

Auto-negociación en Ethernet

Artículo cedido por Next For S.A.,



En las redes de ordenadores de hoy en día, la auto-negociación es tan importante como la tecnología plug-and-play. La auto-negociación es un algoritmo definido en la sección 28 del estándar IEE 802.3 e introducido por primera vez en 1997 como parte del estándar IEEE 802.3u en la Fast Ethernet. La auto-negociación se diseñó para ser también compatible con los estándares de red Ethernet originales. En 1999 se introdujeron nuevas mejoras en la auto-negociación con el nuevo estándar IEEE 802.3ab para la Gigabit Ethernet. La auto-negociación puede definirse como el mutuo acuerdo entre dos dispositivos de red que comparten un cable en cuanto a la velocidad, el dúplex y el control para hacer uso de ese cable. El protocolo de auto-negociación existe estrictamente en la capa física (PHY) del OSI (Modelo de Referencia Open System Interconnection) y se implementa por software, hardware o una mezcla de ambas, y en este artículo vamos a explicar cómo el protocolo negocia la velocidad, el dúplex, el Auto-MDIX (terminación del cable) y el control de flujo.

Como verá en la discusión técnica que sigue, la auto-negociación es un parámetro extremadamente importante en las redes cableadas Ethernet de hoy en día. Para que un enlace funcione correctamente, los dispositivos en ambos extremos del cable deben estar configurados de la misma forma; bien ambos deben estar configurados para hacer uso de la auto-negociación o bien ambos deben tener definidos de forma fija los mismos parámetros de velocidad y dúplex. En un entorno en el que un dispositivo se encuentra configurado para auto-negociar y el otro dispositivo está configurado para usar una velocidad y dúplex fijos, el algoritmo de auto-negociación puede detectar la velocidad y fijarla correctamente, pero la configuración de dúplex del dispositivo remoto no puede ser determinado por el dispositivo realizando la auto-negociación. Y siguiendo el estándar IEEE, el dispositivo configurado para auto-negociar pasará

a usar half-duplex. Esto presenta un problema si el dispositivo remoto está configurado para full-duplex. Habitualmente en este escenario los usuarios se quejarán de conectividad lenta en la red y de timeouts en las aplicaciones. Estos síntomas se explicarán en detalle en la sección que habla del Dúplex.

Por último, debe recordarse que, de acuerdo con la especificación IEEE, el uso de la Ethernet Gigabit requiere el uso de la auto-negociación, por lo que 1000Mb/s no es una configuración fija válida en un dispositivo de red que siga fielmente las especificaciones IEEE.

Velocidad

El IEEE 802.3u introdujo el 100Mb/s a lo que anteriormente eran redes Ethernet únicamente de 10Mb/s. En el momento en que los ordenadores tenían la opción de elegir a qué velocidad querían comunicarse, hacía necesaria la introducción de un procedimiento para gestionar esta decisión. Con la introducción de una tercera velocidad, 1000Mb/s o Gigabit Ethernet, este procedimiento tomó mayor relevancia todavía. Por ello se creó el protocolo de auto-negociación y el algoritmo NWay adaptado para proporcionar una solución plug-and-play a este proceso de toma de decisión mientras se mantenía completa compatibilidad con el antiguo protocolo de 10Mb/s.

El estándar de 10Mb/s detecta un enlace activo con otro dispositivo de red a través de la transmisión y recepción de

pulsos de prueba de la integridad del enlace (Link Integrity Test - LIT) siempre que el dispositivo no está enviando o recibiendo datos de forma activa. Estos pulsos LIT o Normal Link Pulses (NLP), nombre al que posteriormente se cambió, consisten en un pulso unipolar positivo con una duración de 100ns a un intervalo de 16ms con una ventana de ± 8 ms.

El protocolo de auto-negociación introducido con el estándar de 100Mb/s, transmite un pulso de enlace rápido (FLP - Fast Link Pulse) en lugar de uno normal (NLP). Un envío FLP consiste en una serie de 33 pulsos. Cada envío de 33 pulsos tiene una duración de 2ms en total y sigue los mismos intervalos de transmisión de 16ms ± 8 ms. Los pulsos individuales son de 125 μ s con 62,5 μ s $\pm 7\mu$ s entre pulsos. Podemos observar esta temporización en la figura 1. Los pulsos individuales alternan entre pulsos de reloj y pulsos de datos correspondiendo a pulsos de reloj el primer y todos los consecutivos pulsos impares. Cada uno de los 16 pulsos de datos (con cada pulso o ausencia del mismo representando un 1 o un 0, respectivamente) consiste en un bit de datos y en conjunto forman 16bits o 2bytes de datos. Estos 2 bytes forman la palabra de código de enlace (Link Code Word - LCW) que contiene la información necesaria para la auto-negociación.

Existen múltiples formatos de LCW, pero la LCW más importante es la página base (Base Page). Esta página base es la transmisión que indica las posibili-

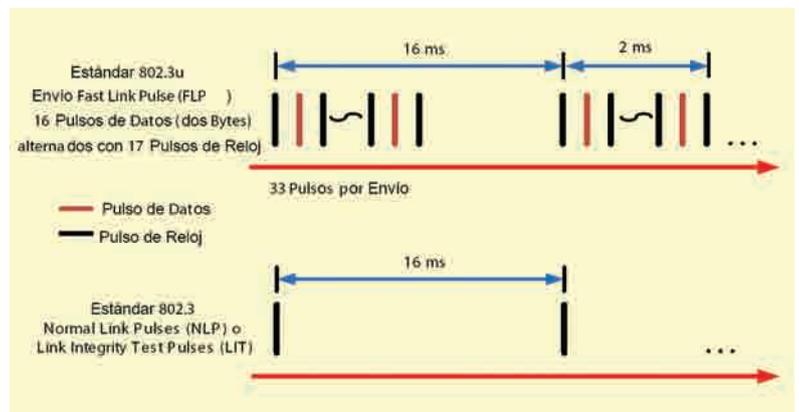


Figura 1. Temporización NLP y FLP.

dades de ese dispositivo. Los primeros 5 bits sólo tiene 2 valores válidos. Indican o bien que usa el IEE 802.3 (Ethernet) o IEEE 802.9 (IsoEthernet sobre par trenzado Cat3). Los siguientes 5 bits indican las combinaciones de velocidad y dúplex que un dispositivo puede utilizar para comunicarse. Los bits A5 y A6 se usan para el Control de Flujo y D14 se usa para confirmar una negociación. El último bit, D15 se usa para denotar la necesidad de usar la página siguiente (Next Page), que es una LCW más avanzada usada para negociar velocidades y control de Gigabit. El diagrama de la figura 2 ilustra esta página base.

Dúplex

Con la introducción del estándar de Fast Ethernet IEEE 802.3u surgió la posibilidad de comunicación simultánea bidireccional. Nuevamente se hizo necesario introducir un protocolo y un método de gestionar esta decisión. Como se comentó en la sección anterior, las negociaciones del dúplex para redes a 100Mb/s se gestionan en la palabra de código de enlace y son parte de la Next Page y en el LCW de mensaje de página para una red a 1000Mb/s.

Si no contamos los problemas de cableado físico o de fallo hardware, la

Las asignaciones incorrectas del dúplex pueden ser difíciles de identificar porque no generan una pérdida total del enlace. A menudo el enlace tendrá un rendimiento suficiente como para evitar que salten las alarmas. Este es especialmente el caso si se está haciendo un uso escaso del enlace. Sin embargo surgirán los problemas una vez que se incrementa la actividad. El problema específico es que un dispositivo half-duplex cree que sólo puede hablar un dispositivo a la vez, así que no se comunicará mientras que el otro dispositivo (full-duplex) esté hablando. El dispositivo full-duplex no se encuentra bajo estas restricciones y cree que ambos dispositivos pueden transmitir simultáneamente. Si el dispositivo half-duplex detecta, mientras está transmitiendo, una transmisión del otro dispositivo, detendrá inmediatamente su transmisión, tirará todas las transmisiones entrantes como inválidas, e iniciará un temporizador de espera para acceder al medio. Entre tanto, el dispositivo full-duplex completa su transmisión y supone su recepción. También recibirá el paquete truncado del dispositivo half-duplex, determinará que es incorrecto y marcará el contador de errores CRC (Cyclical Redundancy Check). El dispositivo half-duplex intentará una retransmisión de su paquete una vez haya vencido el temporizador, pero el dispositivo full-duplex no siente la necesidad de retransmitir, ya que desconoce que el otro equipo tiró sus paquetes) y así el dispositivo half-duplex nunca recibe los paquetes a menos que en una capa de un nivel superior OSI requiera una confirmación y genere una retransmisión.

Los síntomas de esta situación se mostrarán generalmente como un enlace de red lento o una aplicación o aplicaciones con demasiados timeouts. En una conexión correctamente configurada los errores de CRC deberían

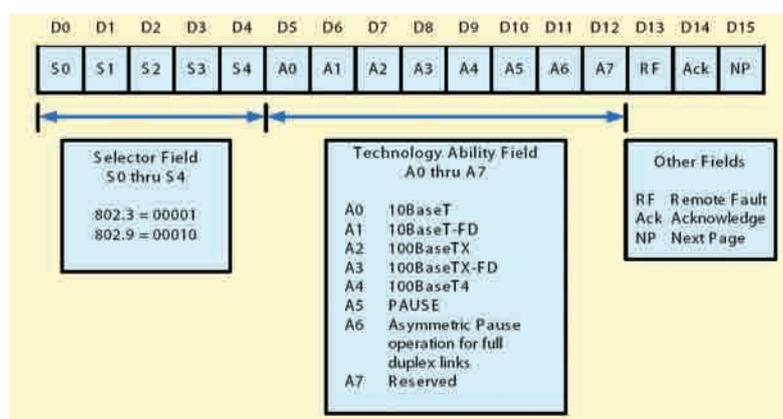


Figura 2. Página Base LCW.

Para que dos dispositivos se pongan de acuerdo en la velocidad de transmisión por el cable, deben transmitirse y recibirse seis LCWs idénticos, 3 de cada extremo. Una vez que un dispositivo ha recibido 3 LCWs idénticos contenidos en FLPs del extremo remoto del cable, el dispositivo local transmitirá un FLP con el bit de confirmación (ACK) a 1. Es importante darse cuenta de que cada dispositivo únicamente está indicando cuáles son sus propias posibilidades y por consiguiente los dos dispositivos deben usar el mismo orden de prioridad para ponerse de acuerdo en la velocidad. Este orden de prioridad es parte del estándar IEEE y puede verse en la figura 3. Una vez que ambos extremos han recibido una respuesta de confirmación, la velocidad queda acordada.

Las FLPs se diseñaron para estar en línea con las NLP de forma que un dispositivo a 10Mb/s detecte señal en la línea al intervalo habitual y sea capaz de comunicarse. Un dispositivo capaz de auto-negociación detectará la existencia de NLPs y, gracias al estándar de compatibilidad, será capaz de comunicarse con NLPs para trabajar a 10Mb/s.

asignación incorrecta del dúplex es la causa más frecuente de problemas de enlace en una red. Los errores de selección de dúplex se deben a la inhabilidad de un dispositivo con auto-negociación de predecir la configuración de un dispositivo con asignación fija del dúplex. Esto se debe a que la transmisión de FLPs está deshabilitada cuando un dispositivo tiene la configuración fija, de acuerdo con las especificaciones IEEE. También de acuerdo con las especificaciones IEEE el dispositivo en auto-negociación se conectará en half-duplex cuando no se pueda determinar la configuración dúplex del otro dispositivo.

Prioridad	Tecnología
1 Máxima	1000BaseT - Full Duplex
2	1000BaseT - Half Duplex
3	100BaseT2 - Full Duplex
4	100BaseTX - Full Duplex
5	100BaseT2 - Half Duplex
6	100BaseT4
7	100BaseTX - Half Duplex
8	10BaseT - Full Duplex
9 Mínima	10BaseT - Half Duplex

Figura 3. Tabla de resolución de prioridad.

Figura 4. Escenario con asignación incorrecta de dúplex.

ser despreciables, así que una elevada cuenta de errores de CRC se suele considerar como un síntoma de una asignación incorrecta del dúplex.

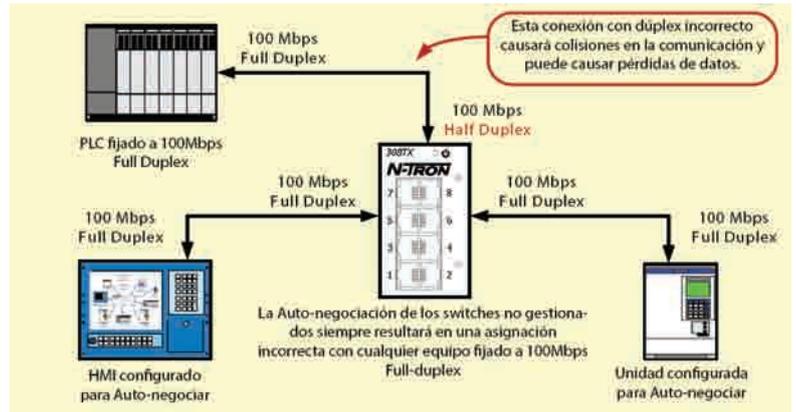
Las asignaciones incorrectas del dúplex pueden ser un problema particularmente difícil con switches no gestionables. Por definición, un switch no gestionable no posee la capacidad de fijar la configuración de un puerto a una velocidad y dúplex específicos y se encuentra siempre en modo de auto-negociación. Si un dispositivo se encuentra configurado con una velocidad o dúplex concretos, el switch no gestionable no podrá detectar un enlace completamente funcional y eventualmente surgirán problemas. Ver Figura 4 para un ejemplo ilustrado.

Auto-MDIX

La introducción del cableado con par trenzado también abrió la posibilidad de conectar el cable de múltiples formas. Un par trenzado categoría-5e contiene 8 hilos distintos de cobre pero hay dos formas estándares de cablear los hilos en los conectores RJ-45. La meta de estos estándares es proporcionar los mismos hilos en ambos extremos. Para cablear un cable cruzado los fabricantes de cables instalan un extremo boca abajo, presentando los pines 1 al 8 en un extremo y pines 8 al 1 en el otro.

Para que un dispositivo pueda conectarse con otro la transmisión (TX) en uno debe estar conectada a la recepción (RX) en el otro y viceversa. Este sistema es necesario para que dos dispositivos puedan comunicarse. Puesto que la mayoría de los cables suelen estar cableados para ser pin-a-pin, se decidió solucionar el problema a nivel del dispositivo. Tradicionalmente, los dispositivos de red y las tarjetas de red de los ordenadores están cableadas a la inversa uno de la otra. La interfaz dependiente del medio (Medium Dependant Interface - MDI) es la orientación en la que la tarjeta de un ordenador se cablea habitualmente y la interfaz dependiente del medio - cruzada (Medium Dependant Interface crossover - MID-X) es la orientación usada en un switch u otro dispositivo de red. Antiguamente esto era suficiente, pero requería el uso de cables especiales para las comunicaciones entre dos ordenadores o entre dos switches.

El Auto-MDI-X es un procedimiento desarrollado y patentado por dos ingenieros de HP e incluido en el es-



tándar Gigabit Ethernet del IEEE en el IEEE 802.3ab. El protocolo Auto-MDI-X elimina la necesidad de cables específicos, ya sean estos cruzados o pin-a-pin, conectando el receptor y el transmisor a ambos hilos del par. Gracias al estándar Gigabit Ethernet el receptor sabe qué es lo que está enviando el transmisor. Eléctricamente resta esa señal de lo que está recibiendo en el cable y usando cancelación de eco el receptor es capaz de calcular lo que se está transmitiendo en el extremo remoto.

Control de flujo

Con el incesante aumento de velocidad a la que los dispositivos transmiten datos, es importante que los backplanes, buffers y puertos switch-a-switch se mantengan al día con este escalado. Si la velocidad del backplane de un switch es mayor que la suma acumulada de las velocidades de todos los puertos del switch solemos denominarle switch con velocidad de red (Wire Speed Switch o Full Wire Speed). Esto suele ser imposible de alcanzar con switches con gran densidad. Un enlace estará saturado cuando la conexión entre dos dispositivos tiene más datos que transmitir que ancho de banda en el que transmitir esos datos. Este es un escenario fácil de producirse en un enlace entre dos switches si el puerto de subida (uplink) va a la misma velocidad que los puertos del usuario. Esto introduce la necesidad de controlar el flujo de datos, proceso que permite a un dispositivo solicitar al otro que detenga la transmisión para que pueda procesar los paquetes. Esta pausa podría incluir un temporizador de reinicio del envío de datos, requerir de una notificación de reinicio de envío, o simplemente ser una táctica de entretimiento con datos falsos para retrasar la comunicación.

En las redes de 10Mb/s los dispositivos que necesitan una pausa en la red simplemente llenan el medio con un paquete falso tras la recepción de cada paquete para ocupar el canal y evitar que le lleguen nuevos datos. A esta técnica se la denomina contrapresión (backpressure). La contrapresión también se utiliza como proceso de pausa en las redes de 100Mb/s en half-duplex. En las redes a 100Mb/s full-duplex y 1000Mb/s full-duplex, se implementa la auto-negociación del protocolo de control de pausa. Un dispositivo que necesite una pausa envía un FLP con el bit apropiado de pausa (bien el A5 o el A6).

Conclusión

En conclusión, el estándar de auto-negociación permite la existencia de un entorno plug-and-play en el mundo de las redes en las que existen múltiples velocidades, dúplex, cableados y controles de flujo. N-TRON recomienda dejar todos los dispositivos de una red fijados a auto-negociación para permitir un fácil despliegue y minimizar la posibilidad de que en el futuro haya la necesidad de introducir configuraciones especiales en la red. Si es indispensable fijar los parámetros, entonces N-TROM recomienda fijar ambos extremos y documentar estas configuraciones para asegurarse que no surgen problemas si en el futuro hay cambios en la red.

Mundialmente reconocida como un líder en el mercado de la Ethernet Industrial, los productos de N-TRON se usan en todo el mundo en una amplia variedad de aplicaciones incluyendo, marítima, control de procesos, parques eólicos, tratamiento de aguas, plantas nucleares, energía solar, y seguridad y vigilancia, donde la fiabilidad es un requisito imprescindible.