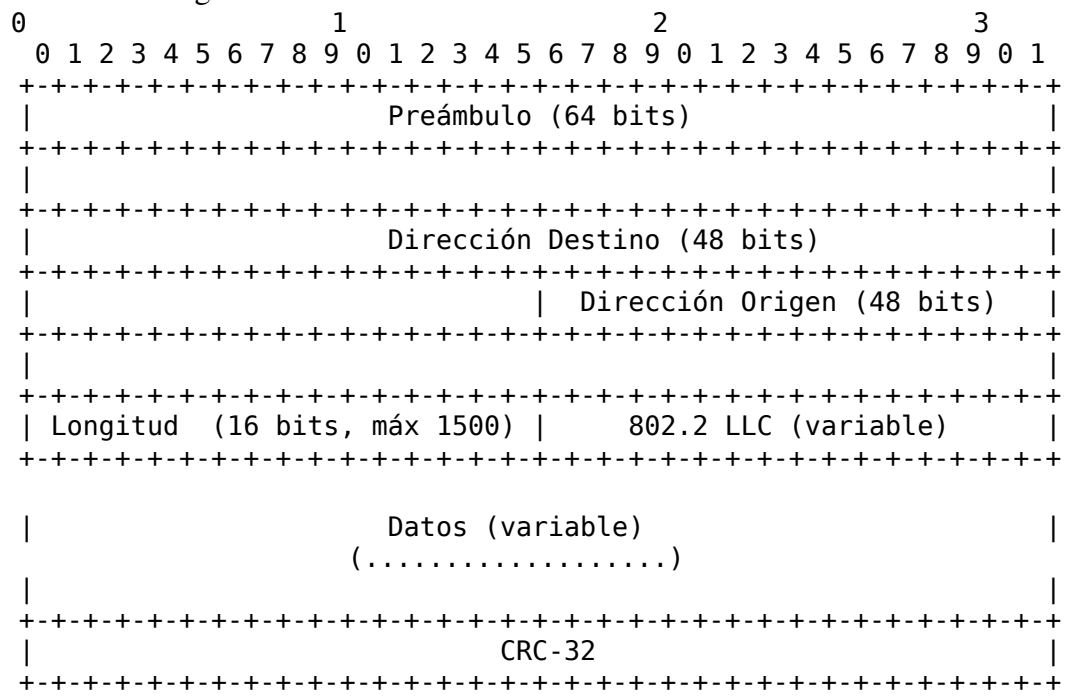
Formato de trama

Hay varias variantes de formato de lo que se conoce hoy en día como ethernet. Todas estas variantes se basan en los mecanismos ya mencionados para controlar el acceso al medio, y difieren principalmente en el formato de la trama. Los dos formatos son el formato Ethernet II y el formato IEEE 802.3. Estos formatos son casi idénticos, pero 802.3 tiene más capacidades (y más complejidad) por la presencia de un segundo juego de encabezados (que esta ausente en Ethernet II). Una trama 802.3 se puede simplificar usando un formato llamado SNAP, que brinda un servicio similar a Ethernet II en una trama 802.3. Todos los formatos 802.3 brindan un servicio no confiable, no orientado a conexión en su capa MAC. Esto puede ser complementado por su capa LLC.

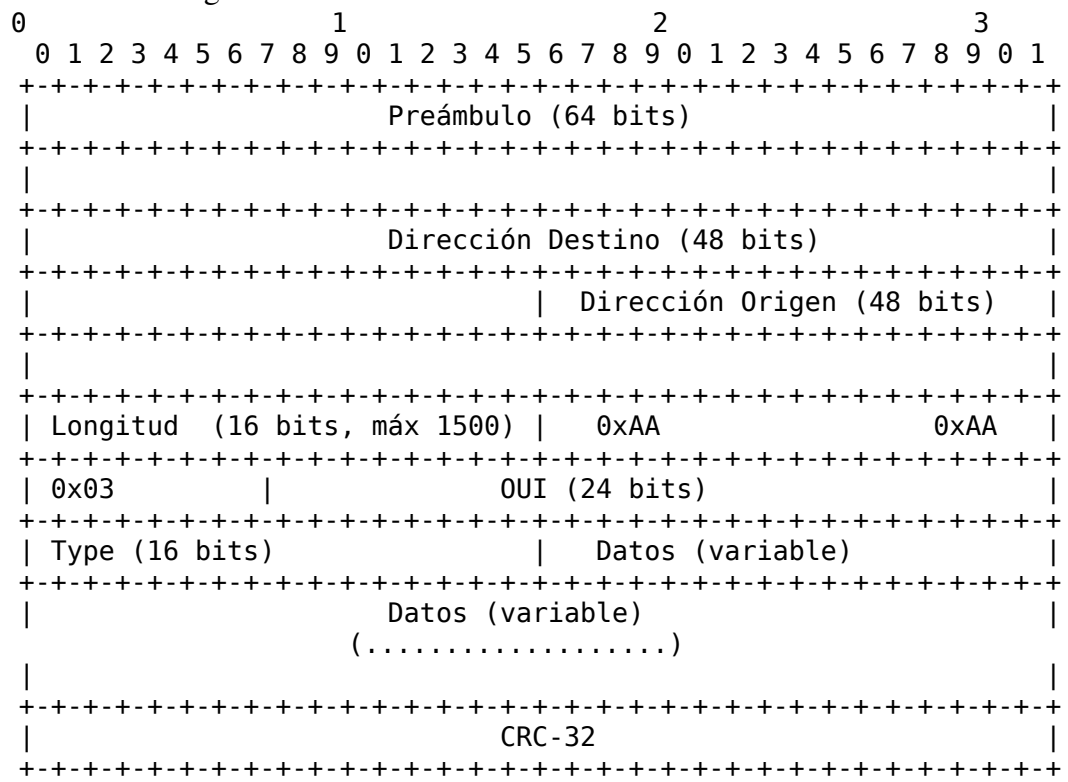
Una trama Ethernet II tiene los siguientes campos



Una trama 802.3 tiene el siguiente formato



Una trama SNAP tiene el siguiente formato



El preámbulo es una serie de bits alternantes(01010101), usado para determinar donde comienza un nuevo frame. Las estaciones receptoras detectan este patrón y pueden sincronizar los relojes. Este patrón se interrumpe en el último byte, transmitiendo en su lugar la secuencia 01010111, lo que indica el comienzo del encabezado.

Despues del preámbulo vienen las direcciones de destino y origen en ese orden. Estas estan presentadas en este orden ya que esto permite que una placa no reciba un frame que no está destinado hacia ella temprano en la lectura del mismo, lo que ahorra recursos en la misma (ya que no necesita procesar campos adicionales). Cabe aclarar que es posible forzar la recepción de todos los frames y pasarlos al sistema operativo, usando el llamado “modo promiscuo”, sin embargo esta forma de operación no es la usual.

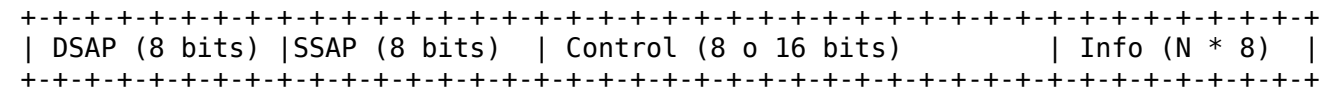
Las direcciones de origen y destino son únicas y están formadas por dos componentes: El OUI (Organisationally Unique Identifier) en los tres primeros bytes, los cuales representan cada uno a un fabricante y un número de serie único por fabricante en los tres últimos bytes. La dirección tiene este formato para evitar que se construyan placas de red con la misma dirección MAC. Usualmente las direcciones se escriben como una serie de seis valores hexadecimales, y es posible (y bastante usual) buscar el fabricante de una placa por medio del OUI. Si bien es posible cambiar la dirección MAC a una placa, no es algo recomendable, ya que bajo circunstancias normales usar la dirección MAC por defecto no afecta el normal funcionamiento de la red, al mismo tiempo que asegura que no haya duplicados.

Una dirección especial es la dirección de broadcast (FF:FF:FF:FF:FF:FF) formada por todos unos, usada para dirigirse a todos los miembros de una LAN, otro tipo especial de direcciones son las direcciones multicast, que se usan para comunicarse con un grupo de máquinas (de los cuales no hablaremos en este curso).

Inmediatamente despues de la dirección origen se encuentra un campo que define si se trata de un frame Ethernet II o 802.3. Este campo es el ethertype o la longitud. Si este campo es menor o igual a

1500, se asume que es una longitud. Si es mayor, se asume un ethertype (que indica el protocolo de nivel superior al que se debe pasar el mensaje). El objetivo del ethertype (o de cualquier campo que mencione al protocolo superior) es permitir que varios protocolos de nivel superior coexistan sin que haya dudas de a quien le corresponde un dato. Este campo existe para no tener que analizar el contenido del campo de datos, lo que sería una violación seria al concepto de capas, asi como inmantenible en la práctica.
En un frame Ethernet II siguen inmediatamente los datos. Si el emisor aplica un padding (para llegar al tamaño mínimo de frame), el receptor no puede determinarlo. En este caso es responsabilidad del nivel superior determinar donde terminan sus datos y empieza el relleno.
En un frame 802.3 sigue un frame LLC (SNAP es un tipo de frame LLC válido). La subcapa LLC es la encargada de ofrecer un servicio similar aunque las capas MAC sean distintas (esto no es tan relevante cuando casi todos usan Ethernet, pero fue pensado para unificar distintos tipos de subcapas MAC bajo una misma interfaz). Usando LLC, es posible simular distintos servicios que la capa MAC no brindaría (por ejemplo: servicio orientado a conexion sobre 802.3).

LLC tiene el siguiente formato



Donde DSAP y SSAP son valores similares al ethertype para el receptor y emisor respectivamente, con bits reservados para saber si es un comando o una respuesta en el SSAP, o si el destino es individual o grupal en el DSAP. Los otros campos son dependientes de estos dos.

Operación full duplex

Cuando se está trabajando con UTP o fibra existen dos canales, fibras o pares de cobre, separados, que se usan para transmitir y recibir respectivamente. Debido a ello, es posible apagar el algoritmo de control de acceso al medio y trabajar con ambos canales independientemente (transmitiendo constantemente en uno y recibiendo en el otro). A este modo se lo llama “full duplex”. Es posible trabajar en este modo solamente si se puede garantizar la exclusividad al medio, ya sea por tratarse de una conexión directa entre hosts, o por haber un dispositivo llamado switch en el medio.Las versiones de 802.3 con velocidades de Gigabit o superiores requieren el uso de este modo. En estos casos también es común usar frames más largos que 1500 bytes, llamados “jumbo frames” (ya que no hay competidores a los que estemos dejando fuera del medio por usarlos).

Switches

Un swich (también conocido como bridge) es un dispositivo de nivel de enlace con varios puertos y capaz de entender los frames que pasan por el, recibiéndolos y forwardeándolos según sea necesario, en base a tablas que indican donde se encuentran las estaciones pertenecientes a la red. Puesto que un switch es capaz de entender los frames a nivel de enlace, este forma una conexión full duplex con cada estación conectada en sus puertos (a menos que en un puerto en particular sea imposible, por ejemplo por estar conectado a un hub). La presencia del switch causa dos efectos importantes en la red: divide el dominio de colisión y cambia la forma en el cual los frames viajan por la LAN.

Puesto que el switch es un dispositivo que entiende el significado de los frames, y trabaja (usualmente) en un modo store and forward, este puede bloquear la propagación de colisiones. Esto significa que una colisión en un puerto no es vista por una estación en otro. Es importante aclarar que aunque el switch divide el dominio de colisiones, no divide el dominio de broadcast, o sea que un broadcast emitido en uno de sus puertos es visto en toda la LAN.

Debido a que un switch conoce cierta información de la red, es capaz de transmitir (en la mayoría de los casos) tráfico a un destino que conoce por el puerto al cual esta estación está conectada, evitando que otras estaciones vean innecesariamente estos frames. Esto permite el establecimiento de múltiples flujos de información usando al máximo los recursos de cada estación involucrada, siempre que no exista un recurso en conflicto (por ejemplo dos estaciones queriendo transmitir a una tercera simultáneamente).

Basandose en estos dos principios, un switch puede intentar optimizar el uso de la red cambiando su modo de operación (siendo “store and forward” el modo usual). Un switch puede, una vez obtenida la dirección destino del frame y determinado que el puerto destino está libre, conectar directamente el puerto de origen (por donde entra el frame) con el puerto destino. Esto evita la necesidad de almacenar el frame (salvo una pequeña parte, que puede quedar en el buffer propio del puerto). Este modo es conocido como “pass thru” A pesar de sus ventajas, este modo no permite la inspección o edición del frame, por lo que el switch podría verse obligado a forwardear un frame corrupto (o peor aún, una colisión). Un segundo modo es llamado “collision free”. Este modo es similar a pass thru, pero no realiza el forward hasta haber recibido correctamente 64 bytes del frame. Esto evita el forwardeo de una colisión (ya que debería haber sido detectado en ese momento), pero aún es posible que pase un frame corrupto. Este modo no es útil si toda la red trabaja en modo full duplex, ya que no es posible tener una colisión bajo esas circunstancias.