The Infrastructure of the Information Infrastructure

Brian Hayes

A reprint from

American Scientist

the magazine of Sigma Xi, the Scientific Research Society

Volume 85, Number 3
May–Junel, 1997
pages 214–218

The Infrastructure of the Information Infrastructure

Brian Hayes

Where is this place called cyberspace? I've been there, along with 50 million other Americans (according to the latest guess), and yet I would have a hard time pointing out on a map just where my virtual travels have taken me. We call the network the Information Highway, or the National Information Infrastructure, but it is not like other highways or other forms of infrastructure. Roads, bridges and power plants all make a conspicuous mark on the landscape; you can go out and have a look at them. But the Internet, even with its extraordinary rate of growth—doubling every 18 months or less—seems to be invisible.

One answer to the question "Where is cyberspace?" is that it has its own geography. The Internet is said to span national boundaries and make neighbors of people who might otherwise be isolated. Thus it is a kind of parallel universe, where distances are measured with a different metric. We may speak of URLs (universal resource locators) and e-mail addresses as if they specified a location in space, but the space is not the one we live in. On the Net, two sites with adjacent addresses need not be nearby in the physical world, and geographical neighbors can have totally different Internet addresses.

Nevertheless, the Net cannot float free of conventional geography. Not a single bit could pass through it without miles of copper wire and glass fiber, as well as tons of computing hardware—all of which is very much situated in the physical world. The cables and routing centers of the Internet have specific coordinates on the earth's surface, even if users of the network seldom give much thought to where their bits are going.

Some weeks ago it occurred to me that I know much more about the abstract protocols of the Internet than I do about the nuts and bolts that hold it together. When I dispatch a message, what route does it follow? How is that route determined? What is the physical form of the message at various points in its journey? I decided to see what I could learn. There were some surprises.

Brian Hayes is a former editor of American Scientist. *Address: 211 Dacian Avenue, Durham, NC 27701. Internet: bhayes@amsci.org.*

Getting to the Bottom of It

Suppose you are browsing in one of the preprint archives maintained on the World Wide Web. When you click on a link, you see a list of titles; another click and the abstract of a paper appears on your screen; still another click and the document itself is downloaded to your computer. It feels like you are conducting a dialogue directly with the remote machine, as if the two computers were connected by a dedicated wire. This impression is an illusion; there is no such wire. What's interesting is that the illusion is not yours alone; your browser software is also fooled, and so is the Web server at the other end.

Every message you send or receive on the Internet is broken down into packets, which are set loose on the network to find their own way. It is like communicating by carrier pigeon. You may want to send a whole book, but each bird can carry only a single page. The pigeons may take a variety of routes and arrive out of sequence. Some of them might get lost, so that their pages have to be sent again. But all this frenetic flapping of wings takes place out of sight. At the far end the pages are re-assembled in the correct order, and the sender and recipient know only that they have transmitted a complete book.

To create the illusion of a stable connection, networks are built in layers, with the upper layers concealing the lower ones from view. At the top of this "protocol stack" is the application layer, which is what you see and manipulate when you send e-mail or view a Web page. A program at this level has no need to master the minutiae of moving data from one machine to another; it only needs to know how to communicate with the next layer down in the stack.

That next layer (in the slightly truncated scheme I shall present here) is the transport layer; it is the home of TCP, the Transport-Control Protocol. TCP is the mechanism that creates the illusion of a dedicated and reliable link between two computers. It rearranges packets if they arrive out of sequence, and retransmits lost packets.

Next below the transport layer is the network layer, where we begin to think about where packets are headed and how to get them there. This is the task of the Internet Protocol, or IP. Packets are steered according to their destination's IP num-

ber, a 32-bit value usually written in "dotted-decimal" notation, as in 127.0.0.1. The steering process, called routing, is a subject I'll return to below.

Descending a step further we come to the data-link layer. Down to this point, all of the protocols have been independent of any specific network hardware; TCP and IP work the same whether you connect over a dial-up telephone line or an Ethernet cable. The data-link software, in contrast, has to talk to the hardware. It also handles such inconveniences as flow control (matching fast senders with slow receivers) and the correction of transmission errors.

Finally, at the bottom of the stack (and often in the basement of the building!) is the physical layer, where the network becomes tangible. Here the bits of a message are no longer regarded as mathematical abstractions; they are not 1s and 0s but voltage levels or modem tones or pulses of light.

The stratified architecture of the network is a means of walling off complexity. Keeping the layers independent means a programmer doesn't have to worry about the details of routing individual packets in a program that transfers whole files. Likewise the lower levels of the protocol stack don't know or care what's in the packets they send. The layering strategy does a remarkably good job of hiding low-level details from view—so good that most people who use the Net are oblivious to how it works.

Here I want to focus on the bottom three layers of the hierarchy: the network layer, where the principal challenge is routing, and the data-link and physical layers, where the task is to deliver the bits (and lots of them).

Finding Your Way through It

The idea of packet switching goes back to the prehistory of the Internet, when the Advanced Research Projects Agency was first planning the ARPANET in the 1960s. The alternative to packet switching is circuit switching, which is the basis of the global telephone network. When you pick up your telephone and dial a number, a continuous pathway is reserved for your exclusive use as long as the conversation lasts, whether or not you have anything to say. The ARPANET designers felt that circuit switching would be inefficient for computer communication, which tends to be "bursty," with long intervals of silence.

Packet switching avoids this waste of network resources, but it aggravates another problem. A circuit-switched network has to find a path for a connection only once, but a packet-switched network has to route every packet individually. It's as if you had to dial a separate telephone call for each word you spoke. You would thereby release the line for others to use during idle moments in your conversation, but you would make much greater demands on the equipment for routing calls to their destination

There are two aspects to routing on the Internet. The first task is simply to move all the packets through the system quickly enough that they don't back up and overflow somewhere. (For network operators, the standard description of a mishap seems to be "bits spilling all over the floor.") When a packet arrives at a node, the router has to store it in a buffer, examine the headers to see where the packet is going, look up the destination in a table to find out how to get there, and finally send the packet on its way—all before the next one comes along. The main strategy for meeting this challenge is simply throwing hardware at it. The routers at heavy-traffic "backbone" sites on the Internet are high-performance computers built specially for the task, with multiple high-speed input and output ports. They are $100,000-class machines.

The second part of the routing process is building and maintaining the table that tells the router where to send each packet. This is essentially a problem in graph theory: Each node needs to calculate the shortest route to every other node. For a small network, you could perform this analysis by hand, and then enter the tables into all the routers. This approach works, but network managers object that "It doesn't scale well." In the world of the Internet, there is no more damning phrase than "It doesn't scale well."

The alternative is to have the routers exchange information and maintain the tables on their own. An early algorithm had each router on the network broadcast its entire routing table every 30 seconds. This scheme is still employed within small networks, but it is another solution that

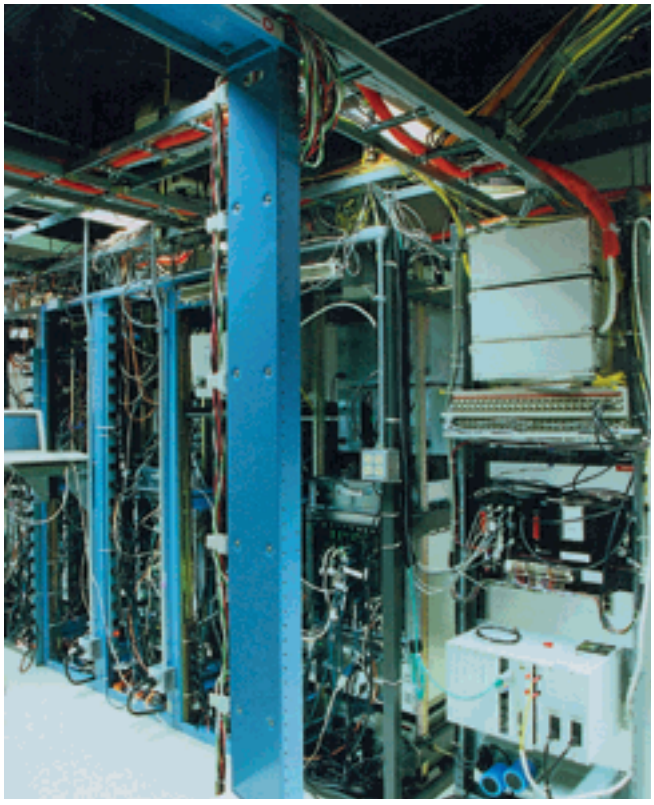

Figure 1. A node on the network at MCNC in North Carolina. (Photographs by the author.)

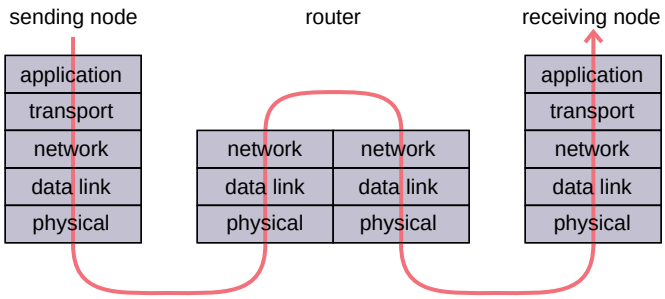

Figure 2. Router links nodes at the network layer of the protocol stack.

"doesn't scale well." If the thousands of routers on the global Internet all adopted this policy, there would hardly be room for any other traffic. The current algorithm for machines on the Internet backbone, called Border Gateway Protocol 4, transmits only essential updates to the tables.

Whose Is It?

A companion question to "Where is cyberspace?" is "Who owns the Internet?" The customary answer is a gleefully anarchic "Nobody!" or "Everybody!" And it's true there is no Internet czar or board of directors. Some of the organizations that set standards for the Net are open to all comers and operate on the principle of "rough consensus and working code." And the documents that define Internet protocols are known by the hyperpolite and self-effacing term "Request for Comments."

But these egalitarian—and occasionally countercultural—principles extend only so far. It turns out the various layers of the protocol stack correspond in a rough way to layers of ownership or custodianship. At the top, the application layer does indeed belong to everyone or no one. Each of us can choose Netscape or Lynx or Mosaic, or write our own application software, and we can run these programs on whatever computer we prefer. (This free choice of platform is no trivial liberation. Many other schemes of networking leave the user captive to proprietary hardware or software.)

The next few levels of the stack belong to the community of hacker volunteers. It is this group that defines protocols such as FTP (for file transfer) and HTTP (for the World Wide Web). Commercial interests have begun to encroach on these layers. For example, Netscape and Microsoft have both bypassed the "rough consensus" process in making extensions to HTTP. But the spirit of volunteerism still thrives and may yet win out.

The network layer and the data-link layer belong to the Internet service providers and the backbone operators. They control these layers because they own the routers and other hardware that runs the network at this level. If you have a private network that you want to connect to the Internet, this is the group whose acceptance you must win.

And finally the physical layer of the Internet belongs mainly to the telephone companies. They are the ones who own most of the cables and optical fibers. They always have. Even in the days of the government-supported ARPANET and NSFNET, the communications infrastructure was built and operated by the phone company. (Who else *could* have built it? Who you gonna call?)

High-capacity transcontinental and transoceanic cables are fabulously expensive, which means the telephone companies are by far the largest financial stakeholders in the Internet. It is therefore no surprise that when you go looking for the Net, where you are most likely to find it is in Ma Bell's basement. Not only do the interconnections run over telephone company cables; many of the backbone routers are housed in "co-lo space"—rooms in telephone switching offices where network operators can lease space to "co-locate" their equipment.

Furthermore, it's not just bare wire, dark fiber and co-lo rooms that the Internet has borrowed from the telephone system. The whole technology of long-distance digital communication is based on telephone standards and practices. We are pouring data through channels that were designed to carry human conversations; as one network guru confided to me, this is "a profoundly subversive act." What makes it all the more convoluted is that the telephone system itself encodes voices in digital data, so that we have data masquerading as voice masquerading as data.

Waiting at the Door

When I set out to see what the physical layer of the Internet looks like, I soon found myself talking to telephone companies and their subsidiaries. They were generous with their time and personnel. I spoke with a Chief Scientist and a Principal Engineer, as well as numerous network operators and technicians. I was shown laboratories and Network Operations Centers—dimly lit rooms with banks of glowing screens, where troubleshooters monitor traffic and repair outages by remote control. But the one place they wouldn't show me was the place I most wanted to see: the room where all the bits flow through.

I was particularly eager to explore one of the major "peering points," where networks come together to exchange traffic. Under ARPA there was no need for such exchanges because the network had a single backbone that tied together all sites, but today the Internet has evolved into an invertebrate, with multiple "transit" networks laced over the countryside. If a customer of UUnet sends e-mail to a customer of BBN Planet, the message must be handed off at a peering point. On the East Coast the two largest peering points are the New York NAP and MAE-East, which are operated respectively by Sprint and Worldcom, the third- and fourth-largest long-distance telephone companies in the U.S. (NAP stands for Network Access Point, MAE for Metropolitan Area Exchange.)

On the Web you can find lots of information about the NAPs and MAEs and other "public" peering points. There are diagrams of the local

networks that distribute traffic among the peers, graphs of throughput and lists of networks connected. But there are no photographs of the buildings or the machinery inside. Something else that's missing in many cases is a street address. For example, Sprint reveals that the New York NAP is in Pennsauken, New Jersey, but gives no finer-scale information on its whereabouts.

My initial requests for a tour of a peering point were refused. One reason given was confidentiality: The companies that keep equipment there expect access to be restricted. There were also security concerns. Telephone switching centers are built without windows, supposedly to protect equipment in a nuclear attack. Major Internet facilities might also merit a place on an enemy's target list, although the threat that seems to be of greatest concern today is not an ICBM but a terrorist with a truckload of fertilizer. (A newly appointed President's Commission on Critical Infrastructure Protection has just held its first meetings.)

These are perfectly good reasons for keeping me out. Nevertheless, after being turned away repeatedly, over a period of several weeks, I began to have an attack of inverse paranoia. This is the feeling that comes over you not when you think everybody is out to get you, but when everybody else thinks you're out to get *them*. Did they worry that I might be an industrial spy? A saboteur? *I* know I'm just a harmless journalist, but perhaps that's bad enough.

This story of frustrated curiosity has a happy ending. MCNC, an organization that operates a network for about a hundred research and educational institutions in North Carolina (including Sigma Xi and *American Scientist*) proved helpful and hospitable. I was given a hasty education in network operations, shown the facilities, and even allowed to photograph the equipment. Then, days before this issue of *American Scientist* went to press, I was finally also admitted to the inner sanctum of MAE-East, the largest of all the exchange points.

What about the New York NAP? I have not been inside, but on a scouting expedition to Pennsauken I found an unmarked and windowless concrete bunker, half buried in an earthen berm, with dual-redundant cooling units and diesel generators on the roof, and an abundance of "call before you dig" placards on the surrounding fences. The adjoining office has a Sprint emblem over the door. If this isn't the NAP, it's a masterful decoy.

The Internet Underground

Those who had barred the door to the machine room had also assured me there was nothing inside worth seeing anyway. They were right, of course. A network nexus is nothing like the bridge of the Starship Enterprise—or even the engineering deck. Maybe it's like the wiring closet of the Enterprise. Still, I was not disappointed. I happen to like wiring closets.

The equipment is mounted in floor-to-ceiling steel racks. The front panels face one way, but no one ever looks at them, as far as I can tell. All the action is on the back side, where the cables plug in. And it is the cables that attract the eye first. There are great multicolored rivers of them—orange and bright yellow fiber optics, creamy thin coaxial cables, gray-jacketed multiconductor bundles of copper—flowing through overhead trays and cascading down the sides of the racks.

Much of the machinery runs on batteries—banks of industrial-grade lead-acid cells, wired in series to produce 48 volts DC. The power is distributed through massive and handsome copper bus bars and cables as thick as a broomstick. Battery power is a tradition that the network engineers have inherited from the telephone companies, which have been running on batteries for well over a century and aren't about to switch.

In cyberspatial geography, MAE-East is a giant hub-and-spoke structure, where tentacles from nearly 100 networks all converge on a single point. No trace of this geometry is visible in the actual floorplan. The equipment racks are arranged in soldierly ranks and files, with just enough space to walk between the rows without too much worry over snagging a cable and spilling bits all over the floor. Half the room is given over to the equipment of the MAE itself,

Figure 3. A windowless concrete building in southern New Jersey may or may not be a Network Access Point.

the most important items being three big Gigaswitch units made by the Digital Equipment Corporation. These provide fiber-optic connections between the participating networks at up to 100 megabits per second. The other half of the room, beyond a wire gate, is co-lo space for customer equipment. The standard kit is centered on a Cisco Systems router in the 7500 series—the top of the line—which just about fills half a rack.

The décor of the machine room is unmarred by extraneous ornament. The room was created by walling off an area of the underground parking garage of a suburban Virginia office tower. The ceiling is low; harsh light pours out of fluorescent tubes; the air is filled with the white noise of a hundred computer cooling fans and a hint of battery fumes. Standing in this crowded space, surrounded by hard-working and very slightly grungy machinery, gives an interesting perspective and sense of scale, which is exactly what I was looking for in coming here. The room is no bigger than a two-car garage, and yet by some estimates more than half the traffic on the Internet passes through here.

The Internet-in-a-garage atmosphere of MAE-East will soon change. The MAE is expanding across the street into a space that has a different mood. The model for the new rooms is the look-but-don't-touch glasshouse where corporate mainframe computers were kept on exhibit for so many years. There is a raised floor to keep the cabling out of sight, so that even on this small scale the wiring of the world will be invisible and untraceable.

Of course décor is utterly inconsequential here—the machines truly don't care one way or another, and most of the time there are no people present. No one has a desk inside the room, and no one works full time there. Like so many modern industrial sites, it has been depopulated. The machines do not require constant attention. They are programmed and monitored by their owners in office cubicles across the street and across the continent.

Does It Scale Well?

Up to now the growth of the communications infrastructure has been driven mainly by the needs of voice telephony; data networking has just gone along for the ride, occupying a sliver of the available bandwidth. Michael O'Dell, Chief Scientist at UUnet, argues that this relation will inevitably change. The growth rate of voice telephony, he points out, is ultimately limited by the growth rate of the human population. Computer communication faces no such impediment. Computers reproduce faster than people, and they also talk faster. Eventually, then, the world will be wired mostly for data, with the occasional human voice piggy-backing on the bitstream.

Can the network-builders keep up with demand? Raw bandwidth is probably not the most serious constraint. Several fiber-optic channels are already carrying Internet traffic at 155 megabits per second, and rates of 622 megabits and 2.5 gigabits are on the horizon. (When the latter rate is used to carry voice, it funnels 32,000 conversations into one optical fiber.) The problem is how to get the bits into and out of the fiber. The present generation of routers is nearly saturated by the traffic on a single 50-megabit-per-second line. Handling 2.5 gigabits per second would take 50 of these routers running in parallel. Even if such an army of routers could be made to march in step, the prospect is logistically unattractive.

Routing has other drawbacks as well. Even with more efficient backbone algorithms, routers have to struggle to keep their tables up to date. The core of the Internet now includes some 45,000 routes—enough to challenge both the memory and the processor capacity of the largest routers. If a router somewhere on the backbone begins "flapping"—issuing a repetitive cycle of route changes—large hunks of the network can bog down and spill bits on the floor.

Bigger and better routers are one answer, but there is growing sentiment that routing just doesn't scale well, and we may need to return to some variation on the circuit-switched principles of the telephone system. Cisco Systems, the maker of most of the backbone routers in service today, has proposed a hybrid scheme called tag switching; another proposal is called IP switching. The basic idea in both plans is that when a node receives a long stream of packets all heading in the same direction, the node shouldn't have to repeat the same routing computation for each one; there should be some way of setting up a path just once and letting all the packets follow it.

Predicting the collapse of the Internet has been a popular sport since the day the Net was born. Bob Metcalfe, who speaks with the authority of a networking pioneer (he is the inventor of Ethernet), made a firm prediction that it would all come crashing down no later than the end of 1996. The failure of his forecast is no guarantee that doomsday won't come tomorrow. Still, for whatever it's worth—and I speak as no more than an Internet tourist—when I look around me at the National Information Infrastructure today, so modest in scope as to be almost invisible, I see a lot of room for growth.

Bibliography

Hafner, Katie, and Matthew Lyon. 1996. *Where Wizards Stay Up Late.* New York: Simon and Schuster.

Dickie, Mark. 1994. *Routing in Today's Internetworks: The Routing Protocols of IP, DECnet, NetWare, and Appletalk.* New York: Van Nostrand Reinhold.

Federal Communications Commission. 1996. "Fiber Deployment Update: End of Year 1995." http://www.fcc.gov/ccb.html

Labovitz, Craig. 1996. "Routing Analysis: Internet Performance Measurement and Analysis (IPMA) Project." http://compute.merit.edu/analysis/routing.html

Manning, Bill. 1997. North American Exchange Points. http://www.isi.edu:80/div7/ra/naps_na.html

Metcalfe, Bob. 1995. From the ether. *InfoWorld*. http://www.infoworld.com/cgi-bin/displayNew.pl?/metcalfe/bm120495.htm (December 4, 1995.)