

# Network Delay And Signal Propagation

By Steve Steinke

05/07/2001 10:17 AM EST

URL: <http://www.networkmagazine.com/shared/article/showArticle.jhtml?articleId=17601083>

There is a widespread impression throughout much of the networking industry that optical networks are inherently faster and more responsive than those based on copper wire, and that wholly photonic transmissions are invariably faster and more responsive than transmissions consisting of signals based on electrons. In order to explore the senses in which this claim is accurate, we must look at some of the fundamentals of physics and communications theory.

The first fundamental fact is that electromagnetic waves propagate at a rate of approximately 300,000 kilometers (km) per second (s), or 186,000 miles per s in a vacuum. (This universal upper limit to the velocity that matter or information can achieve is usually abbreviated by "c.") The velocity of electromagnetic radiation in a vacuum is independent of frequency, so gamma waves, ultraviolet radiation, visible light, infrared (which is the range where practically all optical fiber traffic fits), and radio waves travel through space at the same rate.

The propagation velocity of waves can be dramatically affected by the medium they traverse, however. Recent experiments have actually demonstrated an ability to essentially stop and restart the propagation of light. In more ordinary media, such as certain commercial single-mode optical fiber products, the propagation velocity of a signal is 68 percent of c or 205,000km/s (see [table](#)).

In comparison, electric waves or signals in commonly used copper wire travel at speeds between 55 percent and 80 percent of c. Note that this is the propagation velocity of a signal or a wave, not the velocity of electrons in a wire. Electrons, in fact, move at speeds of about a fraction of a millimeter per s. Compare a wire to a garden hose filled with water; if you turn on the water at one end of the hose, water will flow out the other end almost immediately-in fact, as rapidly as a pressure wave propagates through water. Imagining that the speed of an electric wave in a cable would be constrained by the speed of individual electrons would be similar to counting the time it takes to fill an empty hose with water before it flows out the other end.

## Latency

The concept of latency is important for network and computer operations generally, but it isn't always consistently and tightly defined. In general, latency refers to the delay between the occurrence of two events. One important latency metric of a hard disk is the data access time: the time it takes to return a random block of data to the processor after a request is submitted, typically 8 milliseconds (ms) to 12ms in modern drives. (Drive makers sometimes restrict the term "latency" to a period called "rotational latency," the average time required to read data from a track once the head is positioned over it. Rotational latency typically contributes an average of 4ms or 5ms to a random disk access operation.)

Modems have two prominent latency producers: the time between the initial call request and the time the call is completed with the modems mating successfully (often as much as 30s), and the time between the point where the application sends data to the modem and the point where analog information begins to flow on the circuit (typically some 100ms).

Some of the most widely used latency measures for networks are end-to-end trip time, round-trip time, keystroke response time, and transaction completion time. End-to-end trip time is the time it takes a packet or other unit of data to travel from source to destination. Round-trip time adds the time for a return response or acknowledgment to the end-to-end latency. In the PSTN, round-trip time is almost exactly double the end-to-end time, but in connectionless networks the return path may be substantially different from the original transmission's. Keystroke response times measure from the time that a user presses an "enter" key or issues some other execute command until the time that an entire screen update is completed. This kind of latency is the best indicator of the responsiveness of an interactive application. Transaction completion time can be measured for automatic or "headless" applications, as well as for interactive ones.

For communications networks, propagation time is a significant part of end-to-end latency. The circuit-switched Time Division Multiplexing (TDM)-based PSTN is specifically designed so that once a call is established, there is very little added latency beyond propagation latency. Most telephones, as well as Class 4 switches, convert two-wire full-duplex circuits to four-wire simplex circuits using circuits called hybrids. Imperfect impedance matching in these hybrids creates echo signals, which reflect a speaker's voice back in a delayed form. People experience acoustic-psychological problems with delayed echoes, showing greater sensitivity as the delay increases. Delays of 10ms to 20ms are generally undetectable, but greater delays are more troublesome. U.S. phone companies have traditionally installed echo-cancellation circuitry every 500 miles (800km.) An 800km circuit running over optical fiber would introduce round-trip delays of about 15.6ms.

A commonly cited rule of thumb by voice-over-IP (VoIP) vendors is that round-trip delay times for high-quality voice should be less than 150ms. Round-trip transit time for North American transcontinental calls is about

42ms (San Francisco to Boston over optical fiber). Intercontinental calls may have three or four times as much propagation delay as continental landlines. Geosynchronous satellite communications, where the forward and return paths both involve up and down legs of at least the height of the satellite above the equator-36,000km-have minimum round-trip times of 480ms or so. (Satellite signals do have the advantage of propagating at the full speed of light.)

Connectionless networks, such as the Internet and other IP-based networks, use throughput capacity efficiently, rather than minimizing latency. If I run a traceroute utility from my house in the Bay Area to a server in Boston, I typically see round-trip times of 110ms, almost three times the propagation delay for the round-trip path. You'd better believe that the difference between 42ms and 110ms is accounted for by the 26 intermediate (one-way) nodes between the two hosts. With no congestion at the intermediate points, an IP network's round-trip time could approach the propagation delay time, but an IP network with no congestion or queuing anywhere wouldn't have the cost advantages of traditional TDM phone networks for data transmission, which led to the growth of IP in recent years.

## Throughput

Throughput and latency are completely independent issues. A channel with high throughput can move large quantities of data rapidly, but the first bit of data can never arrive faster than the latency permits. A file transfer operation, for example, may only have to suffer the "latency penalty" once at the beginning of the process, whereupon the full throughput capacity will be available to finish the job. Adding more throughput capacity will speed up the task. A financial transaction, on the other hand, might have numerous back-and-forth data flows, each of which must cope with the channel's latency. Throwing additional throughput at this problem may contribute little or nothing to overall performance. Improving the performance will require either reducing the latency of the individual flows or reengineering the application to reduce "ping-pong" operations.

For example, consider two channels where one has analog 56K modems at each end and the other has ISDN terminal adapters at each end. The best-case throughput of the analog modem channel will be perhaps 53Kbits/sec, while the throughput of an ISDN B channel is 64Kbits/sec. ISDN Terminal Adapters (TAs) have latencies of about 10ms, while analog modems have latencies of 100ms, as I mentioned earlier. So you can expect a 100Kbyte file transfer to take about 15s on the modem channel and 12.5s on the ISDN channel, assuming the calls are already established. The latency values are rounding errors in this scenario.

Now consider a series of 10 round-trip transaction flows of 1Kbyte, where each of the 20 flows can only occur after the previous one concludes. The analog link will take more than 5s for such a transaction, while the ISDN link will take 2.7s. No matter how many additional modem links are added to increase throughput, the transaction can never be completed in less than 2s,

thanks to the cumulative latency of the channel.

With this discussion of latency and throughput in mind, let's return to the initial discussion of optical fiber and copper media. The throughput capacity of optical fiber is, of course, staggering. The theoretical carrying capacity of a single-mode fiber using the 1,550nm "window" is 20Tbits/sec, or 20,000Gbits/sec (assuming 12.5THz bandwidth and a signal-to-noise value of 20dB). Twisted-pair copper wiring is hard-pressed to carry 1Gbit/sec of data farther than 100 meters. Coaxial cable can support somewhat higher throughput than twisted-pair, but the throughput advantages of optical fiber are so great that there has been little development of coax for data-carrying applications. Note that the cable TV operators were the first to deploy optical fiber to residential networks, and they chose to supplement their existing coaxial facilities with fiber.

Optical fiber has few, if any, advantages over copper wires where latency is concerned, however. Propagation latency is more or less the same with either medium. (Free-space optics and radio channels actually propagate 40 percent to 50 percent faster than optical fiber or copper wire.) Queuing, routing, and packetization latencies are also much the same-after all, these operations are performed electronically before they are converted to optical data flows. Any place that optical fiber is installed instead of copper wire, it will be because of its high throughput capacity or some other feature, not because it has superior latency characteristics.

*Steve Steinke, editor-in-chief, can be reached at [ssteinke@cmp.com](mailto:ssteinke@cmp.com).*

---

## **Calculations**

---

---

## **Resources**

---

There is a good set of slides covering this topic by Dr. Pisit Charnkeitkong of Ransit University at [this link](#).

Cisco Systems has a valuable Introduction to Voice and Telephone Technology at [Cisco's site](#).

The October 1999 issue of Communication Systems Design features an article entitled "Echo-cancellation for Voice over IP" by John C. Gammel, which provides a more detailed account of latency, its effects on echo creation, and efforts to minimize the problem of echoes in voice-over-IP (VoIP) systems. See [this link](#).