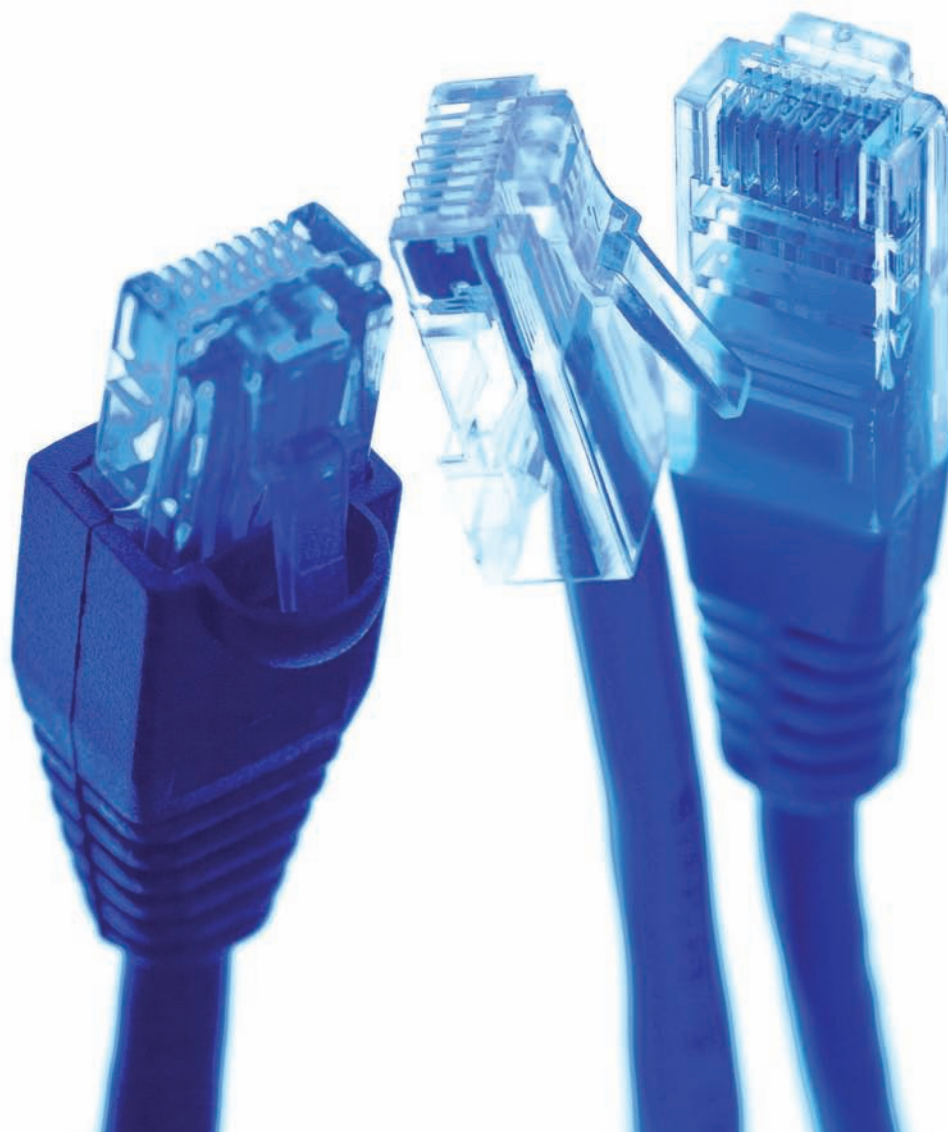


Ethernet como soporte de sistemas de control en red

Juan Ángel Gámiz Caro y Javier Gámiz Caro

Ethernet as a support of online control systems

Foto: Pictelia



RESUMEN

En el área de la automatización de procesos es frecuente encontrar lazos de control que se cierran a través de redes de comunicaciones. Estas arquitecturas de control dan lugar a lo que se conoce con el nombre de *sistemas distribuidos de control*. Cuando, además, los valores de las variables de control y de medida son actualizados, en el peor caso, dentro de un intervalo de tiempo conocido y asumible, se puede añadir a la frase anterior el calificativo de *tiempo real*.

Ethernet es una red de comunicaciones de enorme popularidad debida, entre otras causas, al esquema abierto de su interconexión, su eficiencia en el intercambio de grandes volúmenes de información y al bajo coste de las interfaces requeridas en su implementación. No obstante, su posible elección para soportar el tráfico requerido en aplicaciones de control de procesos no está exenta de riesgos.

En este artículo se exponen los requisitos exigidos a cualquier red que deba servir de soporte en sistemas distribuidos de control y, en particular, se evalúa la conducta temporal de Ethernet para este propósito.

Encargado: 1 de marzo de 2011

Recibido: 5 de junio de 2011

Aceptado: 25 de junio de 2011

Palabras clave

Redes de comunicaciones, buses de campo, Ethernet, telecomunicaciones, sistemas de control

ABSTRACT

In the area of process automation is common to find control loops are closed over communication networks. These control architectures give rise to what is known as distributed control systems. When, moreover, the values of the control variables and measurement are updated, in the worst case, within an interval of time known and manageable, you can add to the preceding sentence the adjective real time.

Ethernet communications network is a hugely popular due to, among other things, to open their interconnection scheme, its efficiency in the exchange of large volumes of information and low cost of the required interfaces in its implementation. However, a possible choice to support the required traffic control applications process is not without risks.

This article sets out the requirements for any system should provide support for distributed control systems and, in particular, we evaluate the temporal behavior of Ethernet for this purpose.

Commissioned: March 1, 2011

Received: June 5, 2011

Accepted: June 25, 2011

Keywords

Communication networks, fieldbus, Ethernet, telecommunications, control systems

Con toda probabilidad el siglo XX pasará a la historia como el *siglo de la información*. Aunque relativamente joven, la ciencia electrónica ha propiciado el imprescindible soporte tecnológico al conjunto de actividades relacionadas con las tecnologías de la información y las comunicaciones (TIC), en particular, y a la mejora del bienestar en general.

De manera similar a lo acontecido en el ámbito de la comunicación humana, la electrónica ha propiciado un desarrollo espectacular de las redes de comunicación industrial en las que dispositivos y equipos intercambian información relevante en las plantas de fabricación con objeto de controlar las variables del proceso tendientes a establecer y homogeneizar las características del producto.

Las actuales redes de comunicación industrial como, por ejemplo, la red Ethernet, pueden considerarse las herederas de aquellas otras más primitivas del pasado. De acuerdo con la tecnología empleada en su diseño, construcción y explotación, es posible clasificar las redes de comunicaciones en dos grupos claramente diferenciados: al primero pertenecerían las más antiguas, esto es, las redes de comunicaciones analógicas y, al

segundo, las redes de comunicaciones digitales.

Redes de comunicaciones analógicas

El ser humano se desenvuelve en un entorno donde las variables físicas que le afectan son fundamentalmente de naturaleza analógica, esto es, variables de magnitud continua definidas en todo tiempo (figura 1a). Este hecho le lleva inicialmente a proyectar, con la tecnología de la época, métodos básicamente analógicos para captar, transmitir y procesar las señales analógicas asociadas a las variables físicas de interés.

Así, en una primitiva red de comunicación analógica la amplitud de la señal eléctrica transmitida, típicamente tensión o intensidad, está muy relacionada con la magnitud de la variable física que tratar. Es una red de comunicación para el intercambio de medidas y acciones de control entre sensores, actuadores y equipos de control usados en sistemas automáticos de poca complejidad. Presenta las siguientes limitaciones:

a) Transmisión unidireccional: la información es transmitida solo en un sentido, esto es, desde el transmisor al receptor. No es posible realizar el inter-

cambio de información en sentido inverso y este hecho merma la capacidad interactiva emisor-receptor.

b) Susceptibilidad al ruido eléctrico: la amplitud de la señal transmitida, portadora de la información, se ve adversamente afectada por las perturbaciones electromagnéticas de su entorno.

c) Excesivo volumen de cableado: habitualmente los enlaces suelen requerir tantas líneas independientes para el intercambio de información como parejas de dispositivos emisores-receptores intervienen en la comunicación. Constituyen los enlaces denominados "punto a punto". Si la señal transmitida es en tensión, solo admiten la conexión de un emisor con un único receptor. En el caso de señales en intensidad (p. ej., de 0 a 20 mA) pueden soportar la conexión de un único emisor y varios receptores.

d) Funcionalidad reducida: presentan una escasa modularidad y complican la descentralización del control, el diagnóstico integrado y el mantenimiento (Lian et al, 2001).

Redes de comunicaciones digitales

Los avances de la tecnología electrónica, particularmente en el desarrollo y comer-

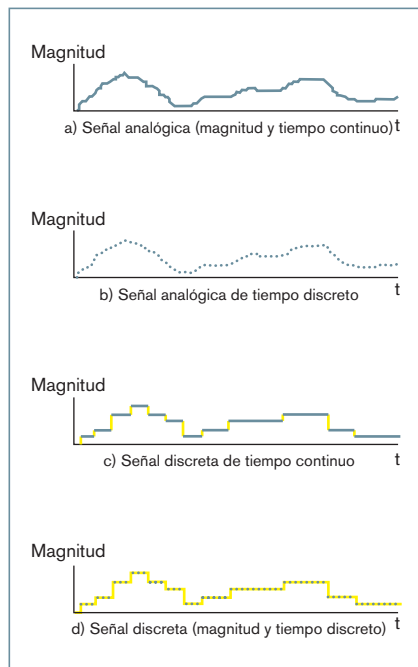


Figura 1. Tipos de señales obtenidas en un proceso de muestreo y cuantificación.

cialización del microprocesador (μP), posibilitaron procesar in situ las señales analógicas y obtener su correspondiente formato digital apto para ser transmitido por una línea de comunicación. Los procesos que permiten convertir una señal analógica (dominio continuo) en discreta (dominio discreto) son básicamente dos: *el muestreo y la cuantificación*.

El proceso de muestreo consiste en tomar periódicamente muestras de la magnitud de la señal analógica de interés, dando lugar a la aparición de una nueva señal denominada *analógica de tiempo discreto* (figura 1b). Cuanto mayor sea el número de muestras tomadas a la señal analógica primitiva, mayor fidelidad de esta poseerá la nueva señal. Nótese que entre muestra y muestra hay una evidente pérdida de información. La operación de muestreo la realiza normalmente un circuito electrónico de muestra y retención denominado *sample and hold*.

Mediante el proceso de cuantificación la amplitud de una señal analógica de tiempo discreto es ahora asociada a un conjunto finito de valores permitiendo obtener, bien una señal denominada *discreta de tiempo continuo* (figura 1c) o, directamente, una señal de magnitud y tiempo discretos (figura 1d). Cuanto mayor sea el conjunto finito de valores elegido en el proceso de cuantificación, más fielmente reproducirá la nueva señal la magnitud de la señal analógica de tiempo discreto. El conjunto finito de valores asociados a la magnitud analógica de partida es el que determina la *resolución* del proceso de cuantificación. La operación de cuantificación la suele realizar un componente electrónico denominado *convertidor analógico-digital* (ADC).

Finalmente, la señal obtenida en el proceso de cuantificación será periódicamente adquirida por un dispositivo electrónico como el microprocesador. A partir de aquí, y aplicando el apropiado proceso de codificación digital, el microprocesador podrá transmitir a través del canal de comunicación el código correspondiente a la magnitud de la variable tratada (figura 2).

Por codificación debe entenderse el proceso que interpreta la información discreta obtenida y la convierte en una combinación de ceros y unos, aptos para ser enviados a través de un canal de comunicación. En la práctica se suelen emplear distintos tipos de codificación en función del medio de transmisión y de las características de los equipos que comunican la información, todos ellos tendentes a conseguir menor distorsión, menor tasa de errores, mayor relación señal/ruido y mejor sincronismo entre emisor y receptor.

La comunicación digital solventa varias de las limitaciones que presentaba la comunicación analógica, aunque aparecen algunos otros inconvenientes que serán tratados más adelante. La comunicación digital presenta las siguientes características:

a) Transmisión y recepción bidireccional: la información puede ser transmitida en ambos sentidos, esto es, desde el transmisor al receptor y viceversa.

b) Poco susceptible al ruido eléctrico: la información de interés es normalmente transmitida mediante su correspondiente código numérico formado por unos y ceros. El código en su conjunto es el portador de la información y no, como sucedía en el caso de las comunicaciones analógicas, la amplitud de la señal transmitida.

c) Mínimo volumen de cableado: a una misma red de comunicación digital se pueden conectar múltiples dispositivos participantes (nodos), permitiendo el intercambio de información entre ellos en la forma que establezca el protocolo de arbitraje de la red.

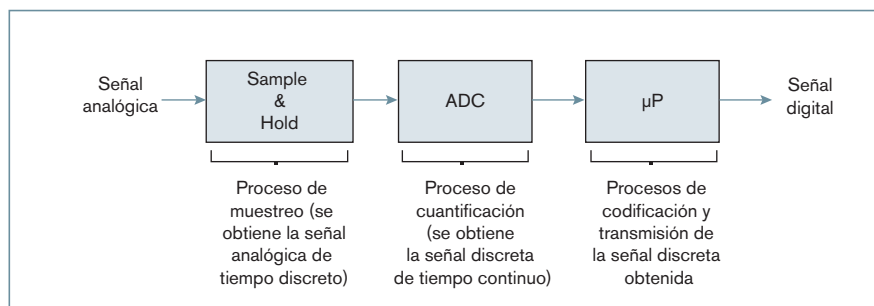
Implementando las redes de comunicación digital en sus procesos de fabricación, las industrias consiguen aplicar métodos de control más sofisticados en los que, inexorablemente, se produce un aumento significativo del número de sensores, actuadores y controladores de la planta pero que, a la vez, mejoran las ratios de producción y revierten en una mayor homogeneidad en la calidad de sus productos.

HART: el híbrido que posibilitó el cambio

En la década de 1980 aparecieron redes de comunicaciones industriales, de naturaleza exclusivamente digital, tendentes a facilitar el intercambio de información entre dispositivos, equipos y sistemas informáticos involucrados en los diferentes niveles del proceso productivo (figura 3). Para adaptarse a los requisitos de comunicación exigidos en cada uno de estos niveles aparecieron soluciones para la comunicación entre equipos y dispositivos en zonas geográficas cercanas, como las redes de área local (*local area network* – LAN), o en zonas más extensas, como las redes de área metropolitana (*metropolitan area network* – MAN) o las redes de área amplia (*wide area network* – WAN).

Las redes de comunicación que operaban en el nivel más próximo a la planta recibieron la denominación genérica de *buses de campo*. Este fue el caso, p. ej., del bus CAN (*controller area network*), desarrollado en 1986 por la firma alemana Robert Bosch GMBH para el sector de la automoción, y del bus PROFIBUS (perfil PA) desarrollado en el año 1987 por las firmas alemanas Bosch, Klöckner Möller y Siemens. En redes LAN surgieron redes como la popular Ethernet

Figura 2. Procesos de discretización de una señal analógica.



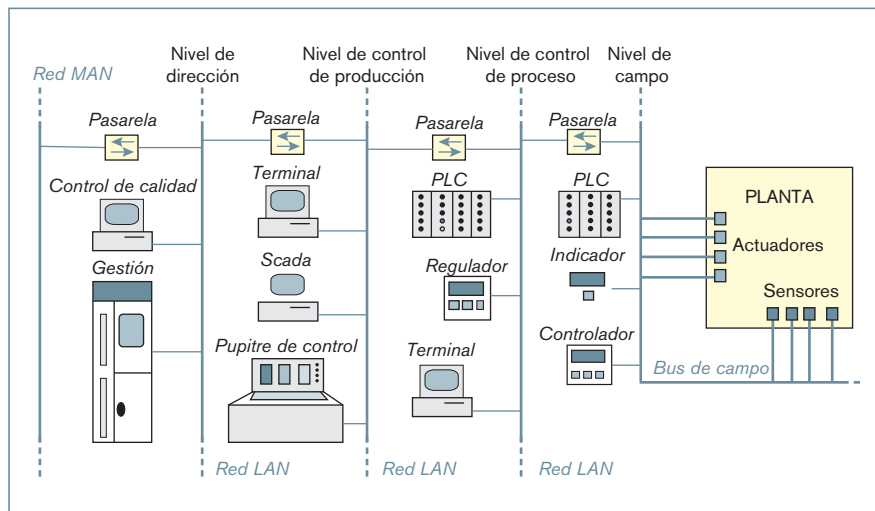


Figura 3. Arquitectura de redes de comunicaciones industriales.

en el año 1985 y PROFIBUS (perfiles DP y FMS) en el año 1987.

Convencidas de las ventajas que entrañaba el uso de las redes de comunicaciones digitales, las empresas se planteaban el cambio radical de toda la instrumentación de la planta con objeto de adaptarla a los nuevos requisitos. Sin embargo, el coste que suponía este cambio podía penalizar el rápido desarrollo de este tipo de redes. Este coste no solo venía dado por el valor de la nueva instrumentación ya que, también implicaba en muchos casos el paro forzoso de los procesos productivos durante el periodo de instalación y prueba de la nueva instrumentación.

En el año 1991 la firma Rosemount estableció una nueva filosofía de comunicación en el nivel de campo, dando lugar a la aparición de los dispositivos HART (*highway addressable remote transmitter*). La intención de Rosemount en aquellos momentos fue la de proporcionar al mercado un transmisor inteligente de campo (*smart transmitter*), compatible con los convencionales analógicos de 4 a 20 mA pero con la capacidad de intercambiar información con otros dispositivos de la planta en sistemas integrados de supervisión de control y adquisición de datos (SCADA). Rosemount trataba así de atender la demanda del mercado en relación con una conexión estándar que permitiera integrar de manera global la instrumentación propia de cualquier proceso industrial. Esta integración implicaba un comportamiento transparente de los dispositivos HART en redes analógicas convencionales, pero con la posibilidad de trabajar intercambiando información digital mediante unos mensajes de control y

operación universalmente aceptados y reconocidos por todos los dispositivos de la planta.

La comunicación digital de un dispositivo HART (Berrie et al, 1995) se establece modulando la frecuencia de una señal analógica (FSK), de amplitud constante de 1 mA ($\pm 0,5$ mA), y superponiéndola a la señal analógica de medida (figura 4). La distorsión que genera la señal FSK de comunicación sobre la analógica de medida es mínima dado que su valor medio es igual a cero y, por tanto, transparente para cualquier otro dispositivo clásico conectado a la línea.

Por efecto de la modulación, un bit a '1' superpone una frecuencia de 1.200 Hz y un bit a '0' una frecuencia de 2.200 Hz. En el equipo receptor se separa la señal de medida de la de comunicación

y esta última es desmodulada hasta obtener la información digital asociada.

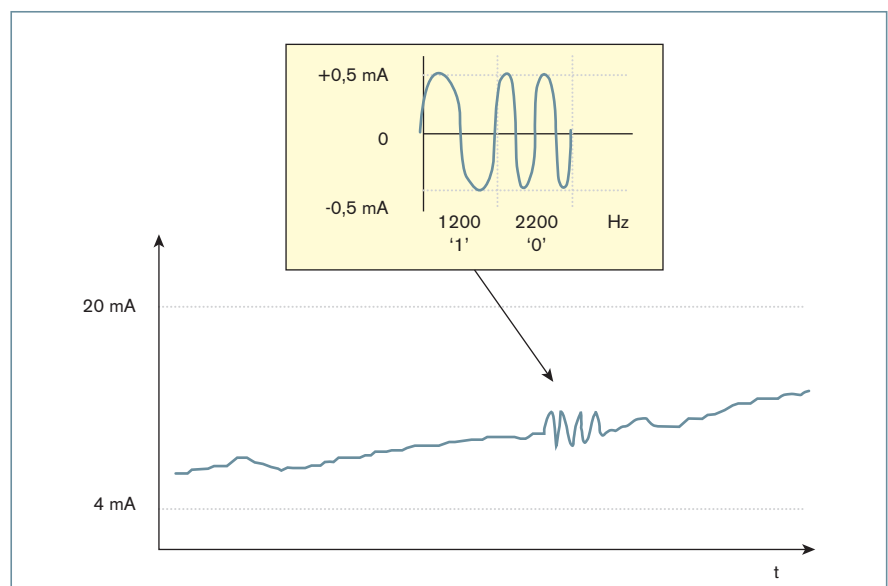
La inclusión de dispositivos HART en redes de comunicaciones analógicas hizo posible la reconversión paulatina de la instrumentación convencional de la planta y, a medio y largo plazo, posibilitó disponer del soporte necesario para establecer, finalmente, una comunicación solo digital entre todos los dispositivos instalados en la planta.

Sistemas distribuidos de control

La proliferación de las redes de comunicaciones digitales permitió pasar de los clásicos sistemas centralizados, típicos en la década de 1970 y normalmente basados en equipos de altas prestaciones y elevado coste, a los actuales sistemas distribuidos de control, también llamados sistemas de control en red (*network control system* - NCS), basados en equipos más sencillos conectados a través de estas redes (Berrie et al, 1995).

El criterio de elección de una red de comunicación digital que satisfaga los requerimientos de control de un cierto proceso industrial debe incluir las diferentes necesidades funcionales del diseño, su eficiencia, su coste, su mantenimiento y su futura ampliación (Eccles, 1998). Un error habitual en la elección de la red es considerar determinante la velocidad de transferencia de la información entre los distintos equipos y dispositivos de la instalación. Aun siendo este un factor importante, hay que tener muy presentes otros, tales como el retraso que sufren los mensajes causados por la posible congestión de la red, el determinismo temporal del

Figura 4. Características de las señales de comunicación en un dispositivo HART.



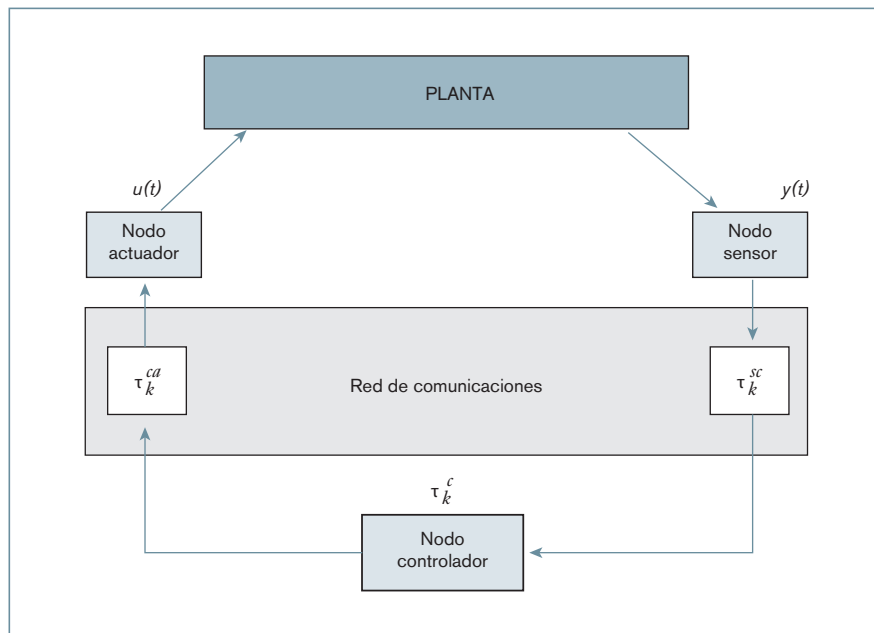


Figura 5. Retrasos de los mensajes en un sistema distribuido de control.

enlace, el número máximo de dispositivos participantes, la integridad en el intercambio de mensajes, etcétera.

El retraso de los mensajes en sistemas distribuidos de control

Frecuentemente, las variables del proceso que son captadas por los sensores de la planta son enviadas a través de la red de comunicación a un nodo controlador que, finalmente, determina y envía la acción de control a través de la red hacia un nodo actuador encargado de aplicar la correspondiente acción de control en el lugar conveniente de la planta. En aplicaciones industriales donde se trata de distribuir el control de la planta a través de una red de comunicaciones digital. Sin embargo, deben considerarse problemas adicionales derivados de su limitado ancho de banda efectivo, del método de acceso al medio empleado por los dispositivos participantes y de la sobrecarga de información que se puede presentar en la propia red.

Por ancho de banda efectivo de una red debe entenderse la cantidad máxima de datos significativos transmitidos por unidad de tiempo. Con independencia de la red utilizada, la transferencia de información a través de cualquier canal de comunicación conlleva un tiempo que dependerá, entre otros factores, del propio ancho de banda efectivo, de la sobrecarga del enlace, de la velocidad de transmisión, del número de participantes y del ruido eléctrico que pueda afectar a la red (Martí et al, 2001). Consecuentemente,

el tiempo empleado en la comunicación de la información puede verse afectado de un retraso que, en muchos casos, tiende a degenerar la estabilidad del control (Yepez, 2002).

En un lazo de control como el de la figura 5 se dan tres tipos de retrasos que deben tenerse en cuenta: el retraso de comunicación entre el nodo sensor y el nodo controlador (τ_k^{sc}), el retraso de cálculo operacional del controlador (τ_k^c) y el retraso de comunicación entre nodo controlador y nodo actuador (τ_k^{ca}).

Desde una perspectiva de control basada en el muestreo de datos es natural pensar en tomar una muestra k de la variable de interés, $y(t)$, y aplicar la correspondiente acción correctora al proceso, $u(t)$. Todo ello, con una cierta periodicidad temporal T . El retraso de muestra y actuación, τ_k , considerado el tiempo que transcurre desde que es tomada una muestra en el nodo sensor hasta que es actualizada la correspondiente señal de control en el nodo actuador, vendrá dado por la suma de los sucesivos retrasos (Sha, 2000):

$$\tau_k = \tau_k^{sc} + \tau_k^c + \tau_k^{ca}$$

Las redes candidatas a soportar el tráfico de información en sistemas distribuidos de control deben cumplir dos requisitos principales: que el retraso de tiempo que pueda sufrir cualquier mensaje sea limitado y que su transmisión esté garantizada. Esto implica que un mensaje debe ser transmitido con éxito en un plazo de tiempo limitado.

Componentes del retraso del mensaje y el determinismo de la red

El retraso máximo (Rm) que puede sufrir el mensaje en una red de comunicación es, a su vez, la suma de tres retrasos básicos (Tindell, 1995): el retraso Jm (jitter) que se produce desde que el mensaje es generado en el nodo hasta que es colocado en su cola de salida, el retraso Im que sufre el mensaje en la cola de salida antes de ser transmitido y, finalmente, el retraso Cm provocado por el tiempo de propagación del mensaje en la red (figura 6). Así, se tiene:

$$Rm = Jm + Im + Cm$$

En la mayoría de aplicaciones la componente Jm no interviene significativamente en el valor global del retraso por cuanto se trata de un tiempo dependiente de la propia electrónica del nodo y, consecuentemente, mínimo y despreciable frente a las componentes Im y Cm .

En aplicaciones de tiempo real en las que los mensajes se generan con una cierta periodicidad (aplicaciones *time-triggered*) es necesario considerar, además, dos parámetros de extraordinaria importancia a la hora de planificar la viabilidad del sistema de comunicaciones. El primero es el periodo, Tm , con el que se genera cada mensaje en particular y, el segundo, el retraso máximo admisible, Dm , para cada uno de los mensajes, contabilizado desde su generación en el nodo remitente hasta su recepción en el nodo destinatario.

Es evidente que para que la red de comunicaciones soporte los tiempos de intercambio de información previstos, para cualquier mensaje del sistema se debe cumplir que $Rm \leq Dm$ y que $Tm \geq Cm$. Esto implica (Gámiz, 2005) las definiciones siguientes:

$$\alpha = \min \left(\frac{Dm}{Rm} \right)$$

$$Cred = \sum_{k=1}^n \frac{Cm_k}{Tm_k}$$

Donde: α es el denominado factor de relajación y toma el valor del menor de los cocientes. Dm/Rm es la relación entre el retraso máximo admisible y el retraso real para un mensaje dado m .

$Cred$ es la carga de la red dada por la suma de cocientes Cm/Tm de los n mensajes del sistema.

n es el número total de mensajes del sistema.

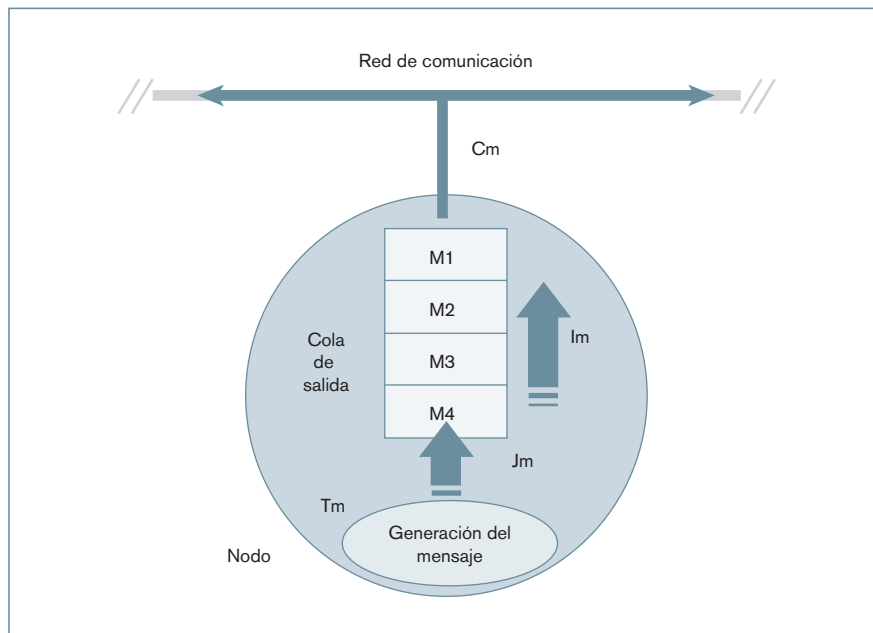


Figura 6. Componentes del retraso en la transmisión de un mensaje.

Nótese que para garantizar que el retraso de cualquier mensaje (R_m) no supera su valor máximo admisible (D_m), el factor de relajación de la red debe ser mayor o igual a la unidad ($\alpha \geq 1$). Asimismo, para asegurar que todos los mensajes generados en los nodos pueden ser transmitidos, la carga de la red debe ser menor o igual a la unidad ($C_{red} \leq 1$). Si estas dos condiciones se cumplen simultáneamente, podrá asegurarse que la red puede ser implementada cumpliendo con los requisitos temporales previstos.

Necesariamente, lo anterior implica que la conducta temporal de la red debe posibilitar la obtención de todos y cada uno de los retrasos de los mensajes (R_m) en el caso más desfavorable. Si esto es así, la red será catalogada de *determinista*.

Red de comunicación Ethernet

En 1985, el IEEE (The Institute of Electrical and Electronics Engineers) produjo un conjunto de normas para las LAN bajo la denominación IEEE 802.X. Una de las normas IEEE 802 es la conocida como Ethernet (IEEE 802.3), adoptada por ISO como ISO 8802-3. Ethernet es una red con topología lógica en bus cuyo estándar especifica una velocidad de transmisión de 10 Mbit/s, aunque con la versión Fast Ethernet se alcanzaron los 100 Mbit/s y con la actual Ethernet gigabit hasta los 10 Gbit/s.

Ethernet es una red de comunicaciones de enorme popularidad debida, entre otras causas, al esquema abierto de

su interconexión, su eficiencia en el intercambio de grandes volúmenes de información, al bajo coste de las interfaces requeridas en su implementación y a su velocidad. Su espectacular difusión le ha permitido presentar batalla para hacerse con cuotas de mercado, incluso, en el nivel de la pirámide de comunicaciones reservado a los buses de campo. En el soporte de numerosos sistemas distribuidos de control se presenta como un serio competidor a las redes deterministas, no obstante, cuando ha de garantizarse el retraso máximo que pueden sufrir los mensajes, principalmente en aplicaciones de tiempo real, su elección no está exenta de riesgos.

Capa de enlace

En el año 1984, bajo la norma ISO 7498, la ISO (International Standards Organization) aprueba el modelo de OSI (Open Systems Interconnection) que describe las reglas por las que deben regirse los equipos de comunicaciones que intercambian información a través de una infraestructura de red. El modelo especifica las reglas atendiendo a su obje-

tivo funcional y las clasifica en siete capas o niveles: *aplicación, presentación, sesión, transporte, red, enlace y físico*.

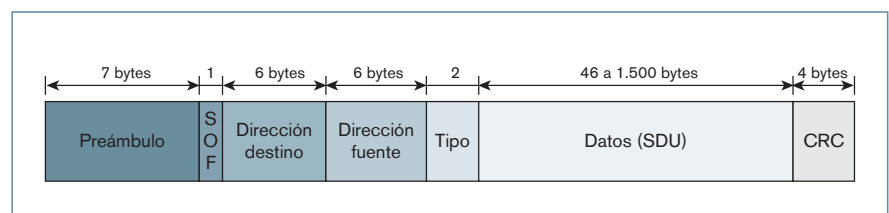
La capa de enlace de datos (*link layer*) de una red estándar Ethernet es la encargada, entre otras cosas, de empaquetar las tramas que conforman los mensajes (figura 7). Cada trama constituye una unidad de datos del protocolo (PDU) que comienza con 8 bytes (preámbulo + SOF), seguidos de 14 bytes para el control de la información, de 46 a 1.500 bytes significativos de datos útiles del mensaje (SDU) y, finalmente, 4 bytes para el control de errores CRC (*cyclic redundancy check*). La trama entera es precedida por un tiempo de inactividad en la red, IFG (*GAP inter-frame*), equivalente al tiempo de transmisión de 12 bytes (*bus idle*).

Método de acceso al medio

Ethernet utiliza en una subcapa de la capa de enlace un protocolo de control de acceso al medio (MAC) denominado CSMA/CD (*carrier sense multiple access with collision detection*). En CSMA/CD, cuando un nodo desea realizar una transmisión de datos observa que el bus esté libre (*bus idle*). En esta situación, si coincide que dos o más nodos intentan una transmisión en el mismo instante, se produce una colisión que corrompe los mensajes enviados, debiendo reintentar la transmisión después de transcurrido un tiempo aleatorio que viene dado por el algoritmo de arbitraje BEB (*binary exponential back-off*) que ejecuta cualquier nodo de la red.

El algoritmo BEB establece que después de producirse una colisión k entre dos o más nodos que intentan comunicar información al mismo tiempo, cada transmisor forma un conjunto de números: $\{0, 1, 2, 3, \dots, (2^k - 1)\}$ (para $k \leq 10$), elige un valor aleatorio R del citado conjunto e intenta reenviar la trama después de transcurrido un tiempo igual al producto de R por el tiempo de trama mínima (51,2 ms para Ethernet de 10 Mbit/s). De la undécima a la decimoquinta colisión k permanece constante

Figura 7. Trama de mensaje generada en Ethernet.



en 10 y a partir de la decimosexta colisión, ante la imposibilidad de transmitir el mensaje, el nodo lo desecha y notifica un error (Tanenbaum, 1996).

Cálculo del retraso del mensaje en Ethernet

El algoritmo de arbitraje BEB empleado en Ethernet estándar hace que esta red posea un comportamiento temporal *no determinista* (Wheelis, 1993), lo cual implica que no será posible calcular exactamente el retraso que pueden llegar a sufrir los mensajes (Lian et al, 2001), pero sí el valor probable de este retraso. Esto es así debido a que R es un valor obtenido al azar y de él depende, básicamente, el tiempo de retransmisión de los mensajes cuando se producen colisiones. Este hecho excluye a Ethernet del grupo de las redes aptas para el control de tiempo real.

Sin embargo, siempre que no se produzca el fenómeno de *captura de Ethernet* (Ramakrishnan, 1994), y para cargas de tráfico en la red medias (< 60%) y bajas (< 10%, habituales en redes Ethernet utilizadas para control), será posible determinar el retraso de los mensajes con una probabilidad elevada de acierto, lo que la capacita para gobernar otros procesos en los que el control no deba ser necesariamente de tiempo real.

Para trabajar con cargas de tráfico bajas o medias hay que limitar necesariamente el número de nodos conectados a la red e intercambiar mensajes con un mismo tamaño de datos útiles. Además, y con tal de apostar por la aleatoriedad *pura* que impone el algoritmo BEB, se debe cuidar que las características de la interfaz de comunicación en todos los nodos sean idénticas.

Considerando las premisas citadas, varios son los casos que se presentan cuando se desea obtener el probable retraso que sufren los mensajes en una red Ethernet:

Caso 1: transmisión continua de un único mensaje en cada nodo

Es el caso en el que se garantiza el empleo de todo el ancho de banda de la red para transmitir los mensajes al ritmo con el que éstos pueden ser “evacuados” de los nodos ($T_m = R_m$). Según la expresión (2), el tiempo de bloqueo Im que un mensaje m debe esperar una vez que el nodo de Ethernet está listo para transmitirlo está en estrecha relación con el grado de ocupación de la red en ese instante. Definiendo el parámetro $U_{rex}(m)$ como el factor de utilización de la red (ancho de banda) reservado para el resto de nodos en el instante en que un nodo trata de transmitir un mensaje m , se obtiene la correlación existente con el número de colisiones que puede llegar a sufrir el mensaje m antes de ser transmitido (figura 8).

Así pues, con valores de $U_{rex}(m)$ de, por ejemplo: 0,5 (2 nodos), 0,75 (4 nodos), 0,80 (5 nodos) y 0,90 (10 nodos), la probabilidad de que el mensaje m sufra una colisión antes de ser transmitido es de 1/2, 3/4, 4/5 y 9/10 respectivamente. Obsérvese que los numeradores de las fracciones anteriores (1, 3, 4 y 9) indican el probable número de colisiones (N_c) que sufre el mensaje m antes de ser transmitido.

Consecuentemente, el número de colisiones que sufren los mensajes en función del factor de utilización de la red, vendrá dado (Gámiz, 2005) por las expresiones:

$$U_{rex}(m) = \frac{\sum_{\forall i \in cex(m)} Cm_i}{\sum_{\forall k \in c(ms)} Cm_k}$$

$$N_c(m) = \frac{U_{rex}(m)}{1 - U_{rex}(m)} = M - 1$$

Donde: Cm es el tiempo de propagación del mensaje a una velocidad de red dada,

$cex(m)$ es el conjunto de mensajes del sistema excluido m ,

$c(ms)$ es el conjunto del total de mensajes del sistema,

$N_c(m)$ es el número de colisiones que sufre el mensaje m ,

M es el número de nodos de la red.

De otra parte, el tiempo de bloqueo del mensaje (Im) estará en estrecha relación con el tiempo que éste debe esperar hasta finalizar la transmisión del mensaje con el que colisiona. Consecuentemente, este retraso se podrá escribir como:

$$Im = \overline{Cm} \cdot N_c(m)$$

Donde: \overline{Cm} es el tiempo de transmisión medio de los mensajes del sistema (idéntico para tramas con un mismo número de datos útiles. En adelante se notará como Cm).

Para una línea de transmisión formada por conductores trenzados y apantallados la componente del retraso de propagación Cm en Ethernet viene dada por la expresión:

$$Cm = \frac{\text{bits fijos} + 8 \cdot \max[Ndata, 46]}{\text{baud_rate}} + l_{\text{línea}} \cdot 6 \cdot 10^{-9}$$

Donde: *bits fijos* son los 304 bits que totalizan la información fija de cada trama (38 bytes),

$\max[Ndata, 46]$ es el valor máximo de los dos indicados entre corchetes ($Ndata \leq 1500$ bytes),

baud_rate es la velocidad de transmisión en bit/s,

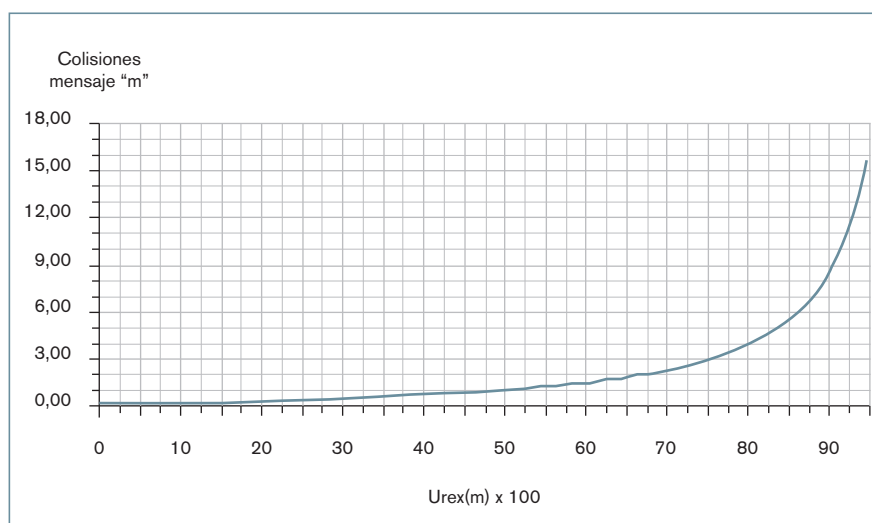
$l_{\text{línea}}$ es la longitud de la línea en metros.

Para Ethernet, y considerando la componente $Jm = 0$, la expresión (2) adquiere el siguiente aspecto:

$$R_m = Im + Cm = M \left(\frac{304 + 8 \cdot \max[Ndata, 46]}{\text{baud_rate}} \right) + l_{\text{línea}} \cdot 6 \cdot 10^{-9}$$

Nótese que, para el caso en estudio y los dos que siguen, el periodo de generación de los mensajes se considera igual para todos e idéntico a la del retraso obtenido ($T_m = R_m$). Esto fija el mínimo tiempo de generación de los mensajes en los nodos y marca el límite por debajo de la cual el

Figura 8. Número de colisiones del mensaje 'm', versus $U_{rex}(m)$.



sistema de comunicaciones no podría cumplir los requisitos temporales exigidos.

Caso 2: transmisión continua de un mismo número de mensajes en todos los nodos

Dado que en Ethernet estándar no existe privilegio alguno en el acceso al medio de los nodos que intercambian información, la determinación del retraso de cada mensaje es inmediata en el caso de que en los nodos haya un número idéntico de mensajes n a transmitir. En el caso más desfavorable, la expresión del retraso (9) es ahora:

$$Rm = n(I_m + C_m) = nM \left(\frac{304 + 8 \cdot \max[Ndata, 46]}{baud_rate} \right) + l_{línea} \cdot 6 \cdot 10^{-9}$$

Donde: n es el mismo número de mensajes en todos los nodos.

Caso 3: transmisión continua de distinto número de mensajes en los nodos

En el caso de que haya un número diferente de mensajes en cada nodo ($\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_M$), la expresión del retraso de los mensajes en cada nodo vendrá dada por:

$$Rm(\lambda_i) = \lambda_i(I_m + C_m) = \lambda_i M \left(\frac{304 + 8 \cdot \max[Ndata, 46]}{baud_rate} \right) + l_{línea} \cdot 6 \cdot 10^{-9}$$

Donde: λ_i es el número de mensajes en el nodo i .

$Rm(\lambda_i)$ es el retraso que sufren los mensajes del nodo i .

Caso 4: transmisión periódica de un único mensaje en cada nodo

Recoge el caso en el que los mensajes pueden ser generados en los nodos con una periodicidad mayor que la de su correspondiente tiempo de propagación por la red ($T_m > C_m$). Esto hace que ahora los parámetros U_{rex} y N_c no sean únicos bajo una determinada situación de carga de la red (idénticos en todos los nodos), sino dependientes de la periodicidad del mensaje en cada nodo. En este caso las expresiones (5) y (6) adquieren (Gámiz, 2005) el siguiente aspecto:

$$U_{rex}(\eta) = \frac{\sum_{\forall i \in cex(\eta)} \frac{C_{m_i}}{T_{m_i}}}{\sum_{\forall k \in c(ms)} \frac{C_{m_k}}{T_{m_k}}}$$

$$N_c(\eta) = \frac{U_{rex}(\eta)}{1 - U_{rex}(\eta)}$$

Donde: $U_{rex}(\eta)$ es el factor de utilización de la red en cada nodo ($\eta = 1, 2, 3, \dots, M$),

C_m es el tiempo de propagación del mensaje a una velocidad de red dada, $cex(\eta)$ es el conjunto de mensajes del sistema excluido el del nodo η ,

$c(ms)$ es el conjunto del total de mensajes del sistema,

$N_c(\eta)$ es el número de colisiones que sufre el mensaje del nodo η ,

M es el número de nodos de la red.

Aplicando un criterio análogo al expuesto en el caso 1, la expresión del retardo que sufre el mensaje en cada nodo vendrá dado por:

$$Rm(\eta) = (N_c(\eta) + 1)C_m = (N_c(\eta) + 1) \left(\frac{304 + 8 \cdot \max[Ndata, 46]}{baud_rate} \right) + l_{línea} \cdot 6 \cdot 10^{-9}$$

Resultados experimentales

Con objeto de validar la expresión (9) se realizó un conjunto de experiencias sobre una arquitectura como la mostrada en la figura 9. La plataforma de ensayo estaba formada por un conjunto de nodos (PC), construyendo, a partir de un HUB, una arquitectura en estrella con segmentos de cable de 5 m de longitud conectados con RJ-45. Los distintos nodos podrían ser considerados los elementos de un hipotético sistema de control (sensores, actuadores, controladores, etcétera) que intercambian la información de interés a través de la red de comunicaciones.

Para minimizar el tiempo de almacenamiento de los mensajes en cada uno de los nodos receptores se creó un disco virtual en RAM y, además, se instaló una aplicación que permitía examinar la velocidad de la red, la cantidad de información recibida y el tiempo que duraba el ensayo. Asimismo, en los nodos transmisores se instaló otra aplicación, confeccionada a medida, que posibilitaba realizar la transmisión de las tramas con el tamaño y la periodicidad deseados.

El procedimiento utilizado en las distintas experiencias se basó en el reenvío continuo de un mensaje, con 46 bytes

útiles de datos, desde el nodo transmisor 2 al nodo receptor 1. Mientras esta transmisión se producía, el flujo de datos era interferido por la información que intercambian otros nodos de la red (3, 4, ...n). De esta forma, aumentando el número de nodos que intercambian información a través de la red se podía obtener el retraso experimental que sufrían los mensajes en el nodo '1' bajo diferentes condiciones de carga en la red. La tabla 1 recoge los resultados de las experiencias realizadas.

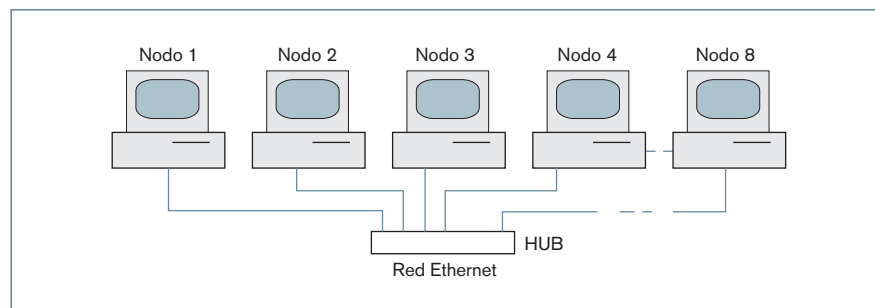
Las distintas experiencias realizadas sobre la plataforma de ensayo confirman plenamente las expresiones que describen el modelo de comportamiento temporal de la red Ethernet presentado anteriormente. La gráfica de la figura 10 valida, asimismo, las expresiones (5) y (6) al reproducir fielmente la relación existente entre el factor de la utilización de la red (U_{rex}) y el número de colisiones (N_c) en cada una de las experiencias realizadas.

¿Ethernet podría soportar tráfico de tiempo real?

Teniendo presente que Ethernet es una red no determinista debido, básicamente, al método probabilístico empleado por los nodos para acceder a la red después de producida una colisión, parece lógico imitar los mecanismos de arbitraje de otras redes deterministas que soportan tráfico de tiempo real. Un método de acceso al medio intrínsecamente determinista en redes con topología de bus es el de *paso de testigo* (*token bus* – IEEE 802.4) (ISO 1985). El símil de comunicación entre humanos se tendría en un método de intervención entre participantes basado en el turno de palabra. Este método evita colisiones dado que, en cada instante, solo *habla* aquel que le corresponde por turno.

El método de arbitraje basado en el paso de testigo evita el fenómeno de colisión de mensajes. Todo nodo de la red está identificado con un número concreto

Figura 9. Plataforma de comunicación en Ethernet.



| Baud-rate experimental (Mbit/s) | Factor de red (Urex) | Im teórico (μs) | Cm teórico (μs) | Número colisiones (Nc) | Rm teórico (μs) | Rm experimental (μs) | Error absoluto (μs) |
|---------------------------------|----------------------|-----------------|-----------------|------------------------|-----------------|----------------------|---------------------|
| 9,13 | 0 | 0,000 | 73,604 | 0,000 | 73,604 | 73,604 | 0,000 |
| 9,05 | 0,0798 | 6,439 | 74,254 | 0,087 | 80,693 | 80,702 | 0,009 |
| 9,19 | 0,1468 | 12,581 | 73,123 | 0,172 | 85,704 | 85,712 | 0,008 |
| 9,08 | 0,2664 | 26,876 | 74,009 | 0,363 | 100,884 | 100,899 | 0,015 |
| 9,16 | 0,3183 | 34,254 | 73,362 | 0,467 | 107,617 | 107,635 | 0,018 |
| 9,22 | 0,4298 | 54,939 | 72,885 | 0,754 | 127,824 | 127,847 | 0,024 |
| 8,93 | 0,5378 | 87,561 | 75,252 | 1,164 | 162,813 | 162,783 | -0,030 |
| 9,29 | 0,6385 | 127,763 | 72,336 | 1,766 | 200,099 | 200,099 | 0,000 |
| 9,19 | 0,6801 | 155,458 | 73,123 | 2,126 | 228,581 | 228,355 | -0,226 |
| 9,29 | 0,7329 | 198,483 | 72,336 | 2,744 | 270,819 | 270,860 | 0,041 |
| 8,83 | 0,8015 | 307,292 | 76,104 | 4,038 | 383,396 | 382,067 | -1,329 |
| 8,94 | 0,9078 | 740,101 | 75,168 | 9,846 | 815,269 | 815,620 | 0,351 |
| 8,86 | 0,9451 | 1.305,693 | 75,847 | 17,215 | 1.381,539 | 1.380,651 | -0,888 |

Tabla 1. Resultados de las diferentes experiencias realizadas.

que es conocido por el resto de participantes. Para poder realizar el intercambio de información el nodo debe esperar a recibir un mensaje único y característico (el testigo) de su antecesor y librarlo a su sucesor, una vez haya intercambiado la información deseada. De esta forma, el testigo pasa de participante a participante emulando virtualmente una topología de red en anillo, posibilitando así el intercambio de información. El protocolo debe garantizar, asimismo, que el testigo pueda circular ágilmente por todos los nodos de la red, evitando que cualquiera de estos pueda abusar del tiempo de posesión del testigo.

Para garantizar la ágil circulación del testigo pueden emplearse diversas técnicas. Por ejemplo, considérese que se establece, como un parámetro más de la red, un tiempo máximo para la rotación del

testigo (TMR) que es común y conocido por todos los nodos participantes. Entiéndase por tiempo de rotación del testigo el periodo que transcurre entre dos visitas del testigo a un mismo nodo de la red. Si, además, en cada nodo se contabiliza el tiempo real que transcurre entre dos recepciones consecutivas del testigo (TRR) y se obtiene la diferencia $TPT = TMR - TRR$, podría pensarse en vincular la transmisión de los mensajes pendientes de acuerdo con el resultado del valor TPT obtenido en cada uno de los nodos.

Cuando el valor TPT resultase mayor de cero (figura 11a), el nodo podría transmitir los mensajes pendientes hasta agotar el tiempo TMR. Si, por el contrario, TPT fuese cero o negativo (figura 11b), se podría pensar en transmitir un único mensaje (el más prioritario).

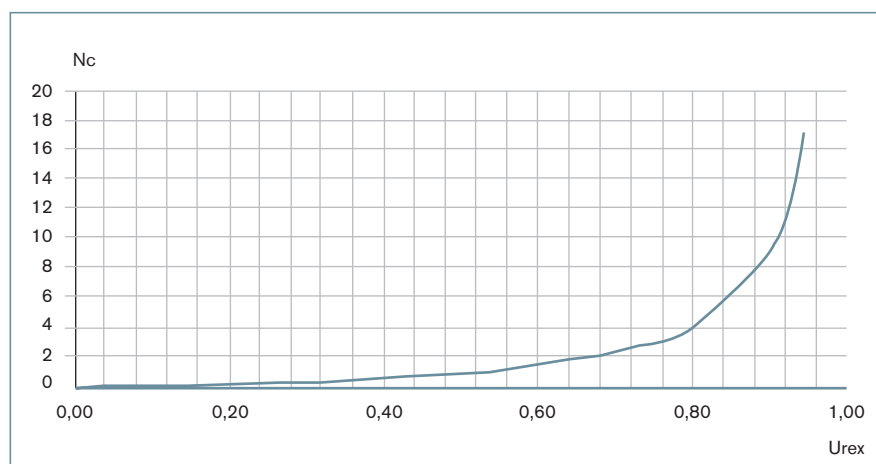
Lo anterior obligaría a asignar una prioridad concreta a cada mensaje de la aplicación que serviría para identificar y decidir el orden de transmisión de los mensajes en las diferentes situaciones que pudieran presentarse. La información que fija la prioridad de cada mensaje podría estar constituida por un conjunto concreto de bytes insertados como un 'identificador' en el campo de datos útiles del mensaje, lo que permitiría su ordenación automática en la cola de salida del nodo mientras el mensaje espera a ser transmitido.

Técnicas como la comentada tienden a regular el abuso en el tiempo de posesión del testigo y a seleccionar y transmitir el mensaje más prioritario cuando un nodo recibe el testigo a tiempo TMR vencido. Además, posibilitan la obtención del valor del retraso máximo que pueden sufrir los mensajes y, consecuentemente, aproximan el comportamiento temporal de la red al de cualquier otra de tipo determinista.

Variantes de las técnicas vistas son propuestas actualmente para conseguir dotar a Ethernet de la capacidad de gestionar el tráfico en aplicaciones de tiempo real. Este es el caso, por ejemplo, de los protocolos RETHER y otros en los que, sin modificar la compatibilidad de Ethernet, se puede intervenir en su subcapa de control de acceso al medio (MAC) para hacer que esta red pueda soportar tráfico de tiempo real.

Se ha de hacer notar, sin embargo, que soluciones como las indicadas se logran a expensas de un rendimiento inferior al que presenta el protocolo

Figura 10. Número de colisiones vs Urex.



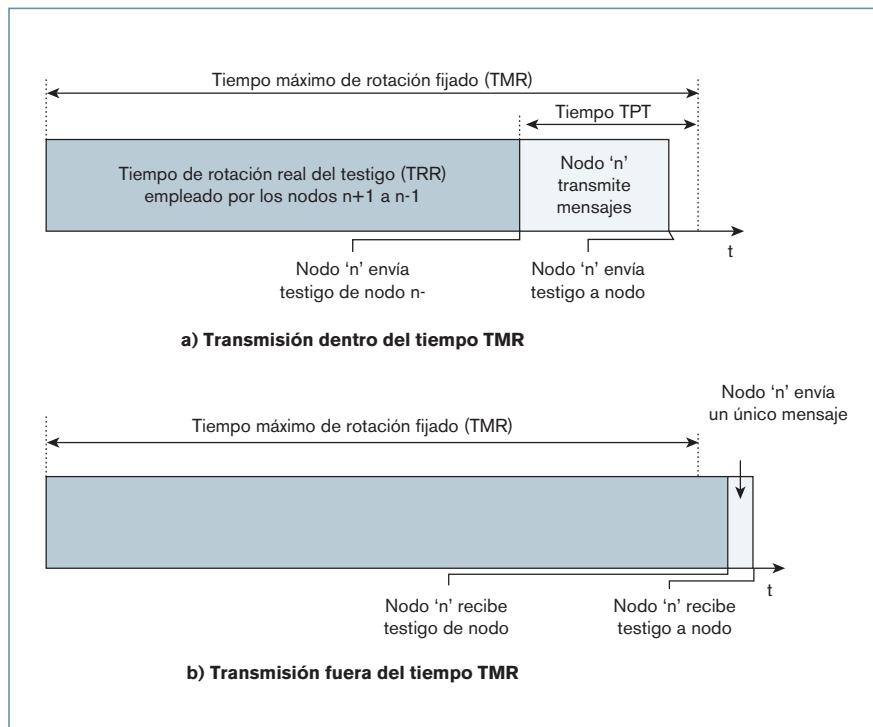


Figura 11. Limitación del tiempo de posesión del testigo (TPT) en el nodo 'n'.

CSMA/CD, sobre todo, cuando la red trabaja con volúmenes de tráfico reducidos (Koubias 1995).

Conclusión

Debido a su amplio ancho de banda, el estándar Ethernet presenta una buena eficiencia y un buen tiempo de respuesta en sistemas de control con un número de nodos reducido, con idénticas longitudes de trama para los mensajes y con intercambios de información dedicados exclusivamente a la aplicación de control. Sin embargo, el comportamiento temporal de la red estándar Ethernet no es determinista y esto puede decidir su exclusión en sistemas distribuidos de control de tiempo real en los que, por su relevancia o peligrosidad, los distintos mensajes del sistema estén sometidos a grandes constreñimientos de tiempo.

En sistemas de control que no requieran un tráfico de tiempo real, puede ser empleada la red estándar Ethernet bajo las premisas expuestas en el artículo. Las expresiones presentadas proporcionan el valor del retraso del mensaje más probable y, como refleja el resultado de las distintas experiencias, este valor se aproxima de manera notable al valor teórico calculado.

Realizando las oportunas modificaciones en el método de acceso al medio y valiéndose de la funcionalidad que proporcionan las interfaces de comunicación

de la red Ethernet estándar, sin embargo, pueden implementarse soluciones que posibilitan el empleo de esta red en aplicaciones de tiempo real.

Bibliografía

- Berrie P, Hennemann M, Schulze J, Schwörer T (1995). *Field Communication for Process Engineering*. p. 70-71, Alemania, Endress+Hauser.
- Eccles L (1998). *A smart sensor bus for data acquisition*. Sensors, vol. 15, no. 3. p. 28-36.
- Gámiz J, Grau A (2005). *Message delay in distributed control systems through Ethernet*. 16th International Foot and Ankle Congress (IFAC05), Praga.
- ISO, 1985). *Token Bus Access Method*. ISO DIS 8802.4.
- Koubias SA, Papadopoulos GD (1995). *Modern field-bus communication architectures for real-time industrial applications*. Comput. Ind., vol. 26, no. 3, p. 243-252.
- Lian F, Moyne J, Tilbury D (2001). *Performance evaluation of control networks: Ethernet, ControlNet, and DeviceNet*. IEEE Control System Magazine, Feb. (1994), pp. 66-83.
- Marti P, Fuertes J, Fohler G (2001). *An Integrated Approach to Real-Time Distributed Control Systems Over Fieldbuses*. 8th IEEE International Conference on Emerging Technologies and Factory Automation. France, October 15-18.
- Raji R *Smart networks for control*. IEEE Spectrum, vol. 31, pp. 49-55, June.
- Ramakrishnan K, Yang H (1994). *The Ethernet capture effect: Analysis and solution*. In Proc. 19th Conf. Local Comp. Networks, Minneapolis, MN, Oct., pp. 228-240.
- Sha L, Liu X, Caccamo M, Buttazzo G (2000). *Online Control Optimization Using Load Driven Scheduling*. 39th IEEE Conference on Decision and Control, Sydney, Australia, December 12-15.
- Tanenbaum A (1996). *Computer Networks*. 3rd ed. Upper Saddle River, Prentice-Hall.
- Tindell K, Burns A, and Wellings AJ *Calculating controller area network (CAN) message response times*. Contr. Eng. Practice, vol. 3, no. 8, pp. 1163-1169, Aug. (1995).

Wheeler J (1993). *Process control communications: Token Bus, CSMA/CD, or Token Ring?* ISA Trans., vol. 32, no. 2, pp. 193-198, July 1993.

Wittenmark B, Nilsson J, Törngren M (1995). *Timing Problems in Real-Time Control Systems*. In Proceedings of the 1995 American Control Conference, Seattle, Washington.

Yepez J, Marti P, Fuertes J (2002). *Control Loop Performance Analysis over Networked Control Systems*. 28th Annual Conference of the IEEE Industrial Electronics Society IECON02 Proceedings, pp.2880-2885, Sevilla 5-8, Noviembre 2002.

Juan Ángel Gámiz Caro

juan.gamiz@upc.edu

Doctor ingeniero en Electrónica y profesor titular de la Escuela Universitaria de Ingeniería Técnica Industrial de Barcelona (EUETIB), centro adscrito a la Universidad Politécnica de Cataluña (UPC). Profesor adscrito al departamento de Ingeniería de Sistemas, Automática e Informática Industrial (ESAI) de la UPC.

Javier Gámiz Caro

javier.gamiz@upc.edu

Ingeniero en Informática e ingeniero técnico en Informática de Sistemas. Profesor titular del departamento de ESAI de la UPC. Jefe de proyectos en el departamento de Sistemas de la empresa Aqua Ambiente (Grupo AGBAR).