TUTORIAL

GIGABIT ETHERNET: From 100 to 1,000 Mbps

Gigabit Ethernet is the latest in the family of high-speed Ethernet technologies. It extends the operating speed of the world's most deployed LAN to 1 billion bits per second while maintaining compatibility with the installed base of 10 Mbps and 100 Mbps Ethernet equipment.

HOWARD FRAZIER
Cisco Systems Inc.
HOWARD JOHNSON
Signal Consulting Inc.

The installed base for Ethernet ranges from PCs and servers to printers, test equipment, telephone switches, cable TV set-top boxes, process control systems, and even high-end theatrical lighting systems. In short, Ethernet is ubiquitous—which is to say, consumers can't get enough of it. The great value of Ethernet is not that one size fits all, but that common protocols and models are available at a wide range of price/performance points. Originally standardized and deployed at 10 megabits per second, Ethernet has always been considered a high-speed network. The newest standard, Gigabit Ethernet,[1] scales up two orders of magnitude to 1,000 Mbps, while maintaining compatibility with systems deployed in the early 1980s.

The Ethernet supplier community has consistently led the technical developments underlying ever-faster operating speeds on the installed base of network cabling—whether it was coaxial cable, Unshielded Twisted Pair, multimode fiber, or single-mode fiber. Each media type presents unique challenges. In this tutorial, we describe the operation of Gigabit Ethernet over fiber-optic cabling and the emerging IEEE standards project known as 1000Base-T, which addresses operation over Category-5 UTP cabling.

LAYERS OF CHANGE

Any tutorial concerning a network interface and access protocol standard must refer first to the ISO seven-layer reference model. Like most specifications produced by the IEEE 802 LAN/MAN standards committee, Gigabit Ethernet (IEEE802.3z) addresses the two lowest layers of the model:

- Layer 2, the Data-link layer, which describes how data are organized into frames and sent over the network, and
- Layer 1, the Physical layer, which describes the network medium and signaling specifications.

1089-7801/99/$10.00 ©1999 IEEE

Figure 1 represents the elements of the Gigabit Ethernet standard as sublayers in this context.

Media Access Layer

The Media Access Control sublayer occupies the lower half of the ISO Data-link layer. The MAC sublayer defines the Carrier Sense Multiple Access with Collision Detection protocol, which made Ethernet famous. The CSMA/CD protocol coordinates data transmission within Ethernet's shared-access physical bus topology. It enables each computer to sense the coaxial cable for electrical carrier signals before a frame is sent (CSMA) and to detect the simultaneous transmission of data (CD).

Shared-Access Topology Enhancements. The Gigabit Ethernet standard includes CSMA/CD with two important modifications:

- A carrier extension was added to overcome an inherent limitation of the CSMA/CD algorithm that mandated the round-trip signal propagation delay between two stations not to exceed the time required to transmit the smallest allowable frame. The carrier extension appends a set of special symbols to the end of short frames so that the resulting transmission is at least 4,096 bit-times in duration (4.096 microseconds), up from the minimum 512 bit-times imposed at 10 and 100 Mbps.
- An optional frame bursting feature was defined to improve the throughput of gigabit CSMA/CD LANs. Frame bursting allows multiple short packets to be transmitted consecutively, up to a limit, without relinquishing control of the signaling channel and reverting to the CSMA/CD protocol between packets. Thus, it offers a way to avoid the overhead associated with the carrier extension technique for all but the first packet of a burst.

Dedicated-Access Topology Enhancements. Widespread deployment of switched networks in the early 1990s has caused a shift from shared to dedicated LAN bandwidth, thus paving the way for the IEEE Standard 802.3x, Full-Duplex Ethernet,[2] which defines a protocol for full-duplex or simultaneous bidirectional communication over the physical signaling channel. When Ethernet operates in full-duplex mode, the CSMA/CD algorithm (including the carrier-extension and frame-bursting features) is disabled. This means that data transmission and reception can occur simultaneously without interference, and with no contention for a shared medium. Easier to implement than the

STANDARD DESIGNATIONS

1000Base-T = IEEE P802.3ab/D5.0 (draft)
Gigabit Ethernet = IEEE Std 802.3z-1998
Full-Duplex Ethernet with Flow Control = IEEE Std 802.3x-1997
Fast Ethernet = IEEE Std 802.3u-1995
Ethernet = IEEE Std 802.3-1998

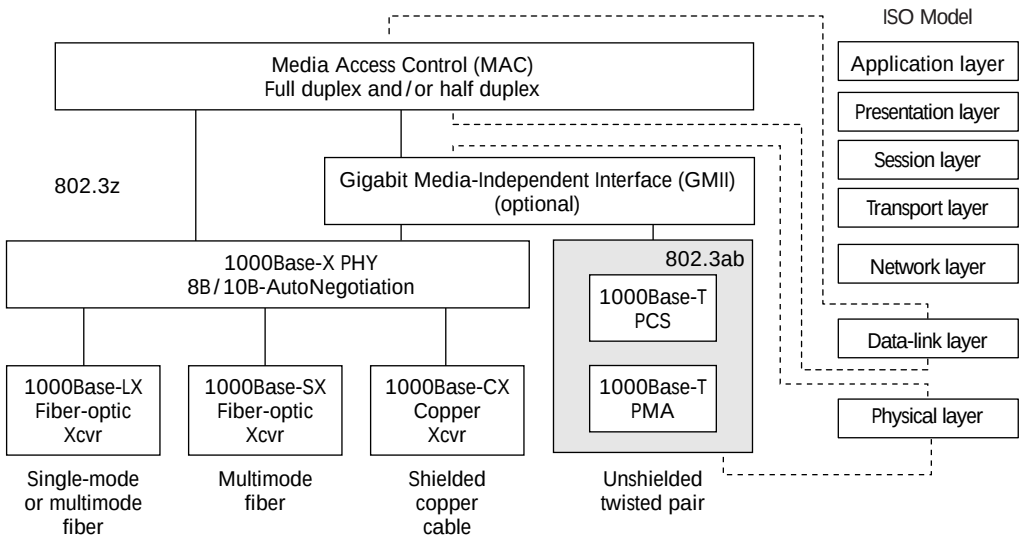

Figure 1. Gigabit Ethernet layer diagram addresses the MAC and PHY layers of the ISO model.

GIGABIT ETHERNET PRODUCT CATEGORIES

Gigabit Ethernet vendor products tend to fall into four categories.

Network Interface Cards are offered with a variety of interfaces to the physical medium, such as 1000Base-SX and 1000Base-LX. NICs also come with various levels of built-in protocol intelligence. Some can off-load tasks from their host computer's CPU.

Hubs can be categorized as either CSMA/CD or full-duplex repeaters. We are unaware of any commercial implementations of a gigabit CSMA/CD hub, but there are several full-duplex repeaters. No formal standard defines the functionality or performance of a full-duplex repeater; however, they generally rely on compliance of the attached stations with Gigabit Ethernet.

Switches may be classified in different ways, but most fall into one of three categories: aggregation devices versus backbone devices, stackables versus modular chassis, or layer-2 versus layer-3 switches. We define an aggregation switch as one that collects a number of lower speed links (10 or 100 Mbps) and combines them into a high-speed (gigabit) link. Aggregation switches are an excellent choice for the periphery of a LAN. They can be either a stackable "pizza box" with a relatively low port density (8 to 48 ports) or a small modular chassis with roughly twice this port density.

Backbone switches generally have a much higher port count and internal bandwidth than aggregation switches. They also provide greater flexibility, accommodating other network interface types, such as ATM or FDDI. They support rich network management capabilities and tend to take the form of modular chassis-based systems with plug-in "blades."

Routers and layer-3 switches can make more sophisticated forwarding decisions than multiport bridges or layer-2 switches. Routers are generally more flexible than layer-3 switches as they can implement a broader range of routing protocols and media interfaces, and can route a variety of network layer protocols. By contrast, a device can be branded a layer-3 switch if it simply performs IP routing between two Ethernets.

half-duplex (CSMA/CD) mode, the full-duplex mode also provides higher bandwidth. Thus, all currently shipping Gigabit Ethernet equipment operates in full duplex.

The Full-Duplex Ethernet standard further supports dedicated channels and full-duplex operation by introducing link-level *flow control*, which keeps switches from losing frames due to buffer overflow. A Pause protocol provides a mechanism whereby a congested receiver can ask the transmitter to inhibit (pause) its transmissions. The protocol is based on the transmission of a short packet known as a *pause frame*. The frame contains a timer value, expressed as a multiple of 512 bit-times, that specifies how long the transmitter should remain quiet. If the receiver becomes uncongested before the transmitter's pause timer expires, the receiver may elect to send another Pause frame to the transmitter with a timer value of zero, allowing the transmitter to resume immediately. Thus, the Pause protocol is like the XON/XOFF flow-control protocol, except that the pause frame really means "XOFF for a while." The protocol operates only between the two endpoints of a point-to-point link. It is not an end-to-end protocol and cannot be forwarded through bridges, switches, or routers.

Gigabit Ethernet builds on the Pause protocol by introducing asymmetric flow control. It uses the AutoNegotiation protocol (described later in the section "The Physical Layer"). This protocol lets a device indicate that it intends to send pause frames to its link partner, but declines to respond to them. If the link partner is willing to cooperate, then Pause frames will flow in only one direction on the link.

Media-Independent Interface

Because Ethernet supports a variety of physical media, it has always included a Media-Independent Interface below the MAC sublayer (see Figure 1). Gigabit Ethernet defines an enhanced MII, termed the Gigabit Media-Independent Interface. The GMII allows industry to develop additional Physical layer (PHY) specifications—for example, support for Category-5 UTP cabling.

Composed of independent 8-bit-parallel transmit-and-receive synchronous data interfaces, the GMII is intended as a chip-to-chip interface that lets system vendors mix MAC and PHY components from different chip manufacturers. Unlike the MII defined for 10- and 100-Mbps operation, the GMII is not exposed at a user-accessible connector, since such exposure would make it prohibitively difficult to maintain the required signal quality.

The management registers defined in the MII specification for IEEE Standard 802.3u, Fast Ethernet,[3] are carried over to the GMII, with appropriate extensions for managing gigabit PHY components. For example, the Pause function requires additional bits to support asymmetric operation and to identify a device capable of 1,000-Mbps operation.

The GMII provides a convenient demarcation between the MAC functions that are best implemented in a high-density circuit process, and the PHY functions that are best implemented in a high-speed circuit process. MAC functions are typically logic intensive, requiring thousands of individual logic elements (gates), wide data paths, and

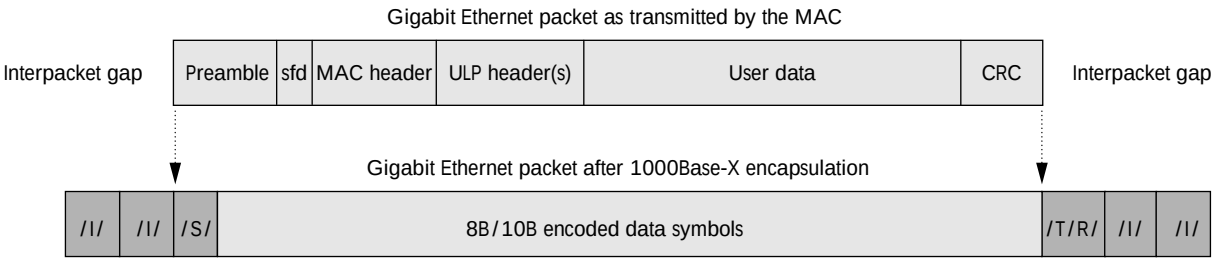

Figure 2. 1000Base-X encapsulation. The normal Ethernet preamble, start-frame-delimiter (sfd), MAC header, upper layer protocol (ULP) headers, user data (if any), and cyclic redundancy check (CRC) are encoded using the 8B/10B code rules.

embedded memory arrays. PHY functions are typically mixed-signal designs, involving both analog and digital functions; and they typically operate at higher clock frequencies. Over time, silicon processes tend to get faster and cheaper, consume less power, and cost less, leading to the eventual integration of MAC and PHY functions on a single die. This has been demonstrated in both 10- and 100-Mbps Ethernet, and is expected to be true for Gigabit Ethernet as well, starting as early as the year 2000. This integration is vitally important to the goal of making Gigabit Ethernet cheap enough to deploy on every desktop.

The Physical Layer

The PHY accepts the data stream from the MAC and converts it into optical or electrical impulses for transmission across a given medium. Within the Gigabit Ethernet standard, the PHY functions are subdivided into a group of logic functions known as the Physical Coding Sublayer and a group of analog-digital mixed signal functions referred to as the Physical Medium Attachment sublayer. Collectively, the PCS and PMA are referred to as the 1000Base-X PHY. Additionally, the 1000Base-X PHY performs a link configuration protocol referred to as AutoNegotiation.

Physical Coding Sublayer. The PCS encodes 8-bit data bytes from the GMII into 10-bit code groups, using a patented block-coding technique known as 8B/10B encoding.[4] The 8B/10B code is robust, well understood, and widely implemented. It is the basis for the ANSI Fibre Channel interface standard for high-performance mass storage devices,[5] and it has excellent properties such as transition density, run-length limiting, direct current (DC) balance, and error robustness. These properties were used to advantage in developing components for the Fibre Channel market, which in turn made components available early for the Gigabit Ethernet equipment market. Thus, system vendors were able to bring product to market in step with development of the Gigabit Ethernet standard (see the sidebar "Gigabit Ethernet Product Categories").

The PCS follows a few simple rules to encapsulate and encode data transmitted by the MAC. As shown in Figure 2, every packet begins with a special code group, referred to as the /S/ code group. This code will not appear in a normal sequence of encoded data, and so it reliably indicates the start of a packet. Following the /S/ code group, the normal Ethernet preamble, start-frame-delimiter, MAC header, upper layer protocol (ULP) headers, user data (if any) and cyclic redundancy check (CRC) are encoded using the 8B/10B code rules. Every packet is terminated with at least a pair of special code groups, referred to as the /T/ and /R/ code groups. Packets containing an odd number of code groups (counting from the /S/ to the first /R/) will be terminated with a second /R/ code group. Between packets, the interpacket gap is filled with a two-symbol idle code group represented as /I/. The idle code group is sent continuously between packets so that the clock recovery circuits associated with the receivers at each end of the link can maintain phase and frequency lock on the recovered clock and data.

With the inherent robustness of the 8B/10B code, the addition of rigorous frame encapsulation rules and a 32-bit CRC, Gigabit Ethernet provides extraordinary robustness against undetected errors, a key requirement to ensure data integrity. All single-, double-, and triple-bit errors in a frame are detected with 100 percent reliability.

Physical Medium Attachment. The 10-bit code-groups produced by the PCS are transmitted one bit at a time in a serial fashion by the PMA using a simple non-return-to-zero (NRZ) line coding. A

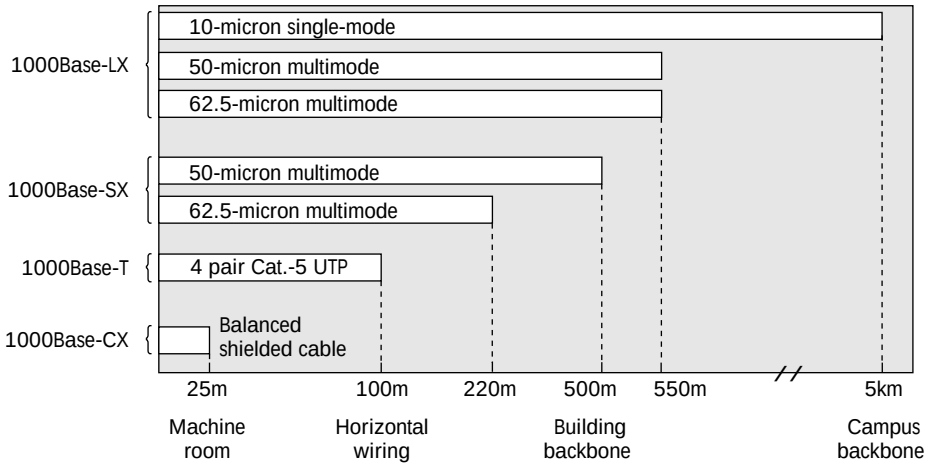

Figure 3. Gigabit Ethernet application environments and supported link distances for fiber-optic and twisted-pair media. (Distances are not drawn to scale.)

logic "one" in 10B code space is transmitted as high voltage (on copper media) or high intensity (on optical fiber); a logic "zero," as low voltage or low intensity. Gigabit Ethernet inherited the 8B/10B code's excellent transition density—that is, its high number of transitions from the logic one to logic zero state, which the phase locked-loop circuits require to perform clock recovery. Likewise, it inherits excellent DC balance—there is no accumulation of DC offset that might cause the DC baseline to wander in the receiver.

AutoNegotiation. Gigabit Ethernet includes an automatic mechanism to ensure that certain link characteristics are properly configured when links are initialized. AutoNegotiation was defined in Fast Ethernet to automatically select operational speeds between 10 and 100 Mbps. It was adapted to Gigabit Ethernet primarily to select between duplex mode and the use of link-level flow control.

On a 1000Base-X link, the configuration information exchanged through AutoNegotiation is encoded into a special sequence of 8B/10B codes (referred to as a /C/ code group) that transfers 16 bits of configuration information at a time. On a typical 1000Base-X link, the AutoNegotiation exchange will be completed roughly 40 milliseconds after the cables are plugged in or the equipment is turned on.

A RANGE OF MEDIA

Figure 3 charts the link distances for three transceiver types supported by Gigabit Ethernet: optical fiber, copper cabling, and UTP.

Fiber-Optic Transceiver Standards

The Gigabit Ethernet standard defines two styles of optical transceivers. These transceivers, 1000Base-LX and 1000Base-SX, support operation on three types of fiber-optic media, as shown in Figure 3. These are the three most popular types of fiber in common use, with 10-micron single-mode fiber dominating in the campus backbone, and 62.5-micron multimode fiber dominating in the building backbone and horizontal wiring environments.

Both transceiver styles defined by the Gigabit Ethernet standard are full duplex—that is, they use two fibers, one in each direction. Distinguished by their operating wavelengths, the LX transmitter emits at wavelengths of 1,270 to 1,355 nanometers, while the SX transmitter emits at wavelengths of 770 to 860 nm. Each receiver optimally receives light only within its respective wavelength band. As a result, the LX and SX devices do not interoperate.

While previous Ethernet standards accommodated only one type of transceiver for each distinct physical transmission medium, the Gigabit Ethernet working group made this exception so that the two transceiver types could fit different requirements in terms of installed base and operating range. LX is intended for longer distance, backbone links, while SX is intended for cheap, short-distance desktop links.

As Figure 3 shows, the LX transceiver works on either single- or multimode fiber, whereas the SX transceiver works only on multimode fiber. In terms of operating range, the LX transceiver, although more costly, goes significantly further

than the SX transceiver for all fiber types. Since it lacks the costly, high-precision optical mount necessary to couple into single-mode fiber, the SX transceiver is less expensive to build.

Shielded Twisted Pair Standard

The Gigabit Ethernet standard defines a transceiver for operation over 150-ohm balanced copper cabling. The 150-ohm balanced transceiver has an operating range of 25 meters. This transceiver is intended for use between two pieces of proximal equipment.

This transceiver is not intended to operate over EIA or ISO-standard in-building wiring. The cabling defined for use with this transceiver is a jumper cord, 25m long or less, physically composed of one differential shielded twisted pair running in each direction.

1000Base-T Supplement for UTP

Targeting the desktop connection market, the 1000Base-T transceiver standard is scheduled for adoption in early 1999. At that time, the possibility of communicating one billion bits per second over ordinary in-building wiring will become a reality.

The 1000Base-T standard supplements Gigabit Ethernet, providing a new transceiver type for use on four pairs of Category-5 UTP cable. The operating range will be 100 meters, in compliance with international building wiring standards EIA-568A and ISO-11801. This new transceiver will support GMII connections to the Gigabit Ethernet MAC layer.

The signaling scheme for 1000Base-T uses each of the four wire pairs, transmitting data on each wire in both directions simultaneously and continuously. Adaptive digital signal processing technology within each receiver cancels the attenuation, distortion, crosstalk from other pairs and self-induced crosstalk on each pair. The signals transmitted on each pair are 5-level pulse-amplitude modulation (PAM) at a signaling rate of 125 MBaud, the same signaling rate (per wire) used for existing 100Base-X transmission at 100 Mbps, but with more signaling levels.

Interoperation with existing 10- and 100-Mbps Ethernet standards requires 1000Base-T transceivers to implement AutoNegotiation. The AutoNegotiation system employs patterns of simple link test pulses on UTP as a means for systems to advertise and respond to various optional capabilities. Once the two ends agree on a common suite of protocols, link communication is established and the AutoNegotiation system goes into hibernation. The 1000Base-T transceivers will understand and communicate AutoNegotiation pulses, which will make it possible to construct devices with multiple personalities that may be switched on or off. In particular, it will be possible to construct a 10-, 100-, and 1,000-Mbps UTP combination port.

NETWORK TOPOLOGY AND MANAGEMENT

The topology rules of Gigabit Ethernet are fairly simple. A half-duplex Gigabit network with CSMA/CD hubs restricts operation to one hub per collision domain, with a maximum collision domain diameter (the distance between the two furthest end stations) of 200 meters. Full-duplex operation affords much greater flexibility because link distances are limited only by the underlying physical link technology.

In producing the update for Gigabit Ethernet, the IEEE took pains to preserve not only the "look and feel" but also the underlying functionality of Ethernet management. The new standard adds only a hub management parameter called "bursts" to help measure network efficiency. For full-duplex implementations, the 10- and 100-Mbps management information is carried over unchanged, although various statistics counters may now increment 10 times faster.

The Simple Network Management Protocol is the basis of most of the world's network management. The SNMP Management Information Base for Ethernet-like devices[6] and Ethernet physical layer devices[7] are defined by the Internet Engineering Task Force, working from the IEEE management specifications. Liaison activity between the IETF and the IEEE was initiated in early 1998 to assist the IETF in producing the MIB updates for Gigabit Ethernet. As of this writing, the IETF expects to complete the Gigabit Ethernet SNMP MIB update by the middle of 1999.

APPLICATION GUIDELINES

The corporate backbone was the initial target for Gigabit Ethernet. The link distances were chosen specifically to address the backbone application space, using the fiber most commonly installed in both building and campus backbones.

Metropolitan Area Networks represent another application area for Gigabit Ethernet. Organizations that own or lease dedicated fiber-optic cable runs may find that Gigabit Ethernet provides a low-cost way to move massive amounts of data between sites. While the maximum link span supported by Gigabit Ethernet is only 5 km on single-

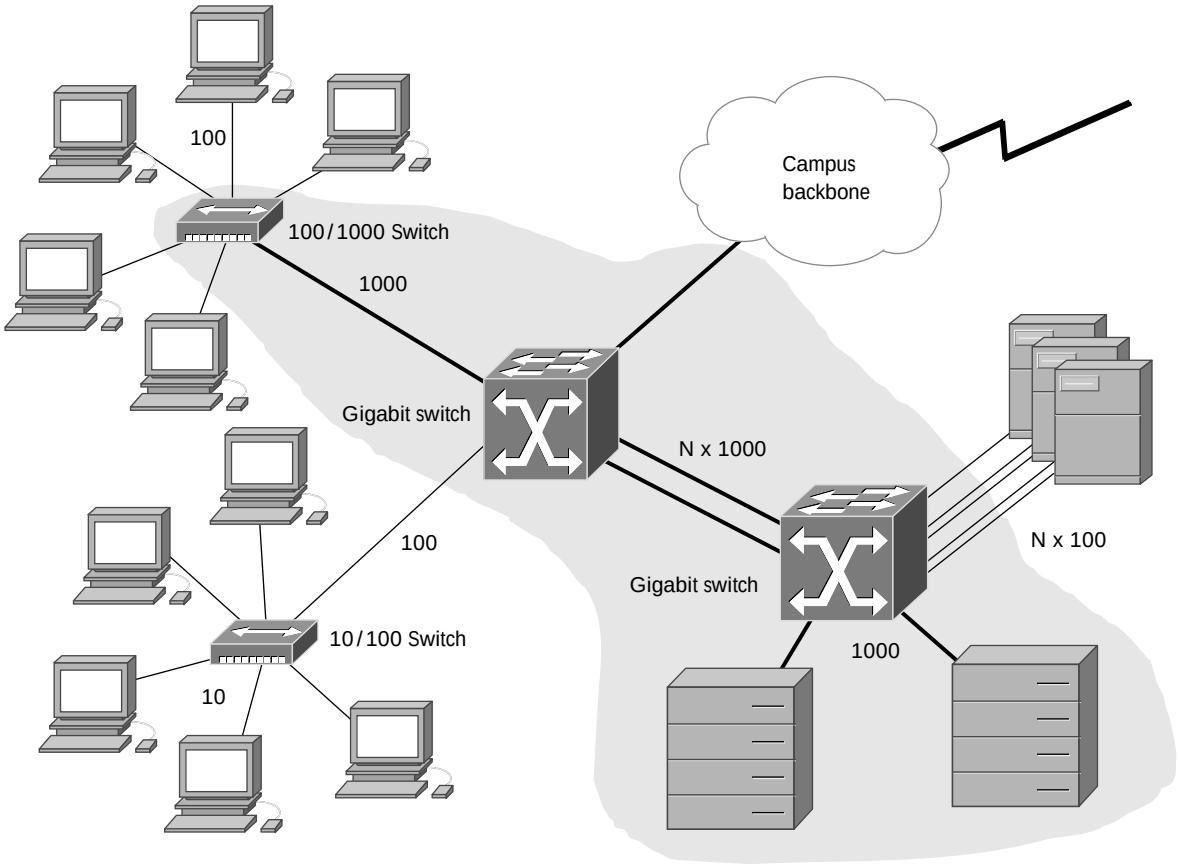

Figure 4. A typical high-performance client-server computing environment based on Ethernet. The shaded region represents the backbone of the network, which is usually the first candidate for an upgrade to Gigabit Ethernet.

mode fiber, longer link distances may be possible based on high-quality laser diodes operating in the 1,550-nm wavelength region.

Gigabit links operating at distances of 100 km between nodes can be achieved today without intermediate repeaters. In time, Gigabit Ethernet may catch on in MANs as an alternative to the services provided by the Public Switched Network.

Its real advantage over competing technologies is that upgrades can be done incrementally, as needed, with minimal network disruption. The keys to smooth upgrades are analyses of bandwidth and cabling requirements:

Bandwidth Requirements

For a network consisting mostly of desktop computers connected to a 10-Mbps Ethernet with a Fast Ethernet backbone, Fast Ethernet connections to servers, and a router to a WAN, it is very simple to analyze where more bandwidth is needed and then to upgrade those links, switches, and server NICs to Gigabit Ethernet. In Figure 4, the components highlighted in bold are candidates for Gigabit Ethernet upgrades. These upgrades are easy because the functionality before and after the upgrade is identical—the bits just move faster.

Identifying the upgrade candidates can be done with a network traffic analyzer or a management application that monitors the utilization of a link over time. For example, if the link between two switches is a Fast Ethernet and the inserted analyzer shows a high average utilization (anything greater than 50 Mbps) with peaks near 100 percent, then this link is ideal for upgrade to Gigabit Ethernet.

Similarly, if any of the servers shown at the right of Figure 4 are exhibiting high utilization on their Fast Ethernet interfaces, then their NICs should be upgraded to Gigabit Ethernet.

Cabling Requirements

Gigabit Ethernet is specified for operation on both multimode and single-mode fiber, with the maximum link spans shown earlier in Figure 2. To have the maximum confidence in a Gigabit Ethernet

installation, network designers should verify that the link spans in their network do not exceed these distances. This can be accomplished either by reviewing the information provided by the cable installer or by reverifying the installation using an Optical Time Domain Reflectometer and a loss test set. With 1000Base-T, it is also important to measure the return loss of the Category-5 UTP links and to verify that the interconnect hardware (outlets, patch panels) meets the Category-5 specification.[8]

CONCLUSION

Applications that operate over Ethernet or Fast Ethernet today will work equally well on Gigabit Ethernet, but the upgraded network will carry far more application traffic, allowing greater centralization and easier management of computational and storage resources. Because Ethernet is so prevalent, most applications and upper layer protocols have been designed to work well on it.

Web traffic is a great example of this. Even though the Internet spans the globe, with data traversing varying speed links and link protocols, the browsing experience from a desktop computer connected via Ethernet to a server connected via Fast Ethernet or Gigabit Ethernet is far more satisfying than browsing from a PC through a dial-up modem. Part of this is due to the bandwidth advantage that Ethernet has over a dial-up modem, and part is due to the fact that Ethernet is such a simple, low-overhead, robust protocol.

Of course, the backward compatibility, simplicity, robustness, and high performance of Gigabit Ethernet would be useless without multivendor interoperability. Early in the development process, a group of vendors convened under the auspices of the Interoperability Lab at the University of New Hampshire to form a Gigabit Ethernet Consortium (see the sidebar, "Gigabit Ethernet Online"). This consortium has produced a suite of conformance and interoperability tests that can be applied to products during development to ensure that they comply with the IEEE standard and, most importantly, interoperate with products from other vendors. Gigabit Ethernet will carry LANs into the next century, while maintaining compatibility with existing networks. It is simple, fast, and easy on the MIS budget. ■

REFERENCES

1. IEEE Std 802.3z-1998, "Media Access Control (MAC) Parameters, Physical Layer, Repeater and Management Parameters for 1000 Mbps Operation," IEEE Press, Piscataway, N.J., 1998.
2. IEEE Std 802.3x-1997, "Full Duplex Operation," IEEE Press, Piscataway, N.J., 1997.
3. IEEE Std 802.3u-1995, "Media Access Control (MAC) Parameters, Physical Layer, Medium Attachment Units, and Repeater for 100 Mbps Operation, Type 100BASE-T," IEEE Press, Piscataway, N.J., 1995.
4. A. Widmer and P. Franaszek, "A DC-Balanced, Partitioned-Block, 8B/10B Transmission Code," *IBM J. Research and Development*, Sept. 1983.
5. ANSI X3.230-1994, "Information Technology–Fibre Channel–Physical and Signaling Interface," ANSI, New York, 1994.
6. J. Flick and J. Johnson, "RFC 2358—Definitions of Managed Objects for the Ethernet-like Interface Types," IETF, 1998; available online at http://www.ietf.org/.
7. K. de Graaf et al., "RFC 2239—Definitions of Managed Objects for IEEE 802.3 Medium Access Units using SMIv2," IETF, 1997; available online at http://www.ietf.org/.
8. ANSI/TIA/EIA 568-A-1995, "Commercial Building Telecom. Cabling Standard," ANSI, New York, 1995.

GIGABIT ETHERNET ONLINE

The home page for the IEEE Gigabit task force is http://grouper.ieee.org/groups/802/3/z IEEE P802.3z. The specification can be downloaded for purchase from http://standards.ieee.org/catalog/.

The Gigabit Ethernet Alliance is a consortium of networking component and system suppliers and users. Its home page is at http://www.gigabit-ethernet.org.

The University of New Hampshire Interoperability Lab provides conformance and interoperability test services to manufacturers of Gigabit Ethernet products. Its home page is at http://www.iol.unh.edu/consortiums/.

Howard Frazier works in the Workgroup Business Unit of Cisco Systems. He was chair of the IEEE 802.3z Gigabit task force, which wrote the Gigabit Ethernet standard. He previously chaired the IEEE 802.3u 100Base-T task force, which developed the Fast Ethernet standard. Frazier graduated from Carnegie Mellon University with a BSEE in 1983.

Howard Johnson is president of Signal Consulting and chief technical editor of the IEEE 802.3z Gigabit Ethernet standard, as well as author of *High-Speed Digital Design: A Handbook of Black Magic* (Prentice Hall, 1993). Johnson received a PhD in electrical engineering from Rice University in 1983.

Readers may contact Frazier at hfrazier@cisco.com and Johnson at howiej@signalintegrity.com.