Brief History of the Internet

Barry M. Leiner, Vinton G. Cerf, David D. Clark, Robert E. Kahn, Leonard Kleinrock, Daniel C. Lynch, Jon Postel, Larry G. Roberts, Stephen Wolff

Origins of the Internet
The Initial Internetting
Concepts Proving the Ideas
Transition to Widespread Infrastructure
The Role of Documentation
Formation of the Broad Community
Commercialization of the Technology
History of the Future
Footnotes
Timeline
References
Authors

Introduction

The Internet has revolutionized the computer and communications world like nothing before. The invention of the telegraph, telephone, radio, and computer set the stage for this unprecedented integration of capabilities. The Internet is at once a world-wide broadcasting capability, a mechanism for information dissemination, and a medium for collaboration and interaction between individuals and their computers without regard for geographic location. The Internet represents one of the most successful examples of the benefits of sustained investment and commitment to research and development of information infrastructure. Beginning with the early research in packet switching, the government, industry and academia have been partners in evolving and deploying this exciting new technology. Today, terms like "bleiner@computer.org" and "http://www.acm.org" trip lightly off the tongue of the random person on the street. [1]

This is intended to be a brief, necessarily cursory and incomplete history. Much material currently exists about the Internet, covering history, technology, and usage. A trip to almost any bookstore will find shelves of material written about the Internet. [2]

In this paper,[3] several of us involved in the development and evolution of the Internet share our views of its origins and history. This history revolves around four distinct aspects. There is the technological evolution that began with early research on packet switching and the ARPANET (and related technologies), and where current research continues to expand the horizons of the infrastructure along several dimensions, such as scale, performance, and higher-level functionality. There is the operations and management aspect of a global and complex operational infrastructure. There is the social aspect, which resulted in a broad community of Internauts working together to create and evolve the technology. And there is the commercialization aspect, resulting in an extremely effective transition of research results into a broadly deployed and available information infrastructure.

The Internet today is a widespread information infrastructure, the initial prototype of what is often called the National (or Global or Galactic) Information Infrastructure. Its history is complex and involves many aspects - technological, organizational, and community. And its influence reaches not only to the technical fields of computer communications but throughout society as we move toward increasing use of online tools to accomplish electronic commerce, information acquisition, and community operations.

Origins of the Internet

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,[4] 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 the TX-2 computer in Mass. to the Q-32 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. Kleinrock's conviction of the need for packet switching was confirmed.

In late 1966 Roberts went to DARPA to develop the computer network concept and quickly put together his plan for the "ARPANET", publishing it in 1967. At the conference where he presented the paper, there was also a paper on a packet network concept from the UK by Donald Davies and Roger Scantlebury of NPL. Scantlebury told Roberts about the NPL work as well as that of Paul Baran and others at RAND. The RAND group had written a paper on packet switching networks for secure voice in the military in 1964. It happened that the work at MIT (1961-1967), at RAND (1962-1965), and at NPL (1964-1967) had all proceeded in parallel without any of the researchers knowing about the other work. The

word "packet" was adopted from the work at NPL and the proposed line speed to be used in the ARPANET design was upgraded from 2.4 kbps to 50 kbps. [5]

In August 1968, after Roberts and the DARPA funded community had refined the overall structure and specifications for the ARPANET, an RFQ was released by DARPA for the development of one of the key components, the packet switches called Interface Message Processors (IMP's). The RFQ was won in December 1968 by a group headed by Frank Heart at Bolt Beranek and Newman (BBN). As the BBN team worked on the IMP's with Bob Kahn playing a major role in the overall ARPANET architectural design, the network topology and economics were designed and optimized by Roberts working with Howard Frank and his team at Network Analysis Corporation, and the network measurement system was prepared by Kleinrock's team at UCLA. [6]

Due to Kleinrock's early development of packet switching theory and his focus on analysis, design and measurement, his Network Measurement Center at UCLA was selected to be the first node on the ARPANET. All this came together in September 1969 when BBN installed the first IMP at UCLA and the first host computer was connected. Doug Engelbart's project on "Augmentation of Human Intellect" (which included NLS, an early hypertext system) at Stanford Research Institute (SRI) provided a second node. SRI supported the Network Information Center, led by Elizabeth (Jake) Feinler and including functions such as maintaining tables of host name to address mapping as well as a directory of the RFC's.

One month later, when SRI was connected to the ARPANET, the first host-to-host message was sent from Kleinrock's laboratory to SRI. Two more nodes were added at UC Santa Barbara and University of Utah. These last two nodes incorporated application visualization projects, with Glen Culler and Burton Fried at UCSB investigating methods for display of mathematical functions using storage displays to deal with the problem of refresh over the net, and Robert Taylor and Ivan Sutherland at Utah investigating methods of 3-D representations over the net. Thus, by the end of 1969, four host computers were connected together into the initial ARPANET, and the budding Internet was off the ground. Even at this early stage, it should be noted that the networking research incorporated both work on the underlying network and work on how to utilize the network. This tradition continues to this day.

Computers were added quickly to the ARPANET during the following years, and work proceeded on completing a functionally complete Host-to-Host protocol and other network software. In December 1970 the Network Working Group (NWG) working under S. Crocker finished the initial ARPANET Host-to-Host protocol, called the Network Control Protocol (NCP). As the ARPANET sites completed implementing NCP during the period 1971-1972, the network users finally could begin to develop applications.

In October 1972, Kahn organized a large, very successful demonstration of the ARPANET at the International Computer Communication Conference (ICCC). This was the first public demonstration of this new network technology to the public. It was also in 1972 that the initial "hot" application, electronic mail, was introduced. In March Ray Tomlinson at BBN wrote the basic email message send and read software, motivated by the need of the ARPANET developers for an easy coordination mechanism. In July, Roberts expanded its utility by writing the first email utility program to list, selectively read, file, forward, and respond to messages. From there email took off as the largest network application for over a decade. This was a harbinger of the kind of activity we see on the World Wide Web today, namely, the enormous growth of all kinds of "people-to-people" traffic.

The Initial Internetting Concepts

The original ARPANET grew into the Internet. Internet was based on the idea that there would be multiple independent networks of rather arbitrary design, beginning with the ARPANET as the pioneering packet switching network, but soon to include packet satellite networks, ground-based packet radio networks and other networks. 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 through a meta-level "Internetworking Architecture". 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.

In an open-architecture network, the individual networks may be separately designed and developed and each may have its own unique interface which it may offer to users and/or other providers. including other Internet providers. Each network can be designed in accordance with the specific environment and user requirements of that network. There are generally no constraints on the types of network that can be included or on their geographic scope, although certain pragmatic considerations will dictate what makes sense to offer.

The idea of open-architecture networking was first introduced by Kahn shortly after having arrived at DARPA in 1972. This work was originally part of the packet radio program, but subsequently became a separate program in its own right. At the time, the program was called "Internetting". Key to making the packet radio system work was a reliable end-end protocol that could maintain effective communication in the face of jamming and other radio interference, or withstand intermittent blackout such as caused by being in a tunnel or blocked by the local terrain. Kahn first contemplated developing a protocol local only to the packet radio network, since that would avoid having to deal with the multitude of different operating systems, and continuing to use NCP.

However, NCP did not have the ability to address networks (and machines) further downstream than a destination IMP on the ARPANET and thus some change to NCP would also be required. (The assumption was that the ARPANET was not changeable in this regard). NCP relied on ARPANET to provide end-to-end reliability. If any packets were lost, the protocol (and presumably any applications it supported) would come to a grinding halt. In this model NCP had no end-end host error control, since the ARPANET was to be the only network in existence and it would be so reliable that no error control would be required on the part of the hosts. Thus, 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). While NCP tended to act like a device driver, the new protocol would be more like a communications protocol.

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.

Other key issues that needed to be addressed were:

- Algorithms to prevent lost packets from permanently disabling communications and enabling them to be successfully retransmitted from the source.
- Providing for host-to-host "pipelining" so that multiple packets could be enroute from source to destination at the discretion of the participating hosts, if the intermediate networks allowed it.
- Gateway functions to allow it to forward packets appropriately. This included interpreting IP headers for routing, handling interfaces, breaking packets into smaller pieces if necessary, etc.
- The need for end-end checksums, reassembly of packets from fragments and detection of duplicates, if any.
- The need for global addressing
- Techniques for host-to-host flow control.
- Interfacing with the various operating systems
- There were also other concerns, such as implementation efficiency, internetwork performance, but these were secondary considerations at first.

Kahn began work on a communications-oriented set of operating system principles while at BBN and documented some of his early thoughts in an internal BBN memorandum entitled "Communications Principles for Operating Systems". At this point he realized it would be necessary to learn the implementation details of each operating system to have a chance to embed any new protocols in an efficient way. Thus, in the spring of 1973, after starting the internetting effort, he asked Vint Cerf (then at Stanford) to work with him on the detailed design of the protocol. Cerf had been intimately involved in the original NCP design and development and already had the knowledge about interfacing to existing operating systems. So armed with Kahn's architectural approach to the communications side and with Cerf's NCP experience, they teamed up to spell out the details of what became TCP/IP.

The give and take was highly productive and the first written version[7] of the resulting approach was distributed at a special meeting of the International Network Working Group (INWG) which had been set up at a conference at Sussex University in September 1973. Cerf had been invited to chair this group and used the occasion to hold a meeting of INWG members who were heavily represented at the Sussex Conference.

Some basic approaches emerged from this collaboration between Kahn and Cerf:

- Communication between two processes would logically consist of a very long stream of bytes (they called them octets). The position of any octet in the stream would be used to identify it.
- Flow control would be done by using sliding windows and acknowledgments (acks). The destination could select when to acknowledge and each ack returned would be cumulative for all packets received to that point.
- It was left open as to exactly how the source and destination would agree on the parameters of the windowing to be used. Defaults were used initially.
- Although Ethernet was under development at Xerox PARC at that time, the proliferation of LANs were not envisioned at the time, much less PCs and workstations. The original model was national level networks like ARPANET of which only a relatively small number were expected to exist. Thus a 32 bit IP address was used of which the first 8 bits signified the network and the remaining 24 bits designated the host on that network. This assumption, that 256 networks would be sufficient for the foreseeable future, was clearly in need of reconsideration when LANs began to appear in the late 1970s.

The original Cerf/Kahn paper on the Internet described one protocol, called TCP, which provided all the transport and forwarding services in the Internet. Kahn had intended that the TCP protocol support a range of transport services, from the totally reliable sequenced delivery of data (virtual circuit model) to a datagram service in which the application made direct use of the underlying network service, which might imply occasional lost, corrupted or reordered packets. However, the initial effort to implement TCP resulted in a version that only allowed for virtual circuits. This model worked fine for file transfer and remote login applications, but some of the early work on advanced network applications, in particular packet voice in the 1970s, made clear that in some cases packet losses should not be corrected by TCP, but should be left to the application to deal with. This led to a reorganization of the original TCP into two protocols, the simple IP which provided only for addressing and forwarding of individual packets, and the separate TCP, which was concerned with service features such as flow control and recovery from lost packets. For those applications that did not want the services of TCP, an alternative called the User Datagram Protocol (UDP) was added in order to provide direct access to the basic service of IP.

A major initial motivation for both the ARPANET and the Internet was resource sharing - for example allowing users on the packet radio networks to access the time sharing systems attached to the ARPANET. Connecting the two together was far more economical that duplicating these very expensive computers. However, while file transfer and remote login (Telnet) were very important applications, electronic mail has probably had the most significant impact of the innovations from that era. Email provided a new model of how people could communicate with each other, and changed the nature of collaboration, first in the building of the Internet itself (as is discussed below) and later for much of society.

There were other applications proposed in the early days of the Internet, including packet based voice communication (the precursor of Internet telephony), various models of file and disk sharing, and early "worm" programs that showed the

concept of agents (and, of course, viruses). A key concept of the Internet is that it was not designed for just one application, but as a general infrastructure on which new applications could be conceived, as illustrated later by the emergence of the World Wide Web. It is the general purpose nature of the service provided by TCP and IP that makes this possible.

Proving the Ideas

DARPA let three contracts to Stanford (Cerf), BBN (Ray Tomlinson) and UCL (Peter Kirstein) to implement TCP/IP (it was simply called TCP in the Cerf/Kahn paper but contained both components). The Stanford team, led by Cerf, produced the detailed specification and within about a year there were three independent implementations of TCP that could interoperate.

This was the beginning of long term experimentation and development to evolve and mature the Internet concepts and technology. Beginning with the first three networks (ARPANET, Packet Radio, and Packet Satellite) and their initial research communities, the experimental environment has grown to incorporate essentially every form of network and a very broad-based research and development community. [REK78] With each expansion has come new challenges.

The early implementations of TCP were done for large time sharing systems such as Tenex and TOPS 20. When desktop computers first appeared, it was thought by some that TCP was too big and complex to run on a personal computer. David Clark and his research group at MIT set out to show that a compact and simple implementation of TCP was possible. They produced an implementation, first for the Xerox Alto (the early personal workstation developed at Xerox PARC) and then for the IBM PC. That implementation was fully interoperable with other TCPs, but was tailored to the application suite and performance objectives of the personal computer, and showed that workstations, as well as large time-sharing systems, could be a part of the Internet. In 1976, Kleinrock published the first book on the ARPANET. It included an emphasis on the complexity of protocols and the pitfalls they often introduce. This book was influential in spreading the lore of packet switching networks to a very wide community.

Widespread development of LANS, PCs and workstations in the 1980s allowed the nascent Internet to flourish. Ethernet technology, developed by Bob Metcalfe at Xerox PARC in 1973, is now probably the dominant network technology in the Internet and PCs and workstations the dominant computers. This change from having a few networks with a modest number of time-shared hosts (the original ARPANET model) to having many networks has resulted in a number of new concepts and changes to the underlying technology. First, it resulted in the definition of three network classes (A, B, and C) to accommodate the range of networks. Class A represented large national scale networks (small number of networks with large numbers of hosts); Class B represented regional scale networks; and Class C represented local area networks (large number of networks with relatively few hosts).

A major shift occurred as a result of the increase in scale of the Internet and its associated management issues. To make it easy for people to use the network, hosts were assigned names, so that it was not necessary to remember the numeric addresses. Originally, there were a fairly limited number of hosts, so it was feasible to maintain a single table of all the hosts and their associated names and addresses. The shift to having a large number of independently managed networks (e.g., LANs) meant that having a single table of hosts was no longer feasible, and the Domain Name System (DNS) was invented by Paul Mockapetris of USC/ISI. The DNS permitted a scalable distributed mechanism for resolving hierarchical host names (e.g. www.acm.org) into an Internet address.

The increase in the size of the Internet also challenged the capabilities of the routers. Originally, there was a single distributed algorithm for routing that was implemented uniformly by all the routers in the Internet. As the number of networks in the Internet exploded, this initial design could not expand as necessary, so it was replaced by a hierarchical model of routing, with an Interior Gateway Protocol (IGP) used inside each region of the Internet, and an Exterior Gateway Protocol (EGP) used to tie the regions together. This design permitted different regions to use a different IGP, so that different requirements for cost, rapid reconfiguration, robustness and scale could be accommodated. Not only the routing algorithm, but the size of the addressing tables, stressed the capacity of the routers. New approaches for address aggregation, in particular classless inter-domain routing (CIDR), have recently been introduced to control the size of router tables.

As the Internet evolved, one of the major challenges was how to propagate the changes to the software, particularly the host software. DARPA supported UC Berkeley to investigate modifications to the Unix operating system, including incorporating TCP/IP developed at BBN. Although Berkeley later rewrote the BBN code to more efficiently fit into the Unix system and kernel, the incorporation of TCP/IP into the Unix BSD system releases proved to be a critical element in dispersion of the protocols to the research community. Much of the CS research community began to use Unix BSD for their day-to-day computing environment. Looking back, the strategy of incorporating Internet protocols into a supported operating system for the research community was one of the key elements in the successful widespread adoption of the Internet.

One of the more interesting challenges was the transition of the ARPANET host protocol from NCP to TCP/IP as of January 1, 1983. This was a "flag-day" style transition, requiring all hosts to convert simultaneously or be left having to communicate via rather ad-hoc mechanisms. This transition was carefully planned within the community over several years before it actually took place and went surprisingly smoothly (but resulted in a distribution of buttons saying "I survived the TCP/IP transition").

TCP/IP was adopted as a defense standard three years earlier in 1980. This enabled defense to begin sharing in the DARPA Internet technology base and led directly to the eventual partitioning of the military and non- military communities. By 1983, ARPANET was being used by a significant number of defense R&D and operational organizations. The transition of ARPANET from NCP to TCP/IP permitted it to be split into a MILNET supporting operational requirements and an ARPANET supporting research needs.

Thus, by 1985, Internet was already well established as a technology supporting a broad community of researchers and developers, and was beginning to be used by other communities for daily computer communications. Electronic mail was being used broadly across several communities, often with different systems, but interconnection between different mail systems was demonstrating the utility of broad based electronic communications between people.

Transition to Widespread Infrastructure

At the same time that the Internet technology was being experimentally validated and widely used amongst a subset of

computer science researchers, other networks and networking technologies were being pursued. The usefulness of computer networking - especially electronic mail - demonstrated by DARPA and Department of Defense contractors on the ARPANET was not lost on other communities and disciplines, so that by the mid-1970s computer networks had begun to spring up wherever funding could be found for the purpose. The U.S. Department of Energy (DoE) established MFENet for its researchers in Magnetic Fusion Energy, whereupon DoE's High Energy Physicists responded by building HEPNet. NASA Space Physicists followed with SPAN, and Rick Adrion, David Farber, and Larry Landweber established CSNET for the (academic and industrial) Computer Science community with an initial grant from the U.S. National Science Foundation (NSF). AT&T's free-wheeling dissemination of the UNIX computer operating system spawned USENET, based on UNIX' built-in UUCP communication protocols, and in 1981 Ira Fuchs and Greydon Freeman devised BITNET, which linked academic mainframe computers in an "email as card images" paradigm.

With the exception of BITNET and USENET, these early networks (including ARPANET) were purpose-built - i.e., they were intended for, and largely restricted to, closed communities of scholars; there was hence little pressure for the individual networks to be compatible and, indeed, they largely were not. In addition, alternate technologies were being pursued in the commercial sector, including XNS from Xerox, DECNet, and IBM's SNA.[8] It remained for the British JANET (1984) and U.S. NSFNET (1985) programs to explicitly announce their intent to serve the entire higher education community, regardless of discipline. Indeed, a condition for a U.S. university to receive NSF funding for an Internet connection was that "... the connection must be made available to ALL qualified users on campus."

In 1985, Dennis Jennings came from Ireland to spend a year at NSF leading the NSFNET program. He worked with the community to help NSF make a critical decision - that TCP/IP would be mandatory for the NSFNET program. When Steve Wolff took over the NSFNET program in 1986, he recognized the need for a wide area networking infrastructure to support the general academic and research community, along with the need to develop a strategy for establishing such infrastructure on a basis ultimately independent of direct federal funding. Policies and strategies were adopted (see below) to achieve that end.

NSF also elected to support DARPA's existing Internet organizational infrastructure, hierarchically arranged under the (then) Internet Activities Board (IAB). The public declaration of this choice was the joint authorship by the IAB's Internet Engineering and Architecture Task Forces and by NSF's Network Technical Advisory Group of RFC 985 (Requirements for Internet Gateways), which formally ensured interoperability of DARPA's and NSF's pieces of the Internet.

In addition to the selection of TCP/IP for the NSFNET program, Federal agencies made and implemented several other policy decisions which shaped the Internet of today.

Federal agencies shared the cost of common infrastructure, such as trans-oceanic circuits. They also jointly supported "managed interconnection points" for interagency traffic; the Federal Internet Exchanges (FIX-E and FIX-W) built for this purpose served as models for the Network Access Points and "*IX" facilities that are prominent features of today's Internet architecture.

To coordinate this sharing, the Federal Networking Council[9] was formed. The FNC also cooperated with other international organizations, such as RARE in Europe, through the Coordinating Committee on Intercontinental Research Networking, CCIRN, to coordinate Internet support of the research community worldwide.

This sharing and cooperation between agencies on Internet-related issues had a long history. An unprecedented 1981 agreement between Farber, acting for CSNET and the NSF, and DARPA's Kahn, permitted CSNET traffic to share ARPANET infrastructure on a statistical and no-metered-settlements basis.

Subsequently, in a similar mode, the NSF encouraged its regional (initially academic) networks of the NSFNET to seek commercial, non-academic customers, expand their facilities to serve them, and exploit the resulting economies of scale to lower subscription costs for all.

On the NSFNET Backbone - the national-scale segment of the NSFNET - NSF enforced an "Acceptable Use Policy" (AUP) which prohibited Backbone usage for purposes "not in support of Research and Education." The predictable (and intended) result of encouraging commercial network traffic at the local and regional level, while denying its access to national-scale transport, was to stimulate the emergence and/or growth of "private", competitive, long-haul networks such as PSI, UUNET, ANS CO+RE, and (later) others. This process of privately-financed augmentation for commercial uses was thrashed out starting in 1988 in a series of NSF-initiated conferences at Harvard's Kennedy School of Government on "The Commercialization and Privatization of the Internet" - and on the "com-priv" list on the net itself.

In 1988, a National Research Council committee, chaired by Kleinrock and with Kahn and Clark as members, produced a report commissioned by NSF titled "Towards a National Research Network". This report was influential on then Senator Al Gore, and ushered in high speed networks that laid the networking foundation for the future information superhighway.

In 1994, a National Research Council report, again chaired by Kleinrock (and with Kahn and Clark as members again), Entitled "Realizing The Information Future: The Internet and Beyond" was released. This report, commissioned by NSF, was the document in which a blueprint for the evolution of the information superhighway was articulated and which has had a lasting affect on the way to think about its evolution. It anticipated the critical issues of intellectual property rights, ethics, pricing, education, architecture and regulation for the Internet.

NSF's privatization policy culminated in April, 1995, with the defunding of the NSFNET Backbone. The funds thereby recovered were (competitively) redistributed to regional networks to buy national-scale Internet connectivity from the now numerous, private, long-haul networks.

The backbone had made the transition from a network built from routers out of the research community (the "Fuzzball" routers from David Mills) to commercial equipment. In its 8 1/2 year lifetime, the Backbone had grown from six nodes with 56 kbps links to 21 nodes with multiple 45 Mbps links. It had seen the Internet grow to over 50,000 networks on all seven continents and outer space, with approximately 29,000 networks in the United States.

Such was the weight of the NSFNET program's ecumenism and funding ($200 million from 1986 to 1995) - and the quality of the protocols themselves - that by 1990 when the ARPANET itself was finally decommissioned[10], TCP/IP had supplanted or marginalized most other wide-area computer network protocols worldwide, and IP was well on its way to becoming THE bearer service for the Global Information Infrastructure.

The Role of Documentation

A key to the rapid growth of the Internet has been the free and open access to the basic documents, especially the specifications of the protocols.

The beginnings of the ARPANET and the Internet in the university research community promoted the academic tradition of open publication of ideas and results. However, the normal cycle of traditional academic publication was too formal and too slow for the dynamic exchange of ideas essential to creating networks.

In 1969 a key step was taken by S. Crocker (then at UCLA) in establishing the Request for Comments (or RFC) series of notes. These memos were intended to be an informal fast distribution way to share ideas with other network researchers. At first the RFCs were printed on paper and distributed via snail mail. As the File Transfer Protocol (FTP) came into use, the RFCs were prepared as online files and accessed via FTP. Now, of course, the RFCs are easily accessed via the World Wide Web at dozens of sites around the world. SRI, in its role as Network Information Center, maintained the online directories. Jon Postel acted as RFC Editor as well as managing the centralized administration of required protocol number assignments, roles that he continued to play until his death, October 16, 1998.

The effect of the RFCs was to create a positive feedback loop, with ideas or proposals presented in one RFC triggering another RFC with additional ideas, and so on. When some consensus (or a least a consistent set of ideas) had come together a specification document would be prepared. Such a specification would then be used as the base for implementations by the various research teams.

Over time, the RFCs have become more focused on protocol standards (the "official" specifications), though there are still informational RFCs that describe alternate approaches, or provide background information on protocols and engineering issues. The RFCs are now viewed as the "documents of record" in the Internet engineering and standards community.

The open access to the RFCs (for free, if you have any kind of a connection to the Internet) promotes the growth of the Internet because it allows the actual specifications to be used for examples in college classes and by entrepreneurs developing new systems.

Email has been a significant factor in all areas of the Internet, and that is certainly true in the development of protocol specifications, technical standards, and Internet engineering. The very early RFCs often presented a set of ideas developed by the researchers at one location to the rest of the community. After email came into use, the authorship pattern changed - RFCs were presented by joint authors with common view independent of their locations.

The use of specialized email mailing lists has been long used in the development of protocol specifications, and continues to be an important tool. The IETF now has in excess of 75 working groups, each working on a different aspect of Internet engineering. Each of these working groups has a mailing list to discuss one or more draft documents under development. When consensus is reached on a draft document it may be distributed as an RFC.

As the current rapid expansion of the Internet is fueled by the realization of its capability to promote information sharing, we should understand that the network's first role in information sharing was sharing the information about its own design and operation through the RFC documents. This unique method for evolving new capabilities in the network will continue to be critical to future evolution of the Internet.

Formation of the Broad Community

The Internet is as much a collection of communities as a collection of technologies, and its success is largely attributable to both satisfying basic community needs as well as utilizing the community in an effective way to push the infrastructure forward. This community spirit has a long history beginning with the early ARPANET. The early ARPANET researchers worked as a close-knit community to accomplish the initial demonstrations of packet switching technology described earlier. Likewise, the Packet Satellite, Packet Radio and several other DARPA computer science research programs were multi-contractor collaborative activities that heavily used whatever available mechanisms there were to coordinate their efforts, starting with electronic mail and adding file sharing, remote access, and eventually World Wide Web capabilities. Each of these programs formed a working group, starting with the ARPANET Network Working Group. Because of the unique role that ARPANET played as an infrastructure supporting the various research programs, as the Internet started to evolve, the Network Working Group evolved into Internet Working Group.

In the late 1970s, recognizing that the growth of the Internet was accompanied by a growth in the size of the interested research community and therefore an increased need for coordination mechanisms, Vint Cerf, then manager of the Internet Program at DARPA, formed several coordination bodies - an International Cooperation Board (ICB), chaired by Peter Kirstein of UCL, to coordinate activities with some cooperating European countries centered on Packet Satellite research, an Internet Research Group which was an inclusive group providing an environment for general exchange of information, and an Internet Configuration Control Board (ICCB), chaired by Clark. The ICCB was an invitational body to assist Cerf in managing the burgeoning Internet activity.

In 1983, when Barry Leiner took over management of the Internet research program at DARPA, he and Clark recognized that the continuing growth of the Internet community demanded a restructuring of the coordination mechanisms. The ICCB was disbanded and in its place a structure of Task Forces was formed, each focused on a particular area of the technology (e.g. routers, end-to-end protocols, etc.). The Internet Activities Board (IAB) was formed from the chairs of the Task Forces.

It of course was only a coincidence that the chairs of the Task Forces were the same people as the members of the old ICCB, and Dave Clark continued to act as chair. After some changing membership on the IAB, Phill Gross became chair of a revitalized Internet Engineering Task Force (IETF), at the time merely one of the IAB Task Forces. As we saw above, by 1985 there was a tremendous growth in the more practical/engineering side of the Internet. This growth resulted in an explosion in the attendance at the IETF meetings, and Gross was compelled to create substructure to the IETF in the form of working groups.

This growth was complemented by a major expansion in the community. No longer was DARPA the only major player in the funding of the Internet. In addition to NSFNet and the various US and international government-funded activities, interest in the commercial sector was beginning to grow. Also in 1985, both Kahn and Leiner left DARPA and there was a significant decrease in Internet activity at DARPA. As a result, the IAB was left without a primary sponsor and increasingly

assumed the mantle of leadership.

The growth continued, resulting in even further substructure within both the IAB and IETF. The IETF combined Working Groups into Areas, and designated Area Directors. An Internet Engineering Steering Group (IESG) was formed of the Area Directors. The IAB recognized the increasing importance of the IETF, and restructured the standards process to explicitly recognize the IESG as the major review body for standards. The IAB also restructured so that the rest of the Task Forces (other than the IETF) were combined into an Internet Research Task Force (IRTF) chaired by Postel, with the old task forces renamed as research groups.

The growth in the commercial sector brought with it increased concern regarding the standards process itself. Starting in the early 1980's and continuing to this day, the Internet grew beyond its primarily research roots to include both a broad user community and increased commercial activity. Increased attention was paid to making the process open and fair. This coupled with a recognized need for community support of the Internet eventually led to the formation of the Internet Society in 1991, under the auspices of Kahn's Corporation for National Research Initiatives (CNRI) and the leadership of Cerf, then with CNRI.

In 1992, yet another reorganization took place. In 1992, the Internet Activities Board was re-organized and re-named the Internet Architecture Board operating under the auspices of the Internet Society. A more "peer" relationship was defined between the new IAB and IESG, with the IETF and IESG taking a larger responsibility for the approval of standards. Ultimately, a cooperative and mutually supportive relationship was formed between the IAB, IETF, and Internet Society, with the Internet Society taking on as a goal the provision of service and other measures which would facilitate the work of the IETF.

The recent development and widespread deployment of the World Wide Web has brought with it a new community, as many of the people working on the WWW have not thought of themselves as primarily network researchers and developers. A new coordination organization was formed, the World Wide Web Consortium (W3C). Initially led from MIT's Laboratory for Computer Science by Tim Berners-Lee (the inventor of the WWW) and Al Vezza, W3C has taken on the responsibility for evolving the various protocols and standards associated with the Web.

Thus, through the over two decades of Internet activity, we have seen a steady evolution of organizational structures designed to support and facilitate an ever-increasing community working collaboratively on Internet issues.

Commercialization of the Technology

Commercialization of the Internet involved not only the development of competitive, private network services, but also the development of commercial products implementing the Internet technology. In the early 1980s, dozens of vendors were incorporating TCP/IP into their products because they saw buyers for that approach to networking. Unfortunately they lacked both real information about how the technology was supposed to work and how the customers planned on using this approach to networking. Many saw it as a nuisance add-on that had to be glued on to their own proprietary networking solutions: SNA, DECNet, Netware, NetBios. The DoD had mandated the use of TCP/IP in many of its purchases but gave little help to the vendors regarding how to build useful TCP/IP products.

In 1985, recognizing this lack of information availability and appropriate training, Dan Lynch in cooperation with the IAB arranged to hold a three day workshop for ALL vendors to come learn about how TCP/IP worked and what it still could not do well. The speakers came mostly from the DARPA research community who had both developed these protocols and used them in day-to-day work. About 250 vendor personnel came to listen to 50 inventors and experimenters. The results were surprises on both sides: the vendors were amazed to find that the inventors were so open about the way things worked (and what still did not work) and the inventors were pleased to listen to new problems they had not considered, but were being discovered by the vendors in the field. Thus a two-way discussion was formed that has lasted for over a decade.

After two years of conferences, tutorials, design meetings and workshops, a special event was organized that invited those vendors whose products ran TCP/IP well enough to come together in one room for three days to show off how well they all worked together and also ran over the Internet. In September of 1988 the first Interop trade show was born. 50 companies made the cut. 5,000 engineers from potential customer organizations came to see if it all did work as was promised. It did. Why? Because the vendors worked extremely hard to ensure that everyone's products interoperated with all of the other products - even with those of their competitors. The Interop trade show has grown immensely since then and today it is held in 7 locations around the world each year to an audience of over 250,000 people who come to learn which products work with each other in a seamless manner, learn about the latest products, and discuss the latest technology.

In parallel with the commercialization efforts that were highlighted by the Interop activities, the vendors began to attend the IETF meetings that were held 3 or 4 times a year to discuss new ideas for extensions of the TCP/IP protocol suite. Starting with a few hundred attendees mostly from academia and paid for by the government, these meetings now often exceed a thousand attendees, mostly from the vendor community and paid for by the attendees themselves. This self-selected group evolves the TCP/IP suite in a mutually cooperative manner. The reason it is so useful is that it is composed of all stakeholders: researchers, end users and vendors.

Network management provides an example of the interplay between the research and commercial communities. In the beginning of the Internet, the emphasis was on defining and implementing protocols that achieved interoperation.

As the network grew larger, it became clear that the sometime ad hoc procedures used to manage the network would not scale. Manual configuration of tables was replaced by distributed automated algorithms, and better tools were devised to isolate faults. In 1987 it became clear that a protocol was needed that would permit the elements of the network, such as the routers, to be remotely managed in a uniform way. Several protocols for this purpose were proposed, including Simple Network Management Protocol or SNMP (designed, as its name would suggest, for simplicity, and derived from an earlier proposal called SGMP) , HEMS (a more complex design from the research community) and CMIP (from the OSI community). A series of meeting led to the decisions that HEMS would be withdrawn as a candidate for standardization, in order to help resolve the contention, but that work on both SNMP and CMIP would go forward, with the idea that the SNMP could be a more near-term solution and CMIP a longer-term approach. The market could choose the one it found more suitable. SNMP is now used almost universally for network-based management.

In the last few years, we have seen a new phase of commercialization. Originally, commercial efforts mainly comprised

vendors providing the basic networking products, and service providers offering the connectivity and basic Internet services. The Internet has now become almost a "commodity" service, and much of the latest attention has been on the use of this global information infrastructure for support of other commercial services. This has been tremendously accelerated by the widespread and rapid adoption of browsers and the World Wide Web technology, allowing users easy access to information linked throughout the globe. Products are available to facilitate the provisioning of that information and many of the latest developments in technology have been aimed at providing increasingly sophisticated information services on top of the basic Internet data communications.

History of the Future

On October 24, 1995, the FNC 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) 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.

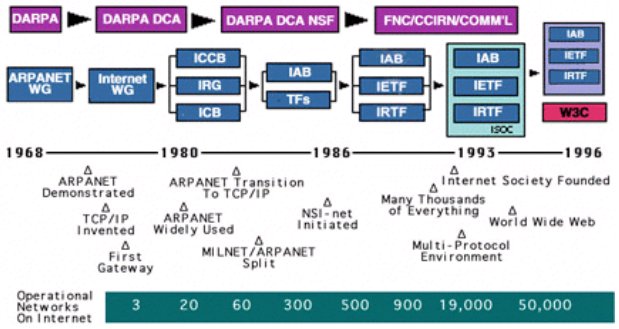
The Internet has changed much in the two decades since it came into existence. It was conceived in the era of time-sharing, but has survived into the era of personal computers, client-server and peer-to-peer computing, and the network computer. It was designed before LANs existed, but has accommodated that new network technology, as well as the more recent ATM and frame switched services. It was envisioned as supporting a range of functions from file sharing and remote login to resource sharing and collaboration, and has spawned electronic mail and more recently the World Wide Web. But most important, it started as the creation of a small band of dedicated researchers, and has grown to be a commercial success with billions of dollars of annual investment.

One should not conclude that the Internet has now finished changing. The Internet, although a network in name and geography, is a creature of the computer, not the traditional network of the telephone or television industry. It will, indeed it must, continue to change and evolve at the speed of the computer industry if it is to remain relevant. It is now changing to provide new services such as real time transport, in order to support, for example, audio and video streams.

The availability of pervasive networking (i.e., the Internet) along with powerful affordable computing and communications in portable form (i.e., laptop computers, two-way pagers, PDAs, cellular phones), is making possible a new paradigm of nomadic computing and communications. This evolution will bring us new applications - Internet telephone and, slightly further out, Internet television. It is evolving to permit more sophisticated forms of pricing and cost recovery, a perhaps painful requirement in this commercial world. It is changing to accommodate yet another generation of underlying network technologies with different characteristics and requirements, e.g. broadband residential access and satellites. New modes of access and new forms of service will spawn new applications, which in turn will drive further evolution of the net itself.

The most pressing question for the future of the Internet is not how the technology will change, but how the process of change and evolution itself will be managed. As this paper describes, the architecture of the Internet has always been driven by a core group of designers, but the form of that group has changed as the number of interested parties has grown. With the success of the Internet has come a proliferation of stakeholders - stakeholders now with an economic as well as an intellectual investment in the network.

We now see, in the debates over control of the domain name space and the form of the next generation IP addresses, a struggle to find the next social structure that will guide the Internet in the future. The form of that structure will be harder to find, given the large number of concerned stakeholders. At the same time, the industry struggles to find the economic rationale for the large investment needed for the future growth, for example to upgrade residential access to a more suitable technology. If the Internet stumbles, it will not be because we lack for technology, vision, or motivation. It will be because we cannot set a direction and march collectively into the future.



Timeline

Footnotes

1 Perhaps this is an exaggeration based on the lead author's residence in Silicon Valley.
2 On a recent trip to a Tokyo bookstore, one of the authors counted 14 English language magazines devoted to the Internet.
3 An abbreviated version of this article appears in the 50th anniversary issue of the CACM, Feb. 97. The authors would like to express their appreciation to Andy Rosenbloom, CACM Senior Editor, for both instigating the writing of this article and his invaluable assistance in editing both this and the abbreviated version.
4 The Advanced Research Projects Agency (ARPA) changed its name to Defense Advanced Research Projects Agency (DARPA) in 1971, then back to ARPA in 1993, and back to DARPA in 1996. We refer throughout to DARPA, the current name.
5 It was from the RAND study that the false rumor started claiming that the ARPANET was somehow related to building a network resistant to nuclear war. This was never true of the ARPANET, only the unrelated RAND study on secure voice

considered nuclear war. However, the later work on Internetting did emphasize robustness and survivability, including the capability to withstand losses of large portions of the underlying networks.
6 Including amongst others Vint Cerf, Steve Crocker, and Jon Postel. Joining them later were David Crocker who was to play an important role in documentation of electronic mail protocols, and Robert Braden, who developed the first NCP and then TCP for IBM mainframes and also was to play a long term role in the ICCB and IAB.
7 This was subsequently published as V. G. Cerf and R. E. Kahn, "A protocol for packet network interconnection" IEEE Trans. Comm. Tech., vol. COM-22, V 5, pp. 627-641, May 1974.
8 The desirability of email interchange, however, led to one of the first "Internet books": !%@:: A Directory of Electronic Mail Addressing and Networks, by Frey and Adams, on email address translation and forwarding.
9 Originally named Federal Research Internet Coordinating Committee, FRICC. The FRICC was originally formed to coordinate U.S. research network activities in support of the international coordination provided by the CCIRN.
10 The decommissioning of the ARPANET was commemorated on its 20th anniversary by a UCLA symposium in 1989.

References

P. Baran, "On Distributed Communications Networks", IEEE Trans. Comm. Systems, March 1964.
V. G. Cerf and R. E. Kahn, "A protocol for packet network interconnection", IEEE Trans. Comm. Tech., vol. COM-22, V 5, pp. 627-641, May 1974.
S. Crocker, RFC001 Host software, Apr-07-1969.
R. Kahn, Communications Principles for Operating Systems. Internal BBN memorandum, Jan. 1972.
Proceedings of the IEEE, Special Issue on Packet Communication Networks, Volume 66, No. 11, November 1978. (Guest editor: Robert Kahn, associate guest editors: Keith Uncapher and Harry van Trees)
L. Kleinrock, "Information Flow in Large Communication Nets", RLE Quarterly Progress Report, July 1961.
L. Kleinrock, Communication Nets: Stochastic Message Flow and Delay, Mcgraw-Hill (New York), 1964.
L. Kleinrock, Queueing Systems: Vol II, Computer Applications, John Wiley and Sons (New York), 1976
J.C.R. Licklider & W. Clark, "On-Line Man Computer Communication", August 1962.
L. Roberts & T. Merrill, "Toward a Cooperative Network of Time-Shared Computers", Fall AFIPS Conf., Oct. 1966.
L. Roberts, "Multiple Computer Networks and Intercomputer Communication", ACM Gatlinburg Conf., October 1967.

Authors

Barry M. Leiner was Director of the Research Institute for Advanced Computer Science. He passed away in April 2003.
Vinton G. Cerf is Vice President and Chief Internet Evangelist at Google.
David D. Clark is Senior Research Scientist at the MIT Laboratory for Computer Science.
Robert E. Kahn is President of the Corporation for National Research Initiatives.
Leonard Kleinrock is Professor of Computer Science at the University of California, Los Angeles, and is Chairman and Founder of Nomadix.
Daniel C. Lynch is a founder of CyberCash Inc. and of the Interop networking trade show and conferences.
Jon Postel served as Director of the Computer Networks Division of the Information Sciences Institute of the University of Southern California until his untimely death October 16, 1998.
Dr. Lawrence G. Roberts is CEO, President, and Chairman of Anagran, Inc.
Stephen Wolff is CTO of Internet2.

Send any comments to Vint Cerf or any of the authors

Copyright © 2012 Internet Society | isoc@isoc.org
1775 Wiehle Avenue, Suite 201, Reston, VA 20190-5108 U.S.A. +1-703-
Galerie Jean-Malbuisson 15, CH-1204 Geneva, Switzerland +41 22 807 1

INTERNET HALL of FAME

We are the world's trusted independent source of leadership for Internet policy, technology standards, and future development. More than simply advancing technology, we work to ensure the Internet continues to grow and evolve as a platform for innovation, economic development, and social progress for people around the world.