Chapter Two

THE INTERNET

2.1 What is this Chapter About?

This chapter attempts to provide a simple description of how the Internet works, and the principles upon which it operates.

Such an exercise may seem superfluous to those who live by computers and whose careers derive from modern digital telecommunications networks. However this may be, it has been our experience that people persist in applying to new phenomena, such as the Internet, models of how the world works based on obsolete assumptions. Many difficulties arise, for those accustomed to telecommunications networks, from the fact that the Internet was designed from its inception to be radically different from telephone networks. Much of the Internet's apparent behaviour cannot be understood unless one sheds assumptions derived from voice-based, circuit switched telephony.

The Internet is an unplanned confluence of two important phenomena: computers and telecommunications. But it is also more than this. Telephone networks were digitized - made to run on digital computers - without being recast, rebuilt, and reorganized. The Internet represents a redesign from first principles of how signals can be made to move, of how people can connect to one another through machines, and what value can be found in communicating.

The Internet is more than a fad in transmission technologies. It embodies the tying together of the available computer power of the world through a common grammar for machines. The means by which this linking has occurred is the subject of this chapter.

In this chapter we examine the following topics:

- The emergence of the Internet

- The basic concepts of how the Internet works

- Why the Internet model has prevailed over other proprietary network models

The first things to notice are that:

- The designers of the Internet were concerned with how to make computers communicate across different platforms, software configurations, and makes;
- the Internet was built around the transmission characteristics of data, and not the human voice; and
- from these two design parameters result most of the differences between the Internet and voice telephone networks.

2.2 What is the Internet?

The Internet is frequently described as a network of networks, a term which confuses more than it clarifies. The official definition was announced in a resolution of October 24, 1995, when the U.S. Federal Networking Council, a body of Internet architects, unanimously passed a resolution

defining the term Internet. This definition was developed in consultation with members of the Internet and intellectual property rights communities.

> RESOLUTION: The Federal Networking Council (FNC) agrees that the following language reflects our definition of the term "Internet".
>
> "Internet" refers to the global information system that –
>
> i. i. is logically linked together by a globally unique address space based on the Internet Protocol (IP) or its subsequent extensions/follow-ons;
>   ii. is able to support communications using the Transmission Control Protocol/Internet Protocol (TCP/IP) suite or its subsequent extensions/follow-ons, and/or other IP-compatible protocols; and
>   iii. provides, uses or makes accessible, either publicly or privately, high level services layered on the communications and related infrastructure described herein.

The terms of importance that will be examined in that definition are:

- Globally unique address space
- Based on the Internet protocol
- Able to support communications using the Transmission Control Protocol/Internet Protocol (TCP/IP)
- Provides services *layered* on the communications infrastructure – the operative word is 'layered'

A term not found in this definition but which is at the core of the Internet is

- *packet switching.*

Another important attribute of the Internet is the extent to which it is

- *private.*

The Internet is distinct from the public switched telephone networks not only by design, but by ownership and legal destination. Tony Rutkowski, a noted Internet authority, defines the Internet as

Although access to the Internet takes place in most cases today via the public switched telephone network, once a signal has been passed through to the Internet, it is traversing a series of privately owned networks and hosts, which are not part of the public switched telephone system or the cable distribution infrastructure of broadcasting.

These private networks number over one million. They include not only the big backbone networks like UUNet and Teleglobe, but also every LAN (local area network) in every corporation, government department, and university. Any LAN that has a permanent connection with an Internet service provider (ISP) is considered an integral part of the global Internet.

The Internet is an amorphous entity which includes every private and public network that has agreed to exchange communication using the TCP/IP protocol, so that things we call "backbone networks" are operated on the same principles and are no different in nature from a corporate or campus LAN. The same shapelessness also makes it difficult to define with precision who is or is not an Internet service provider. Every corporate and academic LAN in the world is potentially an integral and seamless part of the Internet.

As the definition supplied by the Federal Networking Council is more or less authoritative, the underlying ideas and operations of the Internet will be described in those terms.

2.2.1. The Origins of the Internet

The purpose of this chapter is to explore the fundamental ideas incorporated into the architecture of the Internet. Accordingly it touches upon the history of the Internet only for the purpose of relating these ideas. The Internet abounds with web sites depicting the early history of the Internet. The most authoritative source is "A Brief History of the Internet", written by the original network architects themselves. Their version begins in 1962.

> "The first recorded description of the social interactions that could be enabled through networking was a series of memos written by J.C.R. Licklider of MIT in August 1962 discussing his "Galactic Network" concept. He envisioned a globally interconnected set of computers through which everyone could quickly access data and programs from any site. In spirit, the concept was very much like the Internet of today. Licklider was the first head of the computer research program at DARPA (Defense Advanced Research Projects Agency), starting in October 1962. While at DARPA he convinced his successors at DARPA, Ivan Sutherland, Bob Taylor, and MIT researcher Lawrence G. Roberts, of the importance of this networking concept.

> "Leonard Kleinrock at MIT published the first paper on packet switching theory in July 1961 and the first book on the subject in 1964. Kleinrock convinced Roberts of the theoretical feasibility of communications using packets rather than circuits, which was a major step along the path towards computer networking. The other key step was to make the computers talk

together. To explore this, in 1965 working with Thomas Merrill, Roberts connected a computer in Massachusetts to another in California with a low speed dial-up telephone line creating the [first (however small) wide-area computer network ever built](). *The result of this experiment was the realization that the time-shared computers could work well together, running programs and retrieving data as necessary on the remote machine, but that the circuit switched telephone system was totally inadequate for the job.* (emphasis added) Kleinrock's conviction of the need for *packet switching* was confirmed.

The basic idea of the Internet was clear from its beginning: to get computers to communicate, independently of their internal architectures or their manufacturers. Its design philosophy was open communication. From the beginning, its inventors decided that the circuit-switched architecture of the telephone system was inadequate to the task of allowing computers to communicate. They set about to redesign how signals can be made to move.

2.2.2 Packet Switching versus circuit-switching

Packet switching is the first of the major concepts introduced in our examination of the Internet. The term is like 'horseless carriage', or 'wireless'; it tries to define the new thing in terms of what preceded it, and what it is not. In this case, the contradiction lies in the relationship of packets to switches. If you have packets, you do not need switches. You need something different to assist the movement of the signal to its destination, a device which has come to be called the *router*. Routers have the same relationship to telephone switches as highway direction signs have to railway switches. And indeed, the comparison between the directedness, control, and central command needed to run a railway versus the freedom and autonomy of the driver on a highway can be extended to the characteristic differences between the Internet and telephone systems, between a packet-switched system and a circuit-switched system.
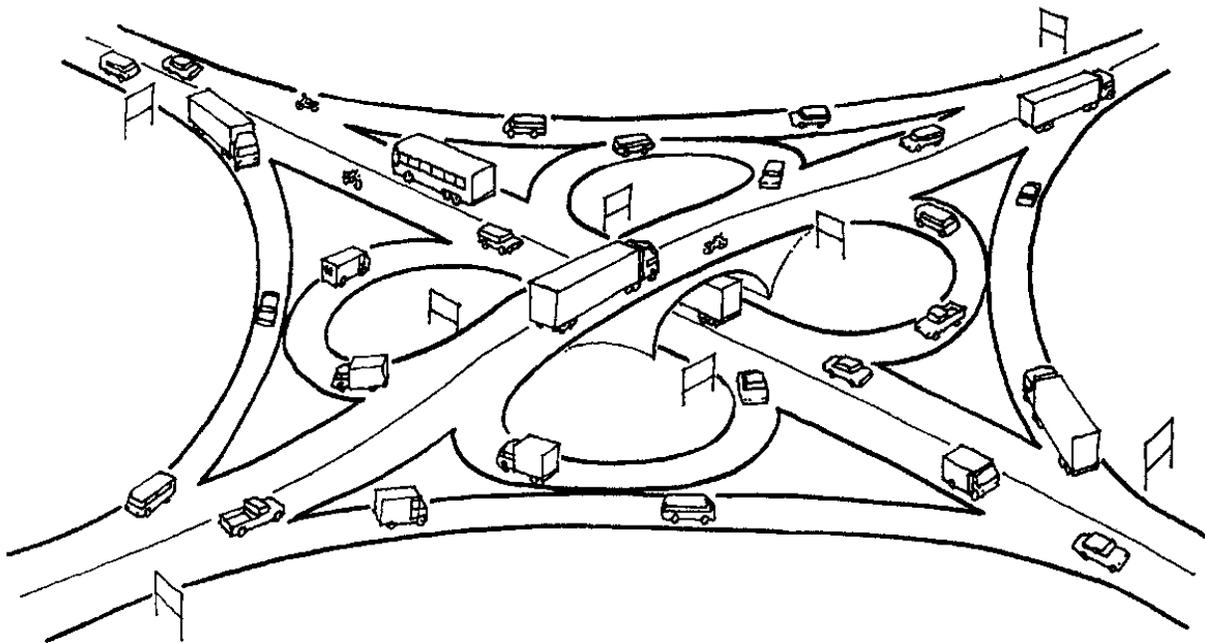
Figure 1

Circuit Switching is like a Railway
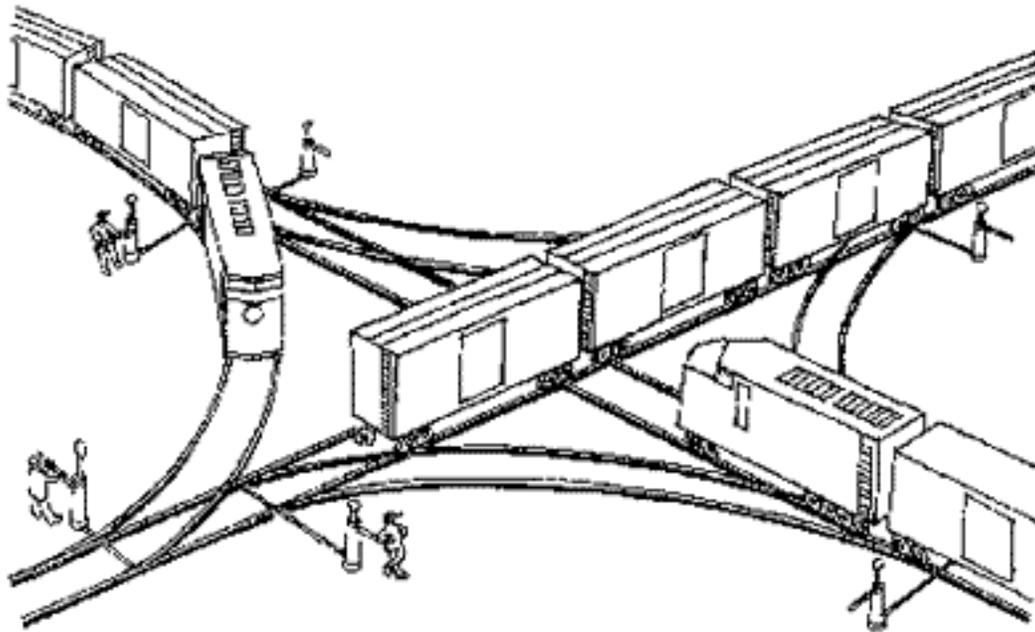
In a circuit-switched system, everything is specified. Control is the essence of making signals move reliably, and everything is designed for reliability. Everything moves by permission.

Figure 2

A Packet/Router System is Like a Highway

In a packet/routing system, very little is specified. Different vehicles (packets) can travel across the same system. They can get on or off the highway without permission. Less reliability = more freedom.

In a packet-switched system, computer power is used to break a signal down into packets. (This will be illustrated further below.) Each packet is individually addressed and routed across the network to its destination where the message is reassembled. Packets that do not arrive at their destination are automatically retransmitted. Each packet is like a car on a highway; it knows where it eventually wants to end up, but it will take direction from the highway signs. The computer address of the person is embedded in the header of each packet. Routers direct the packet towards the address shown in the header. By the same token the highway signs (the routers) can give authoritative direction to the packet to route around obstruction or congestion. Hence the origin of the phrase: "the Internet interprets censorship as obstruction and routes around it." A system designed to withstand the effects of nuclear war (that is, entire cities missing from the loop) turned out to be robust indeed.

Figure 3

The Router Gives Direction, but does not switch the packet
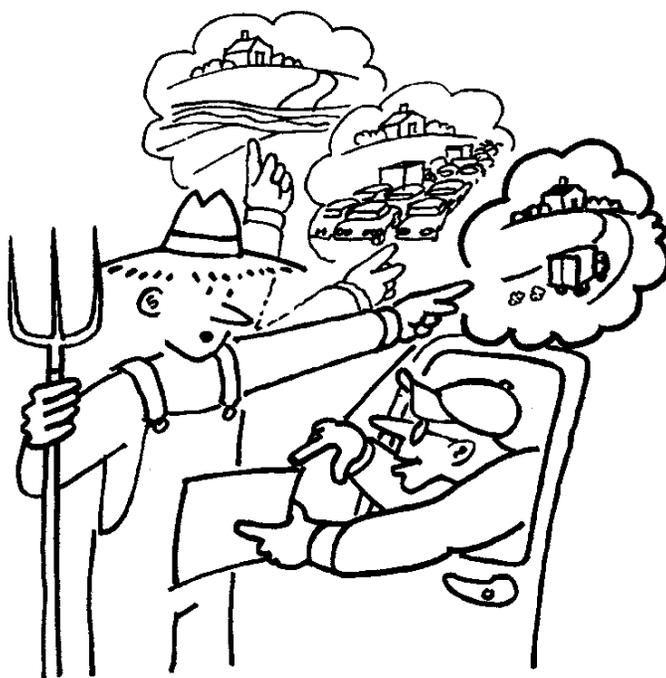
One way of expressing the difference between a packet and a circuit-switched system is to say that, in a packet-switched system, part of the intelligence has been put into the signal itself, whereas in a circuit-switched system, the intelligence is all in the network. The packet and the router work together to deliver the signal.

The same network architecture also means that the Internet tolerates more congestion and crashes than does the circuit-switched telephone system, since the packets have more 'free will' than the data trains inside the circuit-switched phone system. An overburdened Internet slows down, just like traffic at rush hour, generally without signal loss.

Another way of expressing this concept is that the packet-switched system of the Internet uses a connectionless, adaptive routing system, which means that a dedicated end-to-end channel need not be

established for each communication. Circuit switching, no matter how electronic or modern, still relies on an intelligent network to perform these functions, just as railway cars need a switch operator to cause the train to move on to the right tracks.

Figure 4

Routing Allows Different Paths to the Destination
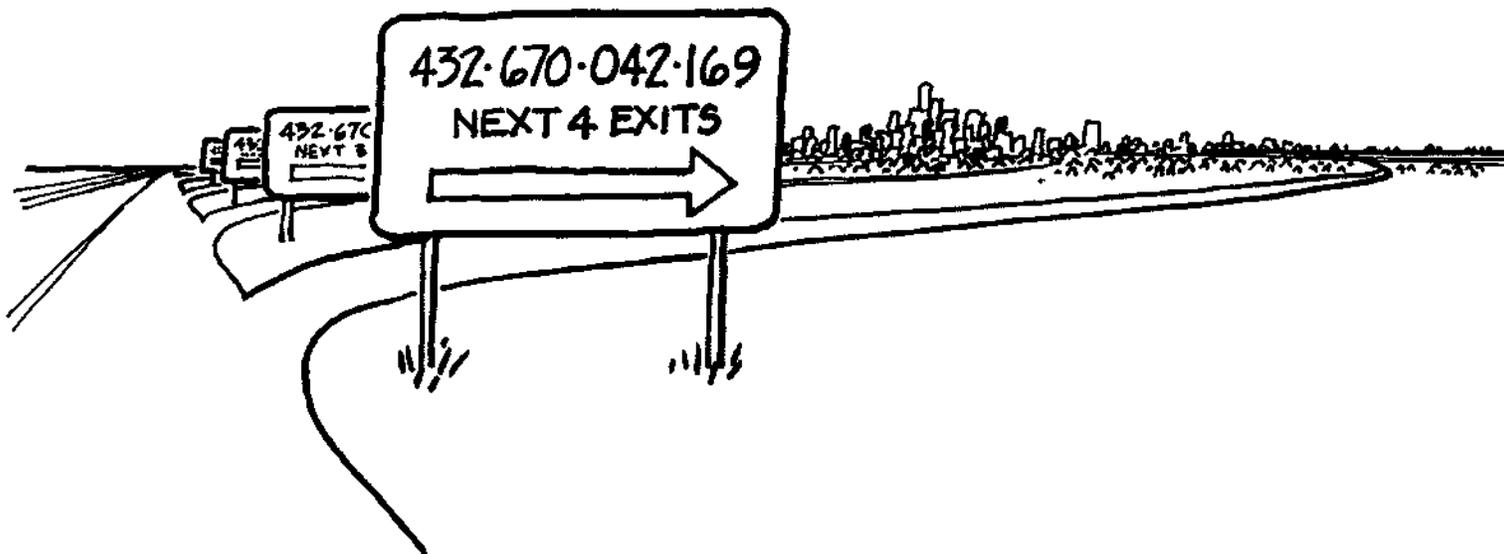
Packets can be in different sizes (most normal Internet traffic) or they can be made into uniform sizes (as in ATM) for efficiency. Many different communications can use the system simultaneously, just as cars, trucks and many different types of vehicle can use a highway simultaneously. The system is *underspecified* to borrow a term from engineering. By contrast, the telephone system is completely specified.

Figure 5

Circuit Switched versus Packet Routed
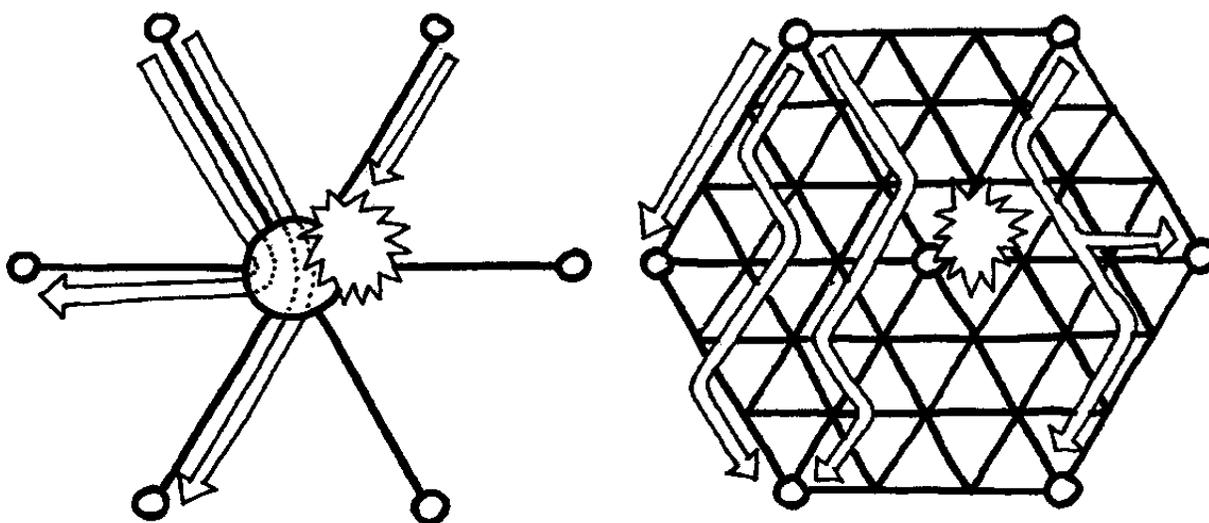
Circuit-Switched Packet-Routed

In a circuit-switched system, the signal depends on the availability of end-to-end bandwidth to allow a connection. The circuit is opened through a series of switches by telephone numbers.

In the packet-routed system, signals can go around a blockage. Packets are routed by the most efficient path. Packets may be sent along different routes and packets may arrive out of order.

Yet another way of saying the same thing is that Internet communications float above the physical facilities used to transport the signal. Layers of signal are disaggregated from the physical layer, be it wire, coaxial cable, or the air, over which the signal is carried. This aspect will be discussed further below in the section on *layers.*

A packet-switched system was inconceivable before computers, since disassembly and reassembly of the signal can only be performed by computers.

The telephone system embodies designs that were necessary when communication was electro-mechanical, and the signal was analog. In an analog signal, the variations in its wave-form have a one-for-one correspondence with the information being transmitted. Subsequent computerization of telephone equipment did not result in a redesign of the basic ideas of what a signal was or how it could be addressed. The functions performed by a telephone company were essential to telecommunications, in this earlier model, since a circuit-switched model requires someone to set up, maintain, and take down the call.

Telephone systems have been built on switched circuits for their entire history. A call is placed, a series of circuits are opened by the interaction of the telephone number with the switching system, and the call is taken down – all the circuits are shut again – when the call is over. The telephone system was engineered on certain premises: that switching and memory were expensive, and that calls would last on average a certain amount of time - the average conversation lasting about three minutes. The public switched telephone network was designed around intermittence of [voice] calling and scarcity of computer memory. Data traffic of the kind generated by the Internet upsets all these assumptions around which the public voice telephone network (PSTN) was engineered.

In the opinion of those who founded the Internet, the then predominant means of conveying signals, which was the opening and closing circuits by means of switches, was inadequate for the job of computer communications, and they set out to devise a new method.

The importance of this discovery is now making itself felt 35 years later. Suppliers of telephone switching equipment are scrambling to change their products into packet switched Internet Protocol devices.

Nortel Networks, one of the world's largest makers of circuit-switched telephone systems, has bought the third largest maker of equipment that links computers to the Internet, Bay Systems, and installed the President of Bay Systems as the heir apparent at Nortel Networks. Nortel Networks is buying its way into

packet switched networks.

AT&T has announced its purchase of the second largest US cable television company, Tele-Communications Inc. The analysis explained the benefits of the deal as follows:

> "Yet both companies know the network future lies not in today's telephone-network or cable-system technologies but in providing consumers and businesses with high-speed network access based on Internet technology – whether for data transfers, voice conversations or, eventually, even TV-quality video. While Internet traffic can flow over phone wires or cable lines or even radio waves, the Internet employs a transmission format that is far more efficient and flexible than conventional telephone or cable systems and seems destined eventually to render those conventional systems obsolete."

2.2.3 Open Architecture based on the Internet Protocol

The next major feature of the Internet to emerge was open architecture, and with it, the idea of communication among peers. The founders of the Internet again:

> "The original ARPANET grew into the Internet….The Internet as we now know it embodies a key underlying technical idea, namely that of open architecture networking. In this approach, the choice of any individual network technology was not dictated by a particular network architecture but rather could be selected freely by a provider and made to interwork with the other networks… Up until that time there was only one general method for federating networks. This was the traditional circuit switching method where networks would interconnect at the circuit level, passing individual bits on a synchronous basis along a portion of an end-to-end circuit between a pair of end locations. Recall that Kleinrock had shown in 1961 that packet switching was a more efficient switching method. Along with packet switching, special purpose interconnection arrangements between networks were another possibility. While there were other limited ways to interconnect different networks, they required that one be used as a component of the other, rather than acting as a *peer* of the other in offering end-to-end service….

> "The idea of open-architecture networking was first introduced by [Bob] Kahn shortly after having arrived at DARPA in 1972….

> Kahn decided to develop a new version of the protocol which could meet the needs of an open-architecture network environment. This protocol would eventually be called the Transmission Control Protocol/Internet Protocol (TCP/IP)….

"Four ground rules were critical to Kahn's early thinking:

- Each distinct network would have to stand on its own and no internal changes could be required to

any such network to connect it to the Internet.

- Communications would be on a best effort basis. If a packet didn't make it to the final destination, it would shortly be retransmitted from the source.

- Black boxes would be used to connect the networks; these would later be called *gateways* and *routers*. There would be no information retained by the gateways about the individual flows of packets passing through them, thereby keeping them simple and avoiding complicated adaptation and recovery from various failure modes.

- There would be no global control at the operations level."

It can be seen that the major features of the Internet were laid down from its inception. Total connectivity of networks was its goal. All vendors and all platforms are treated as equal. All operating systems are treated as equal.

The system is robust and simple. There are no records kept of what passes through the gateways, and, unlike the telephone system, there is no overall control of operations of the system.

2.2.4 Able to support communications using the Transmission Control Protocol/Internet Protocol (TCP/IP)

The goals of the architects of the Internet are achieved by the protocol by which they developed, and by means of which the Internet operates, known as Transmission Control Protocol/ Internet Protocol (TCP/IP, pronounced as the letters would be pronounced in English, tee see pee eye pee). Its key feature was to allow multiple networks to connect to each other. When in the 1970's developers of the UNIX operating code at the University of California at Berkeley added TCP/IP into their software distribution kit, TCP/IP began a rapid growth spurt, especially in academic environments. In the early 1980's, the US Secretary of Defense mandated that all computers connected to the ARPANET had to use TCP/IP. From that moment forward, the Internet was born, because TCP/IP gave to every system in which it was embedded the ability to communicate with any other network transparently. Academics and other specialized users embraced TCP/IP because it was essentially no-cost networking, from a software point of view.

A protocol is a definition or set of rules for how computers will act when talking to each other. Protocol definitions range from how bits are placed on a wire to the format of an electronic mail message. Standard protocols allow computers from different manufacturers to communicate; the computers can use completely different software, providing that the programs running on both ends agree on what the data mean.

IP. The Internet Protocol provides a unique 32-bit address for each machine connected to the Internet, and handles addressing and forwarding of packets.

TCP. The Transmission Control Protocol enables two machines to transmit data back and forth in a manner coherent to the operating system of each. It defines the handling of packets, including segmentation, reassembly, concatenation, separation, and recovery of lost packets.

While there are other protocols for the interconnection of computers, TCP/IP has become the dominant partnership.

2.2.5 "Provides services layered *on the communications infrastructure": Grammar for Machines*

Of all the features of the Internet, the one that needs the most careful explaining and which has the most revolutionary implications is the existence of layers of protocol – which are really software instruction sets in the headers of the messages transported over the networks.

The existence of layers in data communications represents a subtle and powerful method for changing how people are able to take advantage of computer networks. Layers provide agreements among people – and the machines they build and program – as to who will do what, when.

Layers are standards. They consist of agreements about the instructions that will be contained in the headers of messages. They are a form of software. Consequently they partake of the economic characteristics of software.

The importance of this point cannot be emphasized sufficiently. The Internet is an open system where new software – new protocols – may be added by a process of consensus within the industry. The effect of these protocols can be to change how the system of signal transmission works. These protocols can be introduced *without one penny being spent on changing any physical object* within the signal transmission system.

Thus understanding the function of layers helps us to understanding why the Internet is driving technological and business change so effectively.

Knowledge of the existence of layers is fundamental to understanding how the Internet works, and why it works differently from previous signal transport media.

There is an organization called the International Standards Organization (ISO, pronounced eyeso) which has developed definitions of network architecture, called Open Systems Interconnect (pronounced o, ess, eye). There are seven layers in the OSI Reference Model. The seven layer OSI cake is an agreed way for computers to communicate, and can be understood as a grammar for machines.

Layer 1: layer one is the physical layer, where the electrical signals move around.

Layer 2: the data link layer. This is the layer that splits data into packets to be sent across the connection medium. The data link layer handles electromagnetic interference. Analog broadcast signals, for example, are not sent with the benefit of a data link layer, and hence they are much more susceptible to sunspot activity and other forms of interference.

Layer 3: the network layer. This layer gets packets from layer 2 and sends them to the correct network address. If more than one possible route is available for data to travel, the network layer figures the best route. The IP (Internet Protocol) works on this layer.

Layer 4: the transport layer. This layer makes sure that packets have no errors and that all the packets arrive and are in the correct order. The Transmission Control Protocol (TCP) works in this layer.

Layer 5: the session layer. A session is the name for a connection between two computers. This layer establishes and coordinates a session. The other protocols that make up TCP/IP sit on layer 5 and above.

Layer 6: the presentation layer. The presentation layer handles different file formats, so that file transfers can be effected between computers using different file formats.

Layer 7: the application layer. This is the level where ordinary mortals do their work: e-mail, requesting a file transfer, and so forth.

While the Internet uses a layered signal architecture, the founders of the Internet decided not to conform to the ISO seven layer model. Rather TCP/IP takes the top three layers, five through seven, and combines them into one, called the application layer. Finally, it is also useful to realize that the physical and data link layers have nothing to do with TCP/IP but TCP/IP must have these layers below it in order to work. The signal must transport across *something,* even the air.

Figure 6

The Internet Works by Means of Layers of Protocol over a Physical Medium
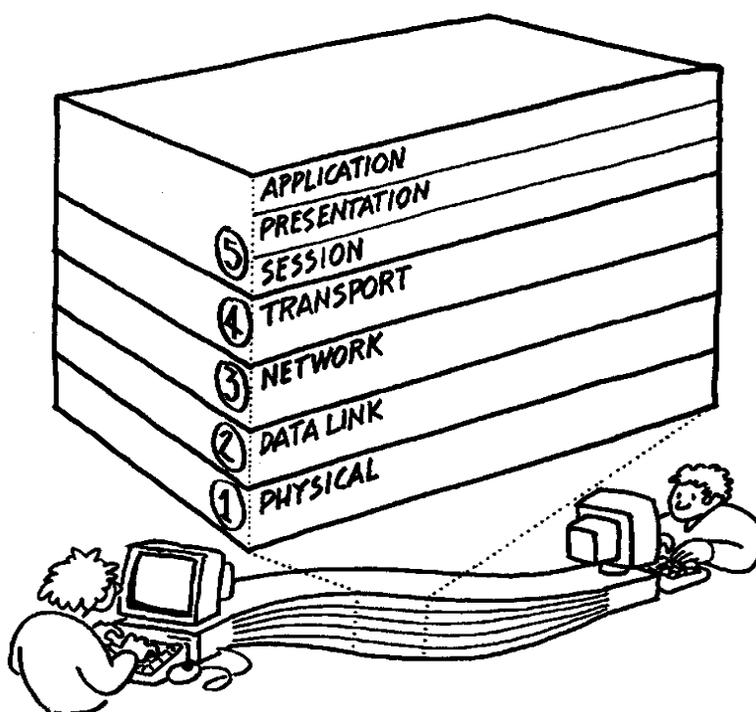
Note that the designers of the Internet combined the top three levels of the OSI's seven-layer reference model into five layers.
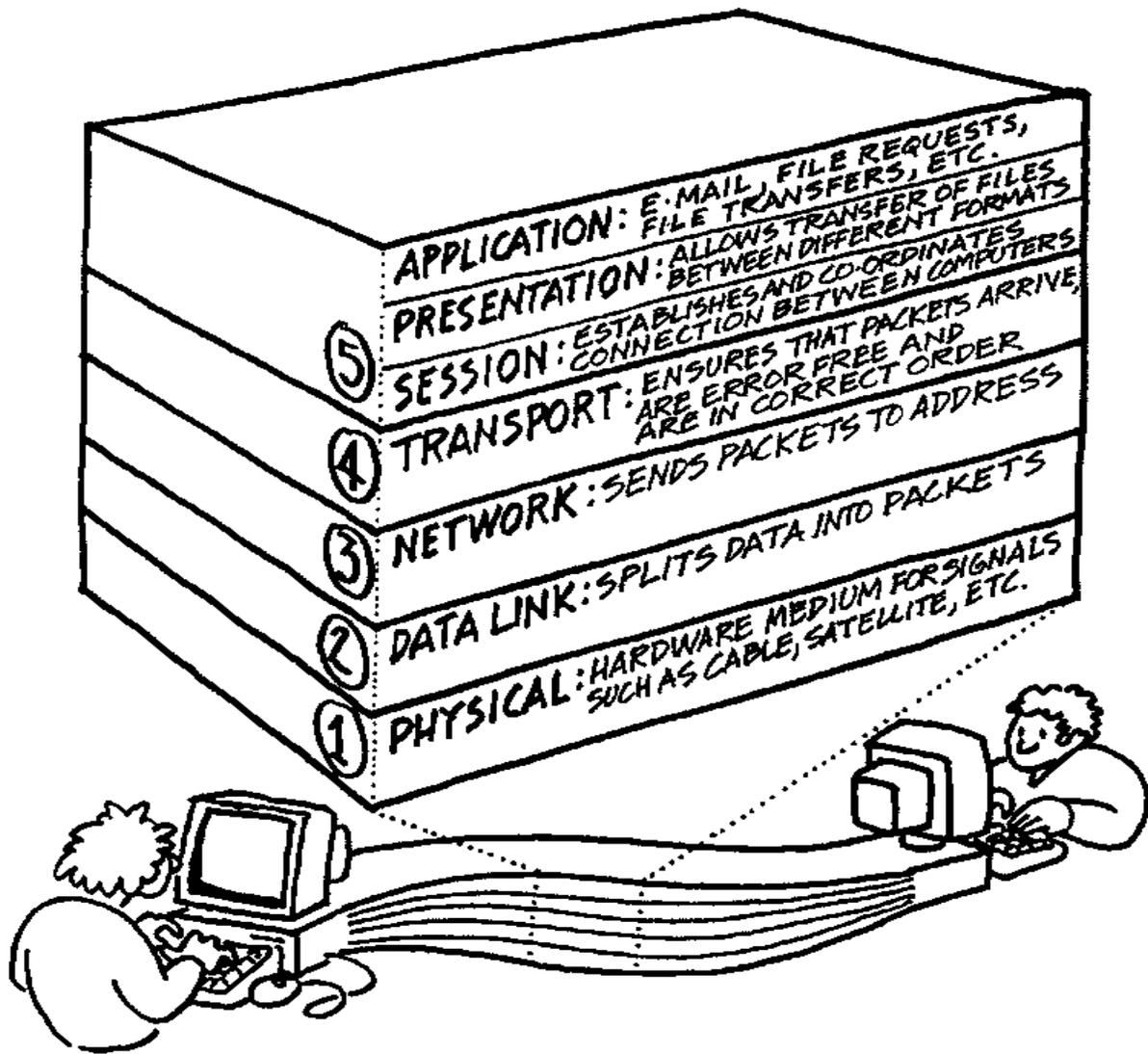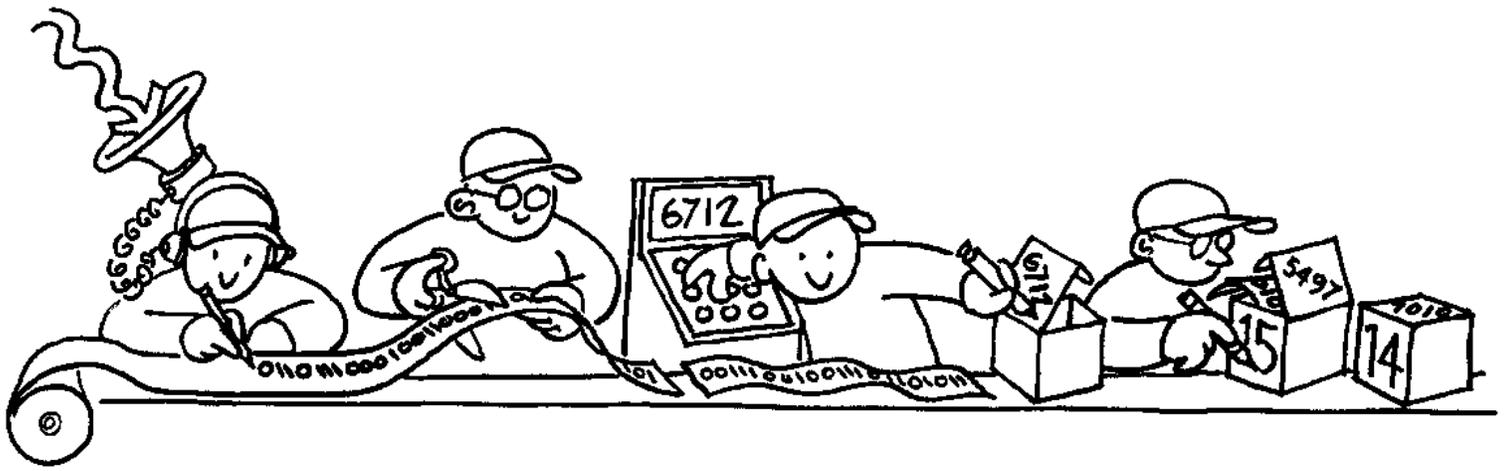
Figure 7

The Layers Perform Different Functions

Figure 8

The Functions of Internet Protocol Version 4

Data Link and Network Layers

Left to Right

An analog signal is digitized at the applications layer. The signal is broken into *packets* by the data link layer. The network layer counts the number of bits (the *checksum*), places the data in envelopes (packets), and gives them addresses, contained in the *header* (information on the boxes). Headers are ordered so that the sequence can be re-established at the receiving end.

Figure 9

The Network Layer
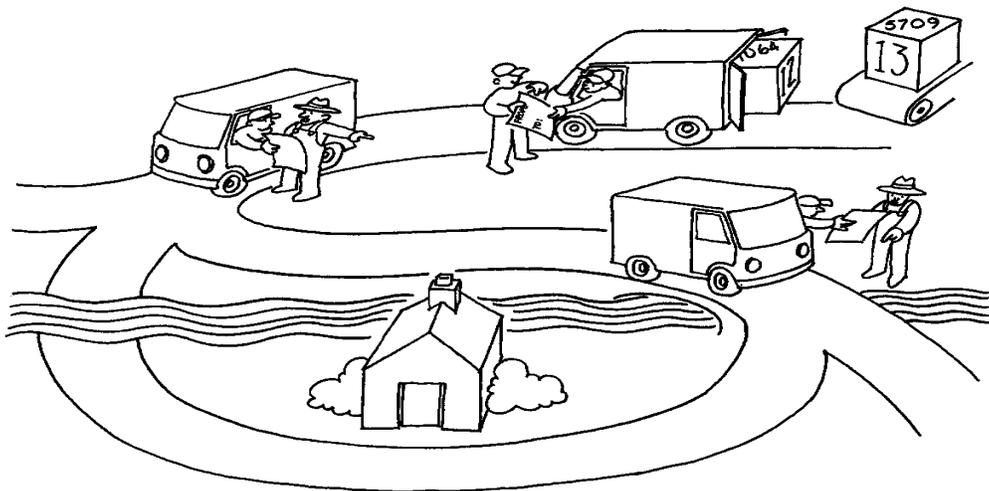
The packets are transported to their destination, with the assistance of routers. Despite the name, it is the network layer and not the transport layer that actually sends the packets to the correct address. They may arrive out of order or corrupted. The transport layer fixes this.

Figure 10

The message is re-Assembled

The Transport Layer

As the packets arrive at their destination, TCP calculates a checksum for each packet. It then compares this checksum that has been sent in the packet. If the checksums do not match, TCP discards the packet and asks that the original packet be retransmitted. When all the correct packets are received by the computer to which the information is being sent, TCP assembles them into their original, unified form.

Figure 11

The Session Layer

The Session Layer

The other protocols that make up TCP/IP sit on layer 5 and above. This layer established and coordinates a *session*, which is the name for a connection between two computers.

Figure 12

The Presentation Layer

The presentation layer works with the operating system and the file system. Files are converted from one format to another, as necessary. Without the presentation layer, file transfer would be restricted to computers with the same file format.

Figure 13

The Application Layer

APPLICATIONS.

This is the layer where people do their work, such as sending e-mail or requesting to transfer a file across the network.

The significance of a layers becomes evident when we see what it enables people to do with computer networks, and it becomes even clear when we contrast what users can do with a computer network running on open protocols versus what cannot be done on proprietary systems. The implications of layers and the software-nature of the Internet will be explored in sections 2.2.7 and 2.2.8 below.

2.2.6 A Globally Unique Address Space

The definition of the Internet made by the Federal Networking Council stated that it was a global information system that

> *is logically linked together by a globally unique address space based on the*
> *Internet Protocol (IP) or its subsequent extensions/follow-ons;*

Addresses were central to making computers communicate with each other. In this subsection of chapter two, we look at how the Internet routes traffic to IP addresses.

Internet Protocol assigns an IP number to every device on the net. If you have no IP number, you are not on the Internet. Every resource on the Internet has a unique IP address.

- Every computer (host) on the Internet is identified by an **IP number**

- Every IP number is **different** - if two computers had the same IP number, the network would not know which one to deliver data to.
- An IP number is a **32-bit binary number**, that is, a string of 32 ones and zeros.

These numbers are converted out of binary notation according a formula that need not concern us.

IP numbers include four address blocks of numbers consisting of numbers between 0 and 256, separated by periods. For an example of IP numbers, go into your Windows system: Start button, to Accessories, to Dial-up Networking. Click on your server, if you have one, and right click on "properties." The first box to click on is "server type", and embedded within that pop-up window will be "server settings". That set of numbers is the IP address of the domain name servers used by your Internet Service Provider. Your own IP number is dynamically assigned by the ISP for the duration of your hook-up from its bank of IP numbers.

The address blocks are separate bytes of a 32-bit address. The growth of the Internet has raised concerns that this number space will eventually be exhausted. As a result, the next version of the Internet's underlying protocol, referred to as IP version 6 or IPv6, includes a much larger 128-bit address space.

The current version of IP gives 256^4 addresses, or more than 4,294, 967,200 addresses.

The next version will accommodate enough IP addresses that the world population will be able to wear or own many devices, each with an IP address.

Domain Names

If the system of IP numbers could be remembered by human beings, we would not have to superimpose on the IP numbering system the system of *domain names* by which products, services, firms and other websites are being called up.

The Domain Name System (DNS) was invented in 1987 by Paul Mockapetris, and stems from work defined by Zaw-Sing Su at the Menlo Park Stanford Research Institute (SRI) and Jon Postel at the Information Sciences Institute (ISI) at the University of Southern California. Jon Postel had been keeping track of "well known numbers" used on the then ArpaNet since 1972 under a US Department of Defense contract.

In March 1992 the US National Science Foundation solicited bids for a five-year contract to run various network registration services. In January 1993 the NSF awarded a contract to a cooperative of Network Solutions, AT&T and a third party which has since become Network Solutions Inc. The contract was set to expire in September 1998.

In May 1994, Joshua Quittner published an article in Wired magazine and described how he registered *mcdonalds.com* and tried to sell it to Burger King. That month, domain registration requests shot up from 2000 a month to 8000 a month and the "great domain name gold rush" began and has not stopped since.

An enormous amount of controversy has arisen inside Internet circles concerning the future of the domain name system. The future management of the Internet is at stake. In legal terms, the issue concerns how the United States government will privatize the remaining functions, including especially domain name management, that have until now been within the jurisdiction of the National Science

Foundation. While these debates are of great importance, their outcome will not affect the fundamental technical nature of how the Internet works. They will help to sort out how much international influence will play in the management of domain name administration. It should be noted that, as the domain name system is privatized, an increasing number of root servers will be located outside North America. This will reduce the asymmetry of traffic considerably.

Basically, there are six top level domains: .com, .net, .org, .mil, .gov, and .edu, of which .mil and .gov are reserved for the United States government. There are also 144 national level domains, such as .ca, .uk, and so forth.

The operation of the domain name system complicates but does not change the basic features of the Internet. A request for a website – for example, APEC's own -apec.org.sg - is sent out from one's computer to one of a very few central computers, most of which are in the United States. These computers return a message to one's server indicating where the website can be found. For example, the top level directory (the *root server*) tells one's server that the supreme *.ca* directory is found at such and such location. One's server then directs its inquiry to the .ca directory. The *.ca* supreme directory then informs one's server that such and such as site, such as *crtc.gc.ca* is found at another server located at an IP address. The connection finally made, the website downloads onto one's screen.

It can be seen that this method ensures that there is plenty of long distance traffic on the Internet – not the PSTN – in the search for any given website. Indeed, more than 90% of Internet traffic transits through the United States because of several factors, one of the most important of which is that most domain name servers are located there.

Internet service providers can sometimes deal with the problem of frequently asked-for websites by storing them at sites closer to the customer. This is referred to as *caching.* An Internet Service Provider automatically will cache websites that customers have browsed on their servers for some period of time before they are deleted. Thus the Toronto *Globe and Mail* for instance will be cached in other Canadian cities because of high demands upon that set of pages from readers across the nation. Internet service providers or their customers can also decide to take the entire contents of a website and store it on their own servers. For instance, popular websites from the United States can be stored in the entirety in Australia, China or Europe, thus avoiding the transoceanic download of data from North America, where most Internet material still originates. This practice of copying a complete website onto a local server and changing its IP address is called *mirroring.*

2.2.7 What is the significance of layers?

The existence of this common grammar for communication between machines allows for people at the periphery of the network – and we are all at the periphery when it comes to machines – *to modify how the network will work.*

The significance of layers can be summarized in the following points:
- ❍ Layers are composed of protocols, which are of their nature software, in all layers above the physical transport layer.
  - ❍ They are therefore not physical objects but instructions and information embedded in the headers of signals and in the machines that read the headers and route them to their

destinations.

- ❍ Layers are developed in a collaborative and open process of commentary upon papers by technical experts. Their acceptance turns them into an industry standard.
- ❍ Changes in one layer will not necessarily affect other layers, unless this is designed into the software.
- ❍ The economics of changing protocols are therefore like the economics of software, the more people use it, the more other it becomes a standard, and once a standard, other software can be designed to run on it, in the same way that programs run on Microsoft Windows.
- ● ❍ The economics of telecommunications, and therefore the players in the game, can now be radically transformed.

For example, suppose you had the technical skill to write programs that would address a consumer's needs. You write a program that solves the problem, say, of knowing whether your friends are on the Internet or not. If you think other people will buy it, you put up a web site and sell it. If enough people buy it, you have created any of the following: a new network standard, a new business, or a new way of communicating. The grammar of machines, TCP/IP, has not changed. But all owners of computers have the ability to buy your product and run it. A common grammar for machines has the effect of creating a common market for all who use those machines. The advantage of the layered model, and a common protocol, is precisely this: no has had to change a single physical device to get a product to work.

The significance of layers can be understood better by seeing the process by which they are created. As was indicated above, layers allow for fundamental improvements in the technical characteristics of the signal transport system because, by segregating various functions from one another, various changes can be made in the protocols of one layer, which will not necessarily affect the operation of others.

The standards that cause the Internet to work are devised by a collaborative group of experts gathered under the title of the Internet Engineering Task Force. The Internet Engineering Task Force (IETF) is a large open international community of network designers, operators, vendors, and researchers concerned with the evolution of the Internet architecture and the smooth operation of the Internet. It is open to any interested individual. The actual technical work of the IETF is done in its working groups, which are organized by topic into several areas (e.g., routing, transport, security, etc.). The creation of new protocols within the session layer proceeds by way of papers put up for comment on email lists. Those with the technical capacity to comment do so.

Currently work is being conducted on the *session layer* within the IETF communities concerned with internet telephony. A recent paper by Professor Henning Schulzrinne, Department of Computer Science, Columbia University, and Jonathan Rosenberg, Bell Laboratories, Lucent Technologies, dated July 2, 1998, titled "Internet Telephony: Architecture and Protocols, an IETF Perspective". Let the authors describe the significance of their work.

> "Internet telephony, also known as voice over IP or IP Telephony, is the real-time delivery of voice (and possibly other multi-media data types) between two or more parties, across networks using Internet protocols, and the exchange of information required to control this delivery. Internet telephony offers the opportunity *to design a global multimedia communications system that may*

> *eventually displace the existing telephony infrastructure*, without being
> encumbered by the legacy of a century-old technology." (emphasis added)

In short, engineers are specifying the protocols by which the signals of the twenty-first century will move about. The exercise of creating protocols is a logical process of specifying what information and functions are to be carried or performed in the headers of the digitized bit streams that constitute data traffic. Below an illustration of the alphabet soup of protocols, called a *protocol stack*, the authors comment:

> "Even though the term Internet telephony is often associated with
> point-to-point service, none of the protocols described here are restricted to a
> single media type or unicast delivery. Indeed, one of the largest advantages of
> Internet telephony compared to the Plain Old Telephone System (POTS) is the
> transparency of the network to the media carried, so that adding new media
> type requires no changes to the network infrastructure."

Layering changes the relationship of users to the network

The effect of layers then translates further into economic power for those who can take advantage of the change. The Internet distinguishes different service layers in a parsimonious way, so that each layer can be applied in the widest possible variety of contexts. Clean functional differentiation among service layers, however, means that simple data transport may become a commodity business.

A particularly insightful observer, David Isenberg, a former employee of AT&T Research, has called this "the rise of the stupid network". Contrasting this with the telephone company paradigm of the "Intelligent Network", Isenberg writes:

> The Intelligent Network is a straight-line extension of …four assumptions … --
> scarcity, voice, circuit switching, and control. Its primary design impetus was
> not customer service. Rather, the Intelligent Network was a telephone company
> attempt to engineer vendor independence, more automatic operation, and some
> "intelligent" new services into existing network architecture. However, even as
> it rolls out and matures, the Intelligent Network is being superseded by a Stupid
> Network, with nothing but dumb transport in the middle, and intelligent
> user-controlled endpoints, whose design is guided by plenty, not scarcity,
> where transport is guided by the needs of the data, not the design assumptions
> of the network.

And further:

> The {telephone} network works as long as engineering assumptions (e.g., the
> length of a call, the number of call attempts, etc.) do not change. But let the
> assumptions change episodically (e.g., Rolling Stones tickets go on sale), or
> structurally (calls to Internet service providers last several times longer than
> voice calls), and the network hits its design limits - completing a call becomes a

matter of try, try again.

What if network design were based on another assumption - that computation and bandwidth were cheap and plentiful?

One of the direct results of layering is that it increases the power of users to configure the network to their purposes. Layers detach the manipulation of the software from the underlying transport facilities. One does not have to build one's own transmission system, nor does one have to modify equipment within it, to change how the system will work. If someone builds a better product or service, including a product or service that changes the way the network operates, then all they have to do is to offer it to the public, over the Internet or however they please. They are not obliged to get into the system and change every black box within the telecommunications network in order to make their idea work. The effect of layers is to allow developers to create new businesses and even new standards if enough people adopt the product or service. Remember that no one has to change the hardware on which the service is offered. If enough people buy your model – be it an e-mail program, a browser, a financial transaction software, then that becomes a standard.

Isenberg again:

A new network "philosophy and architecture" is replacing the vision of an Intelligent Network. The vision is one in which the public communications network would be engineered for "always-on" use, not intermittence and scarcity. It would be engineered for intelligence at the end-user's device, not in the network. And the network would be engineered simply to "Deliver the Bits, Stupid," not for fancy network routing or "smart" number translation.

Fundamentally, it would be a Stupid Network.

In the Stupid Network, the data would tell the network where it needs to go. (In contrast, in a circuit network, the network tells the data where to go.) In a Stupid Network, the data on it would be the boss….

End user devices would be free to behave flexibly because, in the Stupid Network the data is boss, bits are essentially free, and there is no assumption that the data is of a single data rate or data type.

In short, the aspect of the Internet that may have the most subtle and pervasive effect is the breaking down of the communications system into layers. What this accomplishes is to prevent anyone from gaining monopoly rents out of the exclusive possession of distribution channels, be they in "Intelligent Networks" or any other proprietary system. This has dramatic implications for what John Sidgmore, President of UUNet, calls the "central planning models" of telephony and broadcasting.

Figure 14

The two systems place intelligence in different places

Circuit-Switching Packet Switching

Intelligence in the Switch Intelligence at the Periphery

2.2.8. New Business Models

The Internet seems to provide new models for doing business. Much of the Internet and related products consists of software, whose utility is determined largely by how many people use it. Consequently, many successful market strategies for the Internet have involved giving the product away. The goal of every Internet entrepreneur is to turn a software *product* into a *standard*. This can be done by creating a vast installed base of the product, and the most effective way to do this is to give the product away, and let end users figure out its uses and advantages. Several companies are giving away the source codes to their products and letting thousands of users figure out improvements, of which Java and Linux are the foremost examples. Microsoft has given away its Internet Explorer browser software in an effort to catch up with Netscape's product.

Telecommunications networks and computer operating systems show both exhibit two important economic phenomena: network effects -- the benefit of using a given system increases as other people use it -- and economies of scale -- the price of the software can decrease rapidly as more people use it. This combination dramatically favors whichever system has the most users. But this creates the need to get to the market first with as cheap a product as possible, and then gradually add to it.

The Internet allows the separation of services from their underlying transport medium. In the Internet, services are decoupled from the transport layer. Hence layers provide for the openness of the system. In this model, a business can only make money when the services they supply are so good that no one wants to bypass them.

A further implication of the packet switched model is a change in how services are priced. Telecommunications services are priced on circuit-switched assumptions: a call is set up, circuits are

opened, a "call" is made, the longer the call, the more is charged; the more bandwidth is asked for, the higher the price. Services differ in price depending on the nature of the customer – business or residential. In a packet-switched environment, the transport layer is always 'on', there is no such thing as a continuous bit stream recognizable as a "call", resources are not consumed by the duration of any particular packet's travel, and customers cannot be distinguished by opening up the header and finding if they are business or residential IP numbers.

Also implied in the packet-switched model is that no one can open up the header of your message and determine what is being carried: voice, video, or sound. There is therefore no basis for price discriminations based on the nature of the signal traffic, only on the quality of service requested out of the network.

There is currently work underway in the IETF to develop protocols that will allow differentiated services, which would give higher priority to some packets. Such standards, if widely deployed, would enable new services, improve the reliability of existing services, such as Internet videoconferencing, and alleviate congestion by making better use of bandwidth.

There is a further implication of the layered architecture of the Internet for owners of transport facilities. The division of communication protocols into different functional layers may mean that businesses will be unable to extract monopoly rents from the mere possession of transport facilities. Layers allow for open entry by those who do not own transmission facilities into markets potentially as vast as the number of connected computers.

Regulatory policy might work to allow this possibility, or it might suppress it. The ultimate potential of the Internet for allowing competition in all services depends in a great measure on the terms upon which people can gain access to it. Anyone with appropriate equipment is potentially an ISP. The computer revolution is constantly reducing the cost and increasing the power of computers, so it is quite conceivable that the number of servers attaching to the Internet will increase dramatically. Regulatory policy has a choice before it, whether to assist this process, and increase the number of ISPs and non-commercial entities attaching to the Internet, or to allow competition from incumbent carriers to reduce the number of ISPs to match number of providers of facilities in a given region. In that latter vision, only owners of transmission facilities would be able to remain as Internet service providers.

2.2.9 Bandwidth x Computations: The Importance of Technological Change

The cost of computations and bandwidth are dropping like a stone Moore's Law ensures that computations continue to decrease in cost, and the rate of increase of bandwidth and comparable decrease in cost is now overtaking even the astounding improvements in computations. Things are getting cheaper, very, very fast.

Since the Internet is a means for computers to communicate with one another, and they are not bound by the physical limitations of human beings, who have other things todothan talk on the telephone, it follows that Internet traffic growth will not correspond to voice traffic growth,

A couple of illustrations may be helpful.

Graph 1: Drop in Price of Transatlantic Circuits

Rise in Capacity, 1956-1993
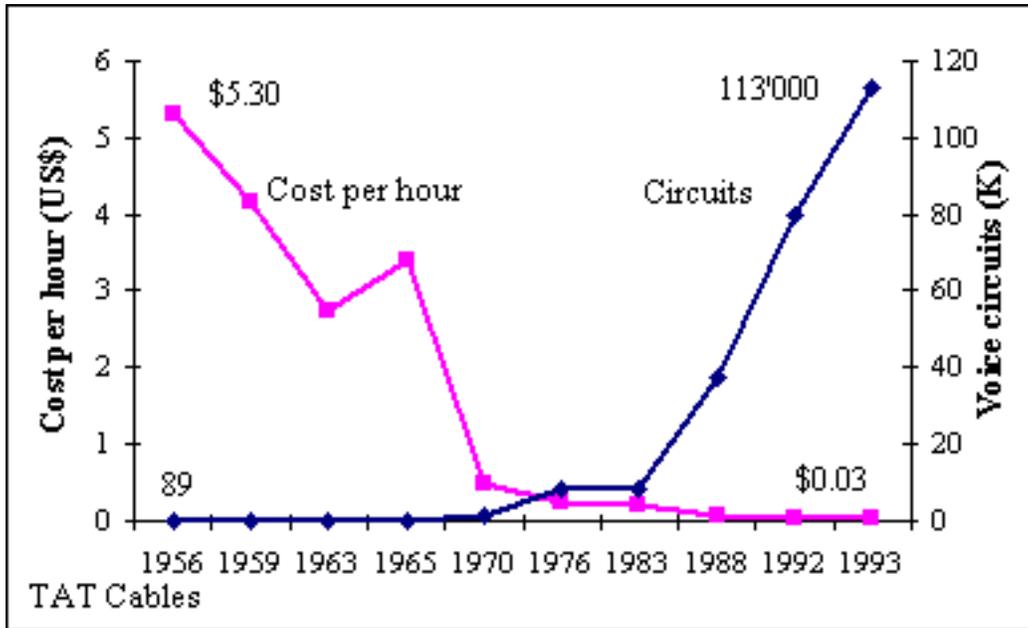
Source: ITU World Telecommunication Development Report 1995
http://www.itu.int/ti/wtdr95/graphics/ov6.gif

Graph 2: Drop in Price of Computer Power

Measured in Millions of Instructions Per Second (MIPS)

Figure 3. Microprocessor Prices Plummet from $230 per MIPS to $3.42 per MIPS in 6 Years
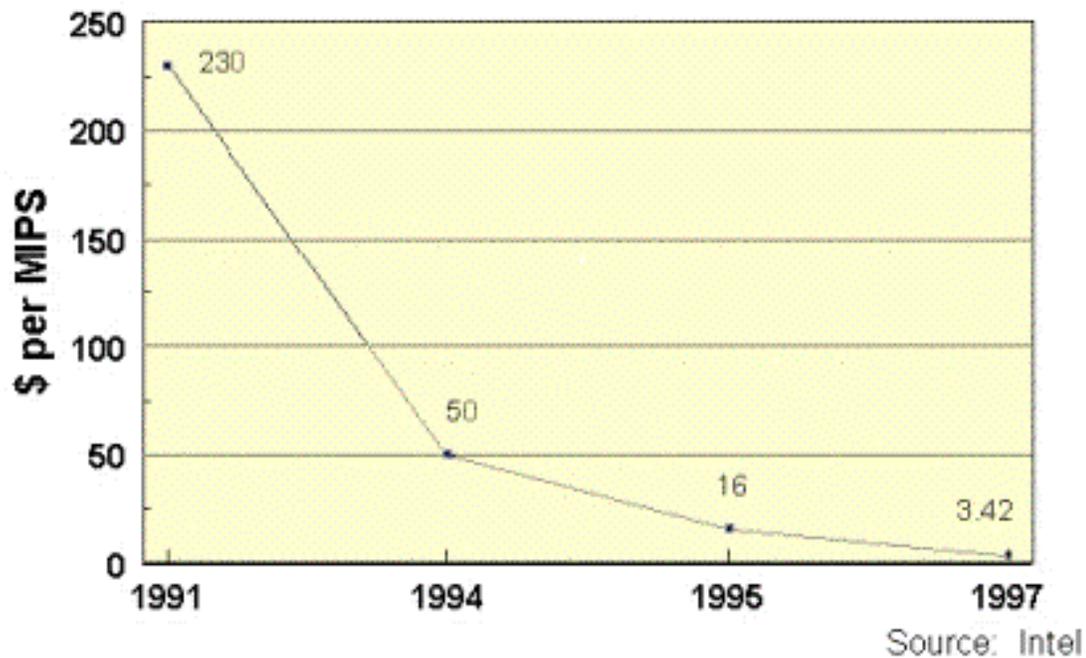
Source: Intel at <http://developer.intel.com/solutions/archive/issue2/focus.htm>

Thus the Internet is designed to take advantage of how the technology is going: faster and cheaper computation, vastly more plentiful bandwidth. This in turn allows entire nationwide telecommunications companies to be built from scratch for ten billion, as opposed to a hundred billion, dollars.

The growth of available bandwidth tells an important story. Generally, increases in voice telecommunications traffic are generally consistent with the growth of the population and economy, in a developed economy. Rates of 5-10% a year are normal. Voice traffic is currently growing around 8% per year. With the Internet, rates of growth are of a different order of magnitude. John Sidgmore, President of UUNet, one of the largest and oldest providers of Internet service, says that his company must double its bandwidth every 3 ½ months to stay abreast of the explosive demand that computer traffic is generating. Doubling every three and a half months is a tenfold increase per year. This is far faster than the operation of Moore's Law on computing power.

The effects will not take long to be seen. No matter how measured, the Internet is growing like crazy. The point of acceleration of Internet growth starts from 1994, from the invention of the World Wide Web, which permitted the transfer of graphics and text on the same file. By the year 2000, the Internet will be half of all bandwith used in the world. By the year 2003, if that rate of growth continues, that figure will be more than 90% and by 2008 more than 99%. "In a way, we won't even know that voice is in there. It will become completely irrelevant", says Sidgmore. Provisioning a network growing tenfold a

year is an enormous challenge. Sidgmore says that his engineers tell him that, "if you are not scared by this, you just don't understand."

A somewhat smaller rate of change was predicted by John MacDonald, Chief Operating Officer of Bell Canada. Bell Canada is the largest telephone company in Canada. Speaking at the Net98/BCIA Conference at Whistler BC, Mr. MacDonald said that, by his calculations

- Internet traffic is increasing at 10% per month,
- By 2005, voice traffic will be less than 20% of network traffic; and
- Bell Canada will migrate or evolve into a "network-centric application and platform developer".

From telephone company to "network-centric application and platform developer" in six years! Telephone companies are faced with a technological revolution, one that they did not start and one which they may be unable to control.

The fate of the telephone company's traditional line of business –switching telephone calls- was represented in the Mr. MacDonald's slide show as the extraction of revenue from the existing plant as its value drops towards zero. Accordingly, it is safe to assume that, while the senior management of telephone companies may not share Mr. Sidgmore's enthusiasm for the Internet, they apprehend it to be a fundamental challenge to their "network-centric" business model.

The point of these illustrations is this: telecommunications are being revolutionized by technologies founded in computers. Successive doublings of the power of computers, and of transmission capacities, are occurring at astonishingly short intervals. It is easy to see that the computer of 2025 will sit in one's hand, have more computing power than all the desktop computers now sitting in Silicon Valley, will resemble a telephone, will run on Internet protocol, and cost $50.

2.3 The World Wide Web

So far this discussion has not mentioned the World Wide Web, the application that has succeeded in making the Internet a household name and a business phenomenon. The omission has been deliberate. All the basic features of the Internet existed before the development of the World Wide Web – the *www* in domain names. But the World Wide Web is a protocol that has greatly added to the power of the Internet – the *killer application* that makes the Internet worth having for millions of people outside university research labs.

The web was originally developed to allow information sharing within internationally dispersed teams, and the dissemination of information by support groups. Tim Berners-Lee, an English physicist working at CERN in Geneva, developed it. It is currently the most advanced information system deployed on the Internet, and embraces within its data model most information in previous networked information systems.

In fact, the web is an architecture which will also embrace any future advances in technology, including new networks, protocols, object types and data formats.

The WWW world consists of documents, and links. Indexes are special documents which, rather than being read, may be searched. The result of such a search is another ("virtual") document containing links to the documents found. A simple protocol ("HTTP", or hypertext transfer protocol) is used to allow a

browser program to request a keyword search by a remote information server.

The web contains documents in many formats. Those documents which are hypertext, (real or virtual) contain links to other documents, or places within documents. All documents, whether real, virtual or indexes, look similar to the reader and are contained within the same addressing scheme.

To follow a link, a reader clicks with a mouse (or types in a number if he or she has no mouse). To search and index, a reader gives keywords (or other search criteria). These are the only operations necessary to access the entire world of data.

The WWW model gets over the frustrating incompatibilities of data format between suppliers and readers by allowing negotiation of format between a smart browser and a smart server.

The development of the World Wide Web, starting about 1989, some fifteen years after the basic protocols of the Internet were devised, illustrates as nothing else could the points made earlier about the fundamental openness of the Internet to new services. The entire hypertext system for locating information from arbitrary nodes via browsers was developed independently of the TCP/IP protocols. The reason most of us we know of the existence of the Internet is through the Web, and yet the two have nothing in common save that one rides upon the other.

Table 1:

Growth in Number of Internet hosts

Date	Hosts
Aug 81	213
May 82	235
Aug 83	562
Oct 84	1,024
Oct 85	1,961
Nov 86	5,089
Dec 87	28,174
Oct 88	56,000
Oct 89	159,000
Oct 90	313,000
Oct 91	617,000

Oct 92	1,136,000
Oct 93	2,056,000
Oct 94	3,864,000
Jan 96	9,472,000
Jan 97	16,146,000
Jan 98	29,670,000
July 98	36,739,000
Jan 99	43,230,000

Source: Network Wizards <http://nw.com

Graph 3

Internet Domain Survey Host Count