APPENDIX

The Internet as System and Spirit

This Appendix explains how the Internet works and summarizes some larger lessons of its remarkable success.

The Internet as a Communication System

The Internet is not email and web pages and digital photographs, any more than the postal service is magazines and packages and letters from your Aunt Mary. And the Internet is not a bunch of wires and cables, any more than the postal service is a bunch of trucks and airplanes. The Internet is a system, a delivery service for bits, whatever the bits represent and however they get from one place to another. It's important to know how it works, in order to understand why it works so well and why it can be used for so many different purposes.

Packet Switching

Suppose you send an email to Sam, and it goes through a computer in Kalamazoo—an Internet *router*, as the machines connecting the Internet together are known. Your computer and Sam's know it's an email, but the router in Kalamazoo just knows that it's handling bits.

Your message almost certainly goes through some copper wires, but probably also travels as light pulses through fiber optic cables, which carry lots of bits at very high speeds. It may also go through the air by radio—for example, if it is destined for your cell phone. The physical infrastructure for the Internet is owned by many different parties—including telecommunications firms in the

U.S. and governments in some countries. The Internet works not because anyone is in charge of the whole thing, but because these parties agree on what to expect as messages are passed from one to another. As the name suggests, the Internet is really a set of standards for **inter**connecting **net**works. The individual networks can behave as they wish, as long as they follow established conventions when they send bits out or bring bits in.

In the 1970s, the designers of the Internet faced momentous choices. One critical decision had to do with message sizes. The postal service imposes size and weight limits on what it will handle. You can't send your Aunt Mary a two-ton package by taking it to the Post Office. Would there also be a limit on the size of the messages that could be sent through the Internet? The designers anticipated that very large messages might be important some day, and found a way to avoid any size limits.

A second critical decision was about the very nature of the network. The obvious idea, which was rejected, was to create a "circuit-switched" network. Early telephone systems were completely circuit-switched. Each customer was connected by a pair of wires to a central switch. To complete a call from you to your Aunt Mary, the switch would be set to connect the wires from you to the wires from Aunt Mary, establishing a complete electrical loop between you and Mary for as long as the switch was set that way. The size of the switch limited the number of calls such a system could handle. Handling more simultaneous calls required building bigger switches. A circuit-switched network provides reliable, uninterruptible connections—at a high cost per connection. Most of the switching hardware is doing very little most of the time.

So the early Internet engineers needed to allow messages of unlimited size. They also needed to ensure that the capacity of the network would be limited only by the amount of data traffic, rather than by the number of interconnected computers. To meet both objectives, they designed a *packet-switched network*. The unit of information traveling over the Internet is a packet of about 1500 bytes or less—roughly the amount of text you might be able to put on a postcard. Any communications longer than that are broken up into multiple packets, with serial numbers so that the packets can be reassembled upon arrival to put the original message back together.

The packets that constitute a message need not travel through the Internet following the same route, nor arrive in the same order in which they were sent. It is very much as though the postal service would deliver only postcards with a maximum of 1500 characters as a message. You could send *War and Peace*, using thousands of postcards. You could even send a complete description of a photograph on postcards, by splitting the image into thousands of rows and columns and listing on each postcard a row number, a column number, and the color of the little square at that position. The recipient could, in principle, reconstruct the picture after receiving all the postcards. What makes the Internet work in practice is the incredible speed at which the data packets are transmitted, and the processing power of the sending and receiving computers, which can disassemble and reassemble the messages so quickly and flawlessly that users don't even notice.

Core and Edge

We can think of the ordinary postal system as having a *core* and an *edge*—the edge is what we see directly, the mailboxes and letter carriers, and the core is everything behind the edge that makes the system work. The Internet also has a core and an edge. The edge is made up of the machines that interface directly with the end users—for example, your computer and mine. The core of the Internet is all the connectivity that makes the Internet a network. It includes the computers owned by the telecommunications companies that pass the messages along.

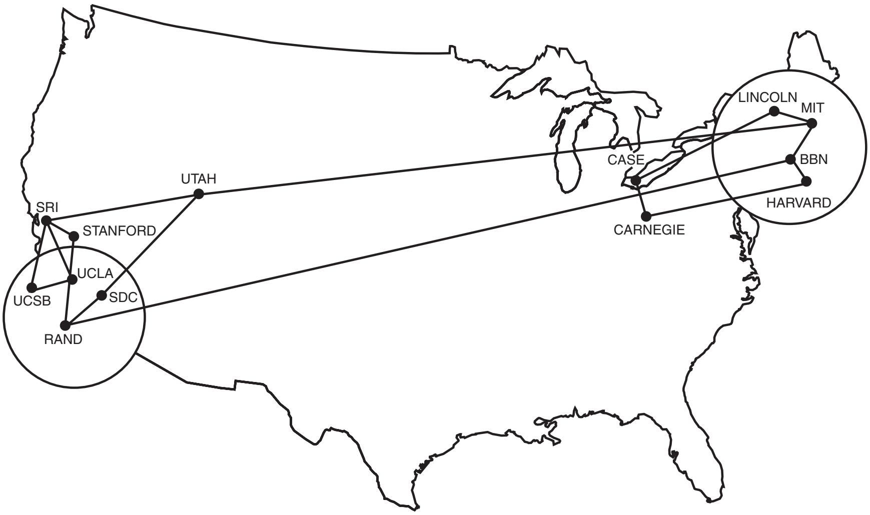
An *Internet Service Provider* or *ISP* is any computer that provides access to the Internet, or provides the functions that enable different parts of the Internet to connect to each other. Sometimes the organizations that run those computers are also called ISPs. Your ISP at home is likely your telephone or cable company, though if you live in a rural area, it might be a company providing Internet services by satellite. Universities and big companies are their own ISPs. The "service" may be to convey messages between computers deep within the core of the Internet, passing messages until they reach their destination. In the United States alone, there are thousands of ISPs, and the system works as a whole because they cooperate with each other.

Fundamentally, the Internet consists of computers sending bit packets that request services, and other computers sending packets back in response. Other metaphors can be helpful, but the service metaphor is close to the truth. For example, you don't really "visit" the web page of a store, like a voyeuristic tourist peeking through the store window. Your computer makes a very specific request of the store's web server, and the store's web server responds to it—and may well keep a record of exactly what you asked for, adding the new information about your interests to the record it already has from your other "visits." Your "visits" leave fingerprints!

IP Addresses

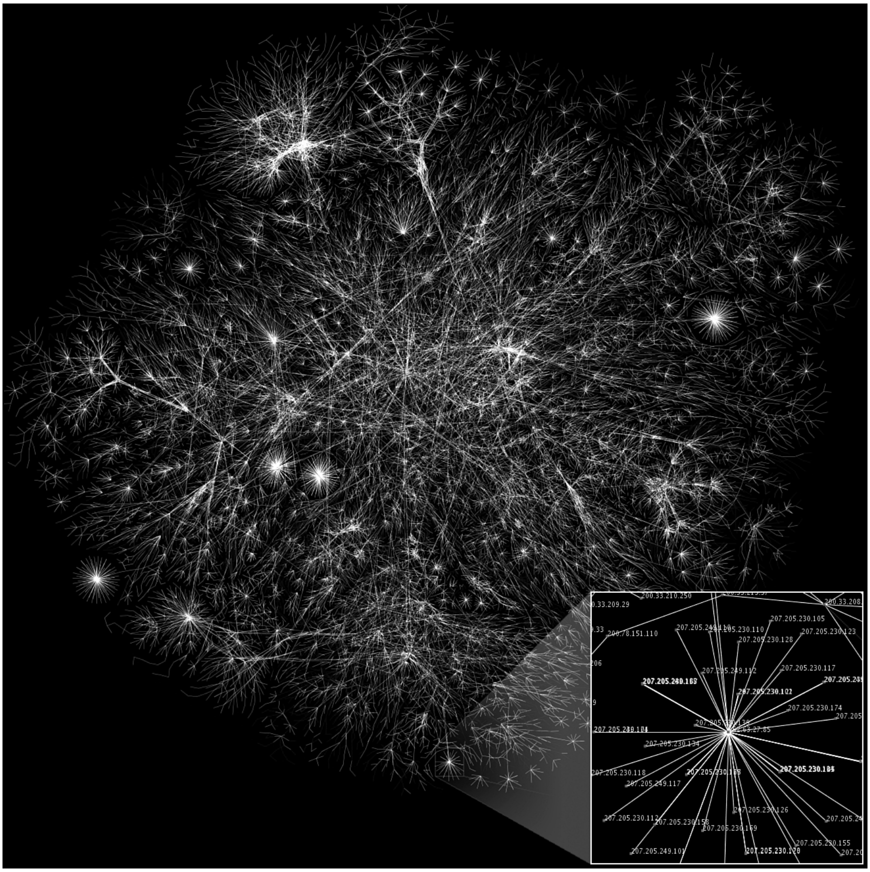
Packets can be directed to their destination because they are labeled with an *IP address*, which is a sequence of four numbers, each between 0 and 255. (The numbers from 0 to 255 correspond to the various sequences of 8 bits, from 00000000 to 11111111, so IP addresses are really 32 bits long. “IP” is an abbreviation for “Internet Protocol,” explained next.) A typical IP address is 66.82.9.88. Blocks of IP addresses are assigned to ISPs, which in turn assign them to their customers.

There are 256 × 256 × 256 × 256 possible IP addresses, or about 4 billion. In the pre-miniaturization days when the Internet was designed, that seemed an absurdly large number—enough so every computer could have its own IP address, even if every person on the planet had his or her own computer. Figure A.1 shows the 13 computers that made up the entire network in 1970. As a result of miniaturization and the inclusion of cell phones and other small devices, the number of Internet devices is already in the hundreds of millions (see Figure A.2), and it seems likely that there will not be enough IP addresses for the long run. A project is underway to deploy a new version of IP in which the size of IP addresses increases from 32 bits to 128—and then the number of IP addresses will be a 3 followed by 38 zeroes! That’s about ten million for every bacterium on earth.



Source: Heart, F., McKenzie, A., McQuillian, J., and Walden, D., ARPANET Completion Report, Bolt, Beranek and Newman, Burlington, MA, January 4, 1978.

FIGURE A.1 The 13 interconnected computers of the December, 1970 ARPANET (as the Internet was first known). The interconnected machines were located at the University of California campuses at Santa Barbara and at Los Angeles, the Stanford Research Institute, Stanford University, Systems Development Corporation, the RAND Corporation, the University of Utah, Case Western Reserve University, Carnegie Mellon University, Lincoln Labs, MIT, Harvard, and Bolt, Beranek, and Newman, Inc.



Source: Wikipedia, `http://en.wikipedia.org/wiki/Image: Internet_map_1024.jpg`. This work is licensed under the Creative Commons Attribution 2.5 License.

FIGURE A.2 Traffic flows within a small part of the Internet as it exists today. Each line is drawn between two IP addresses of the network. The length of a line indicates the time delay for messages between those two nodes. Thousands of cross-connections are omitted.

An important piece of the Internet infrastructure are the *Domain Name Servers*, which are computers loaded with information about which IP addresses correspond to which "domain names" such as `harvard.edu`, `verizon.com`, `gmail.com`, `yahoo.fr` (the suffix in this case is the country code for France), and `mass.gov`. So when your computer sends an email or requests a web page, the translation of domain names into IP addresses takes place before the message enters the core of the Internet. The routers don't know about domain names; they need only pass the packets along toward their destination IP address numbers.

IP Addresses and Crimes

The recording industry identifies unlawful music downloads by the IP addresses to which the bits are sent. But an IP address is rarely the exclusive property of an individual, so it is hard to be sure who is doing the downloading. A provider of residential Internet service allocates an address to a home only temporarily. When the connection becomes inactive, the address is reclaimed so someone else can use it. If NAT is in use or if many people use the same wireless router, it can be impossible to establish reliably who exactly used an IP address. If you don't activate the security on your home wireless router, neighbors who poach your home network signal may get you in serious trouble by their illegal downloads!

An enterprise that manages its own network can connect to the Internet through a single gateway computer, using only a single IP address. Packets are tagged with a few more bits, called a "port" number, so that the gateway can route responses back to the same computer within the private network. This process, called *Network Address Translation* or *NAT*, conserves IP addresses. NAT also makes it impossible for "outside" computers to know which computer actually made the request—only the gateway knows which port corresponds to which computer.

The Key to It All: Passing Packets

At heart, all the core of the Internet does is to transmit packets. Each router has several links connecting it to other routers or to the "edge" of the network. When a packet comes in on a link, the router very quickly looks at the destination IP address, decides which outgoing link to use based on a limited Internet "map" it holds, and sends the packet on its way. The router has some memory, called a *buffer*, which it uses to store packets temporarily if they are arriving faster than they can be processed and dispatched. If the buffer fills up, the router just discards incoming packets that it can't hold, leaving other parts of the system to cope with the data loss if they choose to.

Packets also include some redundant bits to aid error detection. To give a simple analogy, suppose Alice wants to guard against a character being smudged or altered on a post card while it is in transit. Alice could add to the text on the card a sequence of 26 bits—indicating whether the text she has put on the card has an even or odd number of As, Bs, ..., and Zs. Bob can check whether the card seems to be valid by comparing his own reckoning with the 26-bit "fingerprint" already on the card. In the Internet, all the routers do a similar integrity check on data packets. Routers discard packets found to have been damaged in transit.

The format for data packets—which bits represent the IP address and other information about the packet, and which bits are the message itself—is part of the *Internet Protocol*, or IP. Everything that flows through the Internet—web pages, emails, movies, VoIP telephone calls—is broken down into data packets. Ordinarily, all packets are handled in exactly the same way by the routers and other devices built around IP. IP is a "best effort" packet delivery protocol. A router implementing IP tries to pass packets along, but makes no guarantees. Yet guaranteed delivery is possible within the network as a whole—because other protocols are layered on top of IP.

Protocols

A "protocol" is a standard for communicating messages between networked computers. The term derives from its meaning in diplomacy. A diplomatic protocol is an agreement aiding in communications between mutually mistrustful parties—parties who do not report to any common authority who can control their behavior. Networked computers are in something of the same situation of both cooperation and mistrust. There is no one controlling the Internet as a whole. Any computer can join the global exchange of information, simply by interconnecting physically and then following the network protocols about how bits are inserted into and extracted from the communication links.

The fact that packets can get discarded, or "dropped" as the phrase goes, might lead you to think that an email put into the network might never arrive. Indeed emails can get lost, but when it happens, it is almost always because of a problem with an ISP or a personal computer, not because of a network failure. The computers at the edge of the network use a higher-level protocol to deliver messages reliably, even though the delivery of individual packets within the network may be unreliable. That higher-level protocol is called "Transport Control Protocol," or TCP, and one often hears about it in conjunction with IP as "TCP/IP."

To get a general idea of how TCP works, imagine that Alice wants to send Bob the entire text of *War and Peace* on postcards, which are serial numbered so Bob can reassemble them in the right order even if they arrive out of order. Postcards sometimes go missing, so Alice keeps a copy of every postcard she puts in the mail. She doesn't discard her copy of a postcard until she has received word back from Bob declaring that he has received Alice's postcard. Bob sends that word back on a postcard of his own, including the serial number of Alice's card so Alice knows which card is being confirmed. Of course, Bob's confirming postcards may get lost too, so Alice keeps track of when she sent her postcards. If she doesn't hear anything back from Bob within a certain amount of time, she sends a duplicate postcard. At this point, it starts getting complicated: Bob has to know enough to ignore duplicates, in case it was his acknowledgment rather than Alice's original message that got lost. But it all can be made to work!

TCP works the same way on the Internet, except that the speed at which packets are zipping through the network is extremely fast. The net result is that email software using TCP is failsafe: If the bits arrive at all, they will be a perfect duplicate of those that were sent.

TCP is not the only high-level protocol that relies on IP for packet delivery. For "live" applications such as streaming video and VoIP telephone calls, there is no point in waiting for retransmissions of dropped packets. So for these applications, the packets are just put in the Internet and sent on their way, with no provision made for data loss. That higher-level protocol is called UDP, and there are others as well, all relying on IP to do the dirty work of routing packets to their destination.

The postal service provides a rough analogy of the difference between higher-level and lower-level protocols. The same trucks and airplanes are used for carrying first-class mail, priority mail, junk mail, and express mail. The loading and unloading of mail bags onto the transport vehicles follow a low-level protocol. The handling between receipt at the post office and loading onto the transport vehicles, and between unloading and delivery, follows a variety of higher-level protocols, according to the kind of service that has been purchased.

In addition to the way it can be used to support a variety of higher-level protocols, IP is general in another way. It is not bound to any particular physical medium. IP can run over copper wire, radio signals, and fiber optic cables—in principle, even carrier pigeons. All that is required is the ability to deliver bit packets, including both the payload and the addressing and other "packaging," to switches that can carry out the essential routing operation.

IP over Carrier Pigeon

You can look up RFC 1149 and RFC 2549 on the Web, "Standard for the Transmission of IP Datagrams on Avian Carriers" and "IP over Avian Carriers with Quality of Service." They faithfully follow the form of true Internet standards, though the authors wrote them with tongue firmly planted in cheek, demurely stating, "This is an experimental, not recommended standard."

There is a separate set of "lower-level protocols" that stipulate how bits are to be represented—for example, as radio waves, or light pulses in optic fibers. IP is doubly general, in that it can take its bit packets from many different physical substrates, and deliver those packets for use by many different higher-level services.

The Reliability of the Internet

The Internet is remarkably reliable. There are no "single points of failure." If a cable breaks or a computer catches on fire, the protocols automatically reroute the packets around the inoperative links. So when Hurricane Katrina submerged New Orleans in 2005, Internet routers had packets bypass the city. Of course, no messages destined for New Orleans itself could be delivered there.

In spite of the redundancy of interconnections, if enough links are broken, parts of the Internet may become inaccessible to other parts. On December 26, 2006, the Henchung earthquake severed several major communication cables that ran across the floor of the South China Sea. The Asian financial markets were severely affected for a few days, as traffic into and out of Taiwan, China, and Hong Kong was cut off or severely reduced. There were reports that the volume of spam reaching the U.S. also dropped for a few days, until the cables were repaired!

Although the Internet *core* is reliable, the computers on the edge typically have only a single connection to the core, creating single points of failure. For example, you will lose your home Internet service if your phone company provides the service and a passing truck pulls down the wire connecting your house to the telephone pole. Some big companies connect their internal network to the Internet through two different service providers—a costly form of redundancy, but a wise investment if the business could not survive a service disruption.

The Internet Spirit

The extraordinary growth of the Internet, and its passage from a military and academic technology to a massive replacement for both paper mail and telephones, has inspired reverence for some of its fundamental design virtues. Internet principles have gained status as important truths about communication, free expression, and all manner of engineering design.

The Hourglass

The standard electric outlet is a universal interface between power plants and electric appliances. There is no need for people to know whether their power is coming from a waterfall, a solar cell, or a nuclear plant, if all they want to