

Network Working Group
Request For Comments: 1935
Category: Informational

J. Quarterman
S. Carl-Mitchell
TIC
April 1996

What is the Internet, Anyway?

Status of This Memo

This memo provides information for the Internet community. This memo does not specify an Internet standard of any kind. Distribution of this memo is unlimited.

Copyright (c) 1994 TIC

From Matrix News, 4(8), August 1994
Permission is hereby granted for redistribution of this article provided that it is redistributed in its entirety, including the copyright notice and this notice.
Contact: mids@tic.com, +1-512-451-7602, fax: +1-512-452-0127.
<http://www.tic.com/mids>, <gopher://gopher.tic.com/11/matrix/news>
A shorter version of this article appeared in MicroTimes.

Introduction

We often mention the Internet, and in the press you read about the Internet as the prototype of the Information Highway; as a research tool; as open for business; as not ready for prime time; as a place your children might communicate with (pick one) a. strangers, b. teachers, c. pornographers, d. other children, e. their parents; as bigger than Poland; as smaller than Chicago; as a place to surf; as the biggest hype since Woodstock; as a competitive business tool; as the newest thing since sliced bread.

A recent New York Times article quoting one of us as to the current size of the Internet has particularly stirred up quite a ruckus. The exact figures attributed to John in the article are not the ones we recommended for such use, but the main point of contention is whether the Internet is, as the gist of the article said, smaller than many other estimates have said. Clearly lots of people really want to believe that the Internet is very large. Succeeding discussion has shown that some want to believe that so much that they want to count computers and people that are probably *going to be* connected some time in the future, even if they are not actually connected now. We prefer to talk about who is actually on the Internet and on other networks now. We'll get back to the sizes of the various networks later, but for now let's discuss a more basic issue that is at the

heart of much confusion and contention about sizes: what is the Internet, anyway?

Starting at the Center

For real confusion, start trying to get agreement on what is part of the Internet: NSFNET? CIX? Your company's internal network? Prodigy? FidoNet? The mainframe in accounting? Some people would include all of the above, and perhaps even consider excluding anything politically incorrect. Others have cast doubts on each of the above.

Let's start some place almost everyone would agree is on the Internet. Take RIPE, for example. The acronym stands for European IP Networks. RIPE is a coordinating group for IP networking in Europe. (IP is the Internet protocol, which is the basis of the Internet. IP has a suite of associated protocols, including the Transmission Control Protocol, or TCP, and the name IP, or sometimes TCP/IP, is often used to refer to the whole protocol suite.) RIPE's computers are physically located in Amsterdam. The important feature of RIPE for our purposes is that you can reach RIPE (usually by using its domain, ripe.net) from just about anywhere anyone would agree is on the Internet.

Reach it with what? Well, just about any service anyone would agree is related to the Internet. RIPE has a WWW (World Wide Web) server, a Gopher server, and an anonymous FTP server. So they provide documents and other resources by hypertext, menu browsing, and file retrieval. Their personnel use client programs such as Mosaic and Lynx to access other people's servers, too, so RIPE is a both distributor and a consumer of resources via WWW, Gopher, and FTP. They support TELNET interfaces to some of their services, and of course they can TELNET out and log in remotely anywhere they have personal login accounts or someone else has an anonymous TELNET service such a library catalog available. They also have electronic mail, they run some mailing lists, and some of their people read and post news articles to USENET newsgroups.

WWW, Gopher, FTP, TELNET, mail, lists, and news: that's a pretty characteristic set of major Internet services. There are many more obscure Internet services, but it's pretty safe to say that an organization like RIPE that is reachable with all these services is on the Internet.

Reachable from where? Russia first connected to the Internet in 1992. For a while it was reachable from networks in the Commercial Internet Exchange (CIX) and from various other networks, but not from NSFNET, the U.S. National Science Foundation network. At the time,

some people considered NSFNET so important that they didn't count Russia as reachable because it wasn't accessible through NSFNET. Since there are now several other backbone networks in the U.S. as fast (T3 or 45Mbps) as NSFNET, and routing through NSFNET isn't very restricted anymore, few people would make that distinction anymore. So for the moment let's just say reachable through NSFNET or CIX networks, and get back to services.

Looking at Firewalls

Many companies and other organizations run networks that are deliberately firewalled so that their users can get to servers like those at ripe.net, but nobody outside the company network can get to company hosts. A user of such a network can thus use WWW, Gopher, FTP, and TELNET, but cannot supply resources through these protocols to people outside the company. Since a network that is owned and operated by a company in support of its own operations is called an enterprise network, let's call these networks enterprise IP networks, since they typically use the Internet Protocol (IP) to support these services. Some companies integrate their enterprise IP networks into the Internet without firewalls, but most do use firewalls, and those are the ones that are of interest here, since they're the ones with one-way access to these Internet services. Another name for an enterprise IP network, with or without firewall, is an enterprise Internet.

For purposes of this distinction between suppliers and consumers, it doesn't matter whether the hosts behind the firewall access servers beyond the firewall by direct IP and TCP connections from their own IP addresses, or whether they use proxy application gateways (such as SOCKS) at the firewall. In either case, they can use outside services, but cannot supply them.

So for services such as WWW, Gopher, FTP, and TELNET, we can draw a useful distinction between supplier or distributor computers such as those at ripe.net and consumer computers such as those inside firewalled enterprise IP networks. It might seem more obvious to say producer computers and consumer computers, since those would be more clearly paired terms. However, the information distributed by a supplier computer isn't necessarily produced on that computer or within its parent organization. In fact, most of the information on the bigger FTP archive servers is produced elsewhere. So we choose to say distributors and consumers. Stores and shoppers would work about as well, if you prefer.

Even more useful than discussing computers that actually are suppliers or consumers right now may be a distinction between supplier-capable computers (not firewalled) and consumer-capable

computers (firewalled). This is because a computer that is not supplying information right now may be capable of doing so as soon as someone puts information on it and tells it to supply it. That is, setting up a WWW, Gopher, or FTP server isn't very difficult; much less difficult than getting corporate permission to breach a firewall. Similarly, a computer may not be able to retrieve resources by WWW, Gopher, at the moment, since client programs for those services usually don't come with the computer or its basic software, but almost any computer can be made capable of doing so by adding some software. In both cases, once you've got the basic IP network connection, adding capabilities for specific services is relatively easy.

Let's call the non-firewalled computers the core Internet, and the core plus the consumer-capable computers the consumer Internet. Some people have referred to these two categories as the Backbone Internet and the Internet Web. We find the already existing connotations of "Backbone" and "Web" confusing, so we prefer core Internet and consumer Internet.

It's true that many companies with firewalls have one or two computers carefully placed at the firewall so that they can serve resources. Company employees may be able to place resources on these servers, but they can't serve resources directly from their own computers. It's rather like having to reserve space on a single company delivery truck, instead of owning one yourself. If you're talking about companies, yes, the company is thus fully on the core Internet, yet its users aren't as fully on the Internet as users not behind a firewall.

If you're just interested in computers that can distribute information (maybe you're selling server software), that's a much smaller Internet than if you're interested in all the computers that can retrieve such information for their users (maybe you have information you want to distribute). A few years ago it probably wouldn't have been hard to get agreement that firewalled company networks were a different kind of thing than the Internet itself. Nowadays, firewalls have become so popular that it's hard to find an enterprise IP network that is not firewalled, and the total number of hosts on such consumer-capable networks is probably almost as large as the number on the supplier-capable core of the Internet. So many people now like to include these consumer-capable networks along with the supplier-capable core when discussing the Internet.

Some people claim that you can't measure the number of consumer-capable computers or users through measurements taken on the Internet itself. Perhaps not, but you can get an idea of how many actual consumers there are by simply counting accesses to selected servers

and comparing the results to other known facts about the accessing organizations. And there are other ways to get useful information about consumers on the Internet, including asking them.

Mail, Lists, and News

But what about mail, lists, and news? We carefully left those out of the discussion of firewalls, because almost all the firewalled networks do let these communications services in and out, so there's little useful distinction between firewalled and non-firewalled networks on the basis of these services. That's because there's a big difference between these communications services and the resource sharing (TELNET, FTP) and resource discovery (Gopher, WWW) services that firewalls usually filter. The communications services are normally batch, asynchronous, or store-and-forward. These characterizations mean more or less the same thing, so pick the one you like best. The point is that when you send mail, you compose a message and queue it for delivery. The actual delivery is a separate process; it may take seconds or hours, but it is done after you finish composing the message, and you normally do not have to wait for the message to be delivered before doing something else. It is not uncommon for a mail system to batch up several messages to go through a single network link or to the same destination and then deliver them all at once. And mail doesn't even necessarily go to its final destination in one hop; repeated storing at an intermediate destination followed by forwarding to another computer is common; thus the term store-and-forward. Mailing lists are built on top of the same delivery mechanisms as regular electronic mail. USENET news uses somewhat different delivery mechanisms, but ones that are also typically batch, asynchronous, and store-and-forward. Because it is delivered in this manner, a mail message or a news article is much less likely to be a security problem than a TELNET, FTP, Gopher, or WWW connection. This is why firewalls usually pass mail, lists, and news in both directions, but usually stop incoming connections of those interactive protocols.

Because WWW, Gopher, TELNET, and FTP are basically interactive, you need IP or something like it to support them. Because mail, lists, and news are asynchronous, you can support them with protocols that are not interactive, such as UUCP and FidoNet. In fact, there are whole networks that do just that, called UUCP and FidoNet, among others. These networks carry mail and news, but are not capable of supporting TELNET, FTP, Gopher, or WWW. We don't consider them part of the Internet, since they lack the most distinctive and characteristic services of the Internet.

Some people argue that networks such as FidoNet and UUCP should also be counted as being part of the Internet, since electronic mail is

the most-used service even on the core, supplier-capable Internet. They further argue that the biggest benefit of the Internet is the community of discussion it supports, and mail is enough to join that. Well, if mail is enough to be on the Internet, why is the Internet drawing such attention from press and new users alike? Mail has been around for quite a while (1972 or 1973), but that's not what has made such an impression on the public. What has is the interactive services, and interfaces to them such as Mosaic. Asynchronous networks such as FidoNet and UUCP don't support those interactive services, and are thus not part of the Internet. Besides, if being part of a community of discussion was enough, we would have to also include anyone with a fax machine or a telephone. Recent events have demonstrated that all readers of the New York Times would also have to be included. With edges so vague, what would be the point in calling anything the Internet? We choose to stick with a definition of the Internet as requiring the interactive services.

Some people argue that anything that uses RFC-822 mail is therefore using Internet mail and must be part of the Internet. We find this about as plausible as arguing that anybody who flies in a Boeing 737 is using American equipment and is thus within the United States. Besides, there are plenty of systems out there that use mail but not RFC-822.

So what to call systems that can exchange mail, but aren't on the Internet? We say they are part of the Matrix, which is all computer systems worldwide that can exchange electronic mail. This term is borrowed (with permission) from Bill Gibson, the science fiction writer.

Other people refer to the Matrix as global E-mail. That's accurate, but is a description, rather than a name. Some even call it the e-mail Internet. We find that term misleading, since if a system can only exchange mail, we don't consider it part of the Internet. Not to mention not everything in the world defines itself in terms of the Internet, or communicates through the Internet. FidoNet and WWIVnet, for example, have gateways between themselves that have nothing to do with the Internet. Referring to the Matrix as the Internet is rather like referring to the United Kingdom as England. You may call it convenient shorthand; the Scots may disagree.

What about news? Well, the set of all systems that exchange news already has a name: USENET. USENET is presumably a subset of the Matrix, since it's hard to imagine a USENET node without mail, even though USENET itself is news, not mail. USENET is clearly not the same thing as the Internet, since many (almost certainly most) Internet nodes do not carry USENET news, and many USENET nodes are on other networks, especially UUCP, FidoNet, and BITNET.

A few years ago it was popular in some corners of the press to attempt to equate USENET and the Internet. They're clearly not the same. News, like mail, is an asynchronous, batch, store-and-forward service. The distinguishing services of the Internet are interactive, not news.

Asynchronous Compared to Dialup

Please note that interactive vs. asynchronous isn't the same thing as direct vs. dialup connections. Dialup IP is still IP and can support all the usual IP services. It's true that for the more bandwidth-intensive services such as WWW, you'll be a lot happier with a *fast* dialup IP connection, but any dialup IP connection can support WWW. Some people call these on-demand IP connections, or part-time IP access. They're typically supported over SLIP, PPP, ISDN, or perhaps even X.25.

It's also true that it's a lot easier to run a useful interactive Internet supplier node if you're at least dialed up most of the time so that consumers can reach your node, but you can run servers that are accessible over any dialup IP connection whenever it's dialed up. It's true that some access providers handle low-end dialup IP connections through a rotary of IP addresses, and that's not conducive to running servers, since it's difficult for users to know how to reach them. But given a dedicated IP address, how long you stay dialed up is a matter of degree more than of quality. A IP connection that's up the great majority of the time is often called a dedicated connection regardless of whether it's established by dialing a modem or starting software over a hardwired link.

It's possible to run UUCP over a dedicated IP connection, but it's still UUCP, and still does not support interactive services.

Some people object to excluding the asynchronous networks from a definition of the Internet just because they don't support the interactive services. The argument they make is that FTP, Gopher, and WWW can be accessed through mail. This is true, but it's hardly the same, and hardly interactive in the same sense as using FTP, Gopher, or WWW over an IP connection. It's rather like saying a mail-order catalog is the same as going to the store and buying an item on the spot. Besides, we've yet to see anyone log in remotely by mail.

Is IP Characteristic?

We further choose to define the Internet as being those networks that use IP to permit users to use both the communication services and at least TELNET and FTP among the interactive services we have listed.

This requirement for IP has been questioned by some on the basis that there are now application gateways for other protocol suites such as Novell Netware that permit use of such services. This kind of application gateway is actually nothing new, and is not yet widespread. We choose to think of such networks, at least for the moment, as yet another layer of the onion, outside the core and consumer layers of the Internet.

Others have objected to the use of IP as a defining characteristic of the Internet because they think it's too technical. Actually, we find far fewer people confused about whether a software package or network supports IP than about whether it's part of the Internet or not.

Some people point out that services like WWW, Gopher, FTP, TELNET, etc. could easily be implemented on top of other protocol suites. This is true, and has been done. However, people seem to forget to ask why these services developed on top of IP in the first place. There seems to be something about IP and the Internet that is especially conducive to the development of new protocols. We make no apologies about naming IP, because we think it is important.

There is also the question of IP to where? If you have a UNIX shell login account on a computer run by an Internet access provider, and that system has IP access to the rest of the Internet, then you are an Internet user. However, you will not be able to use the full graphical capabilities of protocols such as WWW, because the provider's system cannot display on a bitmapped screen for you. For that, you need IP to your own computer with a bitmapped screen. These are two different degrees of Internet connectivity that are important to both end users and marketers. Some people refer to them as text-only interactive access and graphical interactive access. Some people have gone so far as to say you have to have graphical capabilities to have a full service Internet connection. That may or may not be so, but in the interests of keeping the major categories to a minimum, we are simply going to note these degrees and say no more about them in this article. However, we agree that the distinction of graphical access is becoming more important with the spread of WWW and Mosaic.

Conferencing Systems and Commercial Mail Systems

Conferencing systems such as Prodigy and CompuServe that support mail and often something like news, plus database and services. But most of them do not support the characteristic interactive services that we have listed. The few that do (Delphi and AOL), we simply count as part of the Internet. The others, we count as part of the Matrix, since they all exchange mail.

We find that users of conferencing systems have no particular difficulty in distinguishing between the conferencing system they use and the Internet. CompuServe users, for example, refer to "Internet mail", which is correct, since the only off-system mail CompuServe supports is to the Internet, but they do not in general refer to CompuServe as part of the Internet.

Similarly, users of the various commercial electronic mail networks, such as MCI Mail and Sprint-Mail, seem to have no difficulty in distinguishing between the mail network they use and the Internet. Since they all seem to have their own addressing syntax, this is hardly surprising. We count these commercial mail networks as part of the Matrix, but not part of the Internet. Many of them have IP links to the Internet, but they don't let their users use them, instead limiting the services they carry to just mail.

Russian Dolls

So let's think of a series of nested Chinese boxes or Russian dolls; the kind where inside Boris Yeltsin is Mikhail Gorbachov, inside Gorbachov is Brezhnev, then Kruschev, Stalin, Lenin, and maybe even Tsar Nicholas II. Let's not talk about that many concentric layers, though, rather just three: the Matrix on the outside, the consumer Internet inside, and the core Internet inside that.

	the core Internet	the consumer Internet	the Matrix
interactive services	supplier- capable	consumer- capable	by mail
	stores and shoppers	shoppers	mail order
asynchronous	yes	yes	yes services

Some people have argued that these categories are bad because they are not mutually exclusive. Well, we observe that in real life networks have differing degrees of services, and the ones of most interest share the least common denominator of electronic mail. Thus concentric categories are needed to describe the real world. You can, however, extract three mutually-exclusive categories by referring to the core Internet, the interactive consumer-only part of the Internet, and to asynchronous systems.

Other people have argued that these categories are not sequential. They look sequential to us, since if you start with the core Internet and move out, you subtract services, and if you start at the outside

of the Matrix and move in, you add services.

Outside the Matrix

In addition to computers and networks that fit these classifications, there are also LANs, mainframes, and BBSes that don't exchange any services with other networks or computers; not even mail. These systems are outside the Matrix. For example, many companies have an AppleTalk LAN in marketing, a Novell NetWare LAN in management, and a mainframe in accounting that aren't connected to talk to anything else. In addition, there are a few large networks such as France's Teletel (commonly known as Minitel) that support very large user populations but don't communicate with anything else. These are all currently outside all our Chinese boxes of the core Internet, the consumer Internet, and the Matrix.

DNS and Mail Addresses

There are other interesting network services that make a difference to end users. For example, DNS (Domain Name System) domain names such as tic.com and domain addresses such tic@tic.com can be set up for systems outside the Internet. We used tic.com when we only had a UUCP connection, and few of our correspondents noticed any difference when we added an IP connection (except our mail was faster). This would be more or less a box enclosing the consumer Internet and within the Matrix. But the other three boxes are arguably the most important.

Some people have claimed that anything that uses DNS addresses is part of the Internet. We note that DNS addresses can be used with the UUCP network, which supports no interactive services, and we reject such an equation.

It is interesting to note that over the years various attempts have been made to equate the Internet with something else. Until the mid-1980s lots of people tried to say the Internet was the ARPANET. In the late 1980s many tried to say the Internet was NSFNET. In the early 1990s many tried to say the Internet was USENET. Now many are trying to say the Internet is anything that can exchange mail. We say the Internet is the Internet, not the same as anything else.

Summary

So, here we have a simple set of categories for several of the categories of network access people talk about most these days. Any such categories are at least somewhat a matter of opinion, and other people will propose other categories and other names. We like these categories, because they fit our experience of what real users

actually perceive.

You'll notice we've avoided use of the words "connected" and "reachable" because they mean different things to different people at different times. For either of them to be meaningful, you have to say which services you are talking about. To us, reachable usually means pingable with ICMP ECHO, which is another way to define the core Internet. To others, reachable might mean you can send mail there, which is another way to define the Matrix.

Once we have terms for networks of interest, we can talk about how big those networks are. We think the terms we have defined here refer to groups of computers that people want to use, and that some people want to measure. Many marketers want to know about users. Well, users of mail are in the Matrix, and users of interactive services such as WWW and FTP are in the Internet. Other people are more interested in suppliers or distributors of information. Suppliers of information by mail can be anywhere in the Matrix, but suppliers of information by WWW or FTP are in the core Internet. It is easy to define more and finer degrees of distinctions of capabilities and connectivity, but these three major categories handle the most important cases.

We invite our readers to tell us what distinctions they find important about the various networks and their services.

Security Considerations

Security issues are not discussed in this memo.

Authors' Addresses

John S. Quarterman
Smoot Carl-Mitchell

EMail: tic@tic.com