The Changing Global Internet Service Infrastructure

Michael F. Schwartz and John S. Quarterman

The Internet interconnects over 12,000 networks worldwide and is growing at an exponential rate, as measured by domain registrations and packet counts. Increasingly often, people want to know how fast a particular part of the Internet is growing, to do capacity planning, gauge commercial promise, or simply to understand this important change in our society. Yet, looking only at registrations and packet counts does not uncover the full complexity of the situation. There are a variety of ways that sites can connect to the Internet, each offering different capabilities, costs, technical problems, and security considerations. In this paper we analyze Internet growth based on measurements of which of a dozen common TCP services could be reached at each of over 13,000 sites worldwide, tested four times over the course of 1992. We analyze this data as a function of country, type of institution, and type of service. We also derive mathematical models that can be used to project growth rates for individual countries and the global Internet.

Introduction

The global TCP/IP Internet is growing at a phenomenal rate. In 1982, its predecessor networks (ARPANET and CSNET) connected approximately 150 computers together. Today, the Internet connects over 12,000 networks and 1.5 million computers. Gateways to a variety of electronic messaging services permit Internet users to communicate with at least 15 million educational, commercial, government, military, and other types of users throughout the worldwide Matrix of computer networks that exchange mail or news. During this same time period, the Internet has experienced an increase of four orders of magnitude in the speed and data capacity of the network.

But the Internet is much more than just a far-reaching basis of physical connectivity. A rapidly developing set of networked information discovery and retrieval tools is changing the way people work and interact, by providing users easy access to documents, sounds, images, and other file system data; library catalog and user directory data; weather, geography, telemetry, and other physical science data; and many other types of information (Schwartz et al., 1992).

The rapid growth of the Internet, combined with the creation of new types of applications, promise potentially sweeping changes to the society. In turn, many people have become interested in projecting and understanding Internet growth rates. The commercial sector would like to know what market potential the Internet offers, and when to enter. The research and educational sector is interested both from the pragmatic perspective of when it will offer a true global village of schools, and from the more theoretical perspective of studying an important and rapidly moving force in the society. Governments want to know how to plan for and manage this burgeoning new global infrastructure. Engineers and developers want to know what to expect as they create new applications and network protocols.

A number of studies have measured Internet growth based on counts of registered sites, naming tables, or network provider-collected traffic statistics. Yet, because of the variety of ways that sites can choose to connect to the Internet and the range of different ways that sites use the Internet, these measures are difficult to interpret. For example, while corporate sites often have thousands of hosts registered in the Domain Naming System (Mockapetris, 1987), they typically erect barriers around their networks that limit network traffic (and hence the realm of potential applications and collaboration tools) in a variety of ways.

Because we believe the most far-reaching impact of the Internet will come from novel types of network services, in this paper we analyze growth based on measurements of changes in reachable Internet services. Our analysis takes into account three different phenomena: the rate at which sites initiate some type of network connectivity (perhaps registering a domain name and arranging for mail to be forwarded through periodic dialup connections); the rate at which such sites become directly connected to the Internet, buying into the richer functionality that provides; and the rate at which sites back away from full Internet connectivity, usually for security reasons (Carl-Mitchell & Quarterman, 1992).

The final category (distancing from the Internet) was our initial motivation for performing this study. While the Internet can support many useful types of collaboration, connectivity also opens many avenues for security violations, as evidenced by a number of well publicized events over the past few years (National Research Council, 1991; Spafford, 1989; Stoll, 1988). In response, many sites have imposed mechanisms to limit their exposure to security intrusions. While preferable to the damage that could occur from security violations, taken to an extreme these measures could hinder or prevent the deployment of new types of network services, impeding both research and commercial advancement. At the Fall 1990 Interop conference public session on security, David Clark of MIT referred to this possibility as ‘‘The Great Disconnection’’.

Our measurements provide a quantitative basis to discuss this phenomenon, as well as mathematical models of overall network growth. The data indicate growth and distancing rates as a function of geographic location and type of institution (commercial, educational, etc.), as well as the types of services sites are willing to run (and hence the type of networked collaboration they can support).

Related Research

To the best of our knowledge, no previous study has attempted to measure reachable Internet services. However, a number of studies have considered growth rates based on network provider-collected traffic statistics, lists of registered sites, or naming tables.

Network Operation Centers typically measure network traffic characteristics, such as packet counts and protocol usage measurements. Regional networks use these measurements to determine when equipment upgrades are needed. NSFNET makes monthly measurements of NSFNET backbone traffic available via the Internet (NNSC, 1989).

Various researchers have collected information about network traffic usage, typically to parameterize models of routing and other low-level network issues, or to help indicate potential areas of improvement for these issues (Braun et al., 1993; Caceres et al., 1991; Heimlich, 1990; Paxson, 1991).

Claffy has used network measurements to analyze policy issues (1993). The paper presents figures for traffic between every pair of countries on the Internet, and figures for traffic into and out of the U.S.

Lottor has performed the most extensive measurements of changes in the Internet host name space. He recently published the results of a ten year study that counted the number of hosts that have IP addresses (Lottor, 1992a). In the early years he extracted the data from host tables maintained by the U.S. NIC. With the transition to the distributed Domain Naming System, he later collected measurements using his *ZONE* software, which recursively traverses the name tree and retrieves host information using ‘‘zone transfers’’ (Lottor, 1990).

Many of the hosts counted by Lottor's study are hidden behind gateways that selectively forward only mail, hosts that are infrequently connected to the Internet, or hosts that in other ways are not directly and continuously connected to the Internet. Therefore, Lottor's study really indicates the spread of IP and the Domain Naming System at sites connected to the Internet. We believe the current study offers a more meaningful measure of Internet size, because it is through reachable network services that all Internet sites gain the biggest advantages of connectivity.

A variety of other people have used recursive Domain Naming System traversals to monitor network growth. Ganatra's *census* software provides this functionality, and is freely available via the Internet (1992). The European IP research organization (Blokzijl, 1990) uses recursive traversals to monitor growth rates of member countries in Europe, in some cases followed by attempting to ''ping'' (Muuss, 1983) hosts to determine network-level reachability. Schwartz uses recursive traversals as one means of discovering new sites for inclusion in the *seed database* used by the Netfind Internet user directory service (Schwartz & Tsirigotis, 1991a; Schwartz & Pu, 1993).

Rutkowski used Lottor's data as well as data collected by the U.S. Network Information Center (NIC),[1] Merit, and Sprint, to compile a range of statistics about host and network registrations, backbone traffic type and geographical destination, and other types of information (1993).

Landweber maintains a map of Internet- and mail- accessible sites, and provides maps of these sites in the quarterly *Internet Society Newsletter*. Quarterman compiles and maps detailed information about the Internet and other networks (MIDS, 1993).

Schwartz et al. have developed a system called Fremont, which automates the process of discovering various network characteristics and problems (Wood, Coleman & Schwartz, 1993).

For discussions of the size of the set of computer networks interconnected for at least mail or news service, see (MIDS, 1993; Quarterman, 1990). For a measure of the diameter of the interpersonal communication graph enabled by electronic mail, see (Schwartz & Wood, 1993). Finally, we urge anyone considering performing an Internet measurement study to read (Cerf, 1991).

Definitions

Network Name Space

To understand Internet growth, we must first define what we wish to measure. At one time, connectivity via the IP protocol suite defined the Internet. Since a number of protocols now coexist on the Internet, some people have suggested defining the Internet instead by a common name space (perhaps the Domain Naming System or X.500 (CCITT/ISO, 1988)). However, such a definition ignores differences between various types of physical connectivity, which determine what types of application services are available to users. In particular, it does not distinguish the parts of the network that can support interactive applications (like remote login) from parts based on dialup connections that support only mail and news. Given the advantages of interactive connectivity and the growing popularity of IP, in this paper we consider only the interconnected IP Internet. We refer to the larger collection of networks that can exchange mail or news as the Matrix.

Throughout this paper when we refer to ''domains'' we mean either leaf or interior nodes in the Domain naming tree (Mockapetris, 1987). When we refer to ''sites'' we mean organizational groupings inferred by the Domain Naming System. For example, the machine ''abingdon.eng.sun.com'' falls within the site ''eng.sun.com'', which is a different site than ''central.sun.com'', even though both domains belong to a single corporation (Sun Microsystems, Inc.). Hosts are identified as domains having IP addresses registered in the Domain Naming System. Both ‘‘abingdon.eng.sun.com’’ and ‘‘eng.sun.com’’ are domains (or, domain names). We use the term ‘‘institution’’ to refer to a collection of sites related to a single organization (Sun in the above example).

The purpose of distinguishing between sites and institutions is to permit a more fine-grained analysis of the patterns of Internet disconnection and growth. In particular, a number of institutions allow direct Internet access to some of their sites, while restricting such access to other sites (e.g., allowing Internet access for a research branch of a company, while restricting such access for a product development branch). Our measurements reflect this level of detail.

Styles of Connection and Distancing

There are many ways that sites can reduce their closeness of association with the Internet. We review them here, as a preface to our definition of Internet connectivity.

The most extreme measure is to avoid Internet connectivity. This approach is taken by particularly security-conscious sites, such as certain military sites. Because of the tremendous advantages of Internet access, however, many sites prefer less extreme measures.

An increasingly popular measure is the use of a ‘‘firewall’’ gateway, which allows only certain types of traffic into a site, such as electronic mail and news (Carl-Mitchell & Quarterman, 1992). In some cases the gateway allows connections directly to internal machines for selected services (e.g., allowing mail but disallowing remote logins). In other cases, the gateway accepts incoming mail and then retransmits it along a separate internal connection, for added security. Many sites or hosts use filters that prohibit accessing particular services from selected domains or networks (Carl-Mitchell & Quarterman, 1993). Some sites also use access control lists, to allow only connections initiated by particular sites or people.

A less restrictive arrangement is to permit all types of network traffic into a site, but not run certain network services within the site. For example, many sites support a full spectrum of remote login, mail, and file transfer services, but disable the ‘‘finger’’ service (Zimmerman, 1990), because of security problems that were present in older versions of that software. As another possible arrangement, some sites permit network traffic only if it was initiated from within that site (i.e., they do not provide services accessible to outside network users).

Finally, some sites support only periodic dialup access to the Matrix, through networks like UUCP[2] and FidoNet (Quarterman, 1990). These restrictions typically arise for economic reasons, rather than because of security concerns. The asynchronous nature of the network connections they support limits sites that use these networks to mail and news service. While it is possible to perform some other services via electronic mail interfaces, this style of access is quite limited. For example, there are mail interfaces to FTP (Postel & Reynolds, 1985), Archie (Emtage & Deutsch, 1992), and WAIS (Kahle & Medlar, 1991), but the interfaces have more limited functionality than their interactive TCP-based counterparts, and only a small number of servers can be accessed through mail interfaces. Many other services, such as Gopher (McCahill, 1992) and Netfind (Schwartz & Tsirigotis, 1991a), cannot be accessed by mail interfaces. It is for these reasons that we consider the distinction important between the Internet and the rest of the Matrix.

The Service-Reachable Internet

As will be discussed in the next section, we determined Internet connectivity through a series of connection attempts spread over a set of 1-2 day periods (which we term measurement *cycles*). Because of this approach, we define Internet connectivity as a site's being reachable via any of the tested services at some time during the measurement cycle. This definition does not include certain types of Internet connectivity, such as periodic SLIP (Romkey, l988) connections, and sites that were unreachable throughout a particular measurement cycle (e.g., because of external gateway problems). Practically speaking, this definition is reasonable: if a site is not reachable for such a long period of time, most of its network services will not be usable. Most network services (such as telnet and FTP) attempt retransmissions for only a few minutes. While electronic mail and news use longer timeouts, connectivity that only permits only these services is essentially the same as that provided by dialup UUCP links. We do not consider hosts connected by such means to be on the Internet.

Our Internet definition also excludes sites that only allow traffic to be initiated locally. While such sites are legitimate members of the Internet community, they do not contribute to the network service infrastructure. Note also that measuring such client-only sites would require monitoring traffic to detect their existence, which we did not perform for this study.

Experimental Methodology

The data collection phase of the study consisted of a set of runs of a program over the span of a one to two day measurement cycle, repeated four times over the course of 1992. Each program run attempted to connect to 13 different TCP services at each of 13,749 Internet sites worldwide (indicated in Table 1)[3], recording the failure/success status of each attempt. The program attempted no data transfers in either direction. If a connection was successful, it was simply closed and counted.[4] The machines on which connections were attempted were selected at random from a large list of machines in the Internet, constrained such that at most 3 machines were contacted in any particular site. We generated the site list from a broad variety of sources. In total, the list contained 90,888 hosts, gathered from USENET (Quarterman, 1990) news headers, FTP and mail access logs collected at the University of Colorado, the U.S. Network Information Center-maintained Internet host table (Feinler et al., 1982), and the output of Lottor's recursive Domain Naming System traversals. The breadth of these information sources is important, as it helps ensure that a wide cross-section of sites were tested, without perturbing our results because of limitations of particular site lists.

Individual connection attempts were timed-out after 20 seconds. If a site experienced 3 timeouts on any port (on 3 different machines), the measurement software gave up trying that site for the duration of that measurement run. A measurement cycle consisted of several such program runs, executed successively until the service reachability counts between two runs differed by no more than 1%. This required between three and six runs, each spanning 4-30 hours (longer in earlier runs, when more sites were tested). Therefore, during each measurement cycle, each site was given a total of three 20-second timeouts per set of connection attempts, and several sets of attempts over the period of a day or two.

The services to which connections were attempted are indicated in Table 2. This list was chosen to span a representative range of service types, each of which can be expected to be found on any machine in a site (so that probing random machines would be meaningful). The one exception is the Domain Naming System, for which the

Top-Level Domain Name	Description	Sub-Domains Tested	Top-Level Domain Name	Description	Sub-Domains Tested
ar	Argentina	7	in	India	3
is	Iceland	27	int	International	1
arpa	Obsolete Names	202	it	Italy	120
at	Austria	94	jp	Japan	638
au	Australia	496	kr	Korea	27
be	Belgium	28	lk	Sri Lanka	1
br	Brazil	16	mil	U.S. Military	230
ca	Canada	561	mx	Mexico	14
ch	Switzerland	93	my	Malaysia	6
cl	Chile	5	na	Namibia	1
cn	China	1	net	Network Administrations	187
co	Columbia	2	ni	Nicaragua	1
com	U.S. Commercial	3,082	nl	The Netherlands	231
cr	Costa Rica	1	no	Norway	171
cs	Former Czechoslovakia	5	nz	New Zealand	76
de	Germany	551	org	Non-profit	625
dk	Denmark	192	ph	Philippines	2
edu	U.S. Educational	3,735	pl	Poland	1
ee	Estonia	1	pr	Puerto Rico	3
es	Spain	144	pt	Portugal	10
fi	Finland	168	se	Sweden	531
fr	France	173	sg	Singapore	6
gb, uk	United Kingdom	642	su	Former Soviet Union	44
gov	U.S. Government	255	th	Thailand	1
gr	Greece	13	tn	Tunisia	1
hk	Hong Kong	4	tw	Taiwan	6
hu	Hungary	1	us	United States	240
ie	Ireland	29	yu	Former Yugoslavia	2
il	Israel	24	za	South Africa	18

Table 1: Domains Tested by Measurement Process

Port Number	Service Name	Function
13	daytime	Time-of-Day
15	netstat	Network Connection Information
21	FTP	File Transfer Protocol
23	telnet	Remote Login Protocol
25	SMTP	Mail Transfer Protocol
53	Domain Naming System	Name Service
79	finger	User Information
111	Sun portmap	Service Name-to-Port Information
513	rlogin	Remote Login (BSD)
514	rsh	Remote Command Interpreter (BSD)
540	UUCP-over-TCP	Data Transfer Protocol
543	klogin	Kerberos-authenticated Remote Login
544	krcmd, kshell	Kerberos-authenticated Remote Command Interpreter

Table 2: Network Services Tested

machines to probe were selected from information obtained from the Domain system itself. Only TCP services were tested, since the TCP connection mechanism provides an application-independent means to determine if a server is running.

It would have been possible to retrieve ‘‘Well Known Service’’ records from the Domain Naming System, as a somewhat less ‘‘invasive’’ measurement approach. However, these records are not required for proper network operation, and often are inaccurate. The only way to collect the data for this study was to measure it using attempted connections.

Another experimental design choice we made was to test a fixed set of sites throughout each study cycle. An alternative approach that we tried during our preliminary study (Schwartz, 1991b) was to incorporate newly discovered/registered sites in the site list between each run, in an attempt to measure connectivity of the growing Internet. We opted away from this approach in the current study because it is difficult to distinguish between the effects of uneven growth of the site list and actual service reachability changes. To differentiate between these two phenomena, one would need to create a complete list of Internet sites for each measurement cycle, so that site unreachability would only reflect disconnections. However, given the decentralized nature of the Internet, amassing a complete list would be quite difficult. By using a large, fixed site list for the study, we could measure the rate at which previously existing sites connect to or disconnect from the Internet.

To help illuminate the meaning of our chosen Internet size metric, Figure 1 plots the set of reachable domains from October 1992 (as squares) and Lottor's host count data from January 1993 (as circles) geographically. For the purposes of comparison, we only included those hosts in Lottor's data that corresponded to domains tested by our study. In the figure, icon sizes are logarithmic: each increase in size represents a factor of ten increase in count. Each icon plots data for sites/hosts located roughly at the center of the icon.

In this figure, the count of reachable sites is necessarily smaller than the count of hosts, since most sites have multiple hosts. However, the wide fluctuation in the order-of-magnitude differences between the two metrics indicates that in some cases host counts markedly over-represent the set of service-reachable sites. In many cases (such as South East Asia), there are many hosts but few service-reachable sites. In some cases (such as Moscow), there are hosts but no service-reachable sites. The plot also provides a rough intuitive feeling for the scope and geographical distribution of the sites tested by of our measurement study. Note that the circles should be used for this purpose, as squares indicate only *reachable* sites, not all tested sites.

As another point of comparison, we found that only about 35% of the sites in Lottor's January 1991 list were reachable by our January 1992 measurement cycle (Schwartz, 1992).

We now turn to more general considerations of our study methodology. A study of this nature and magnitude raises a number of potential concerns. Below we discuss considerations of network and remote site load, mechanisms used to control the data collection process, and our efforts to inform sites measured by this study, along with concomitant network appropriate use and privacy issues.

Network and Remote Site Load

The measurement software was careful to avoid generating unnecessary Internet packets, or congesting the Internet with too much concurrent activity. As described in the methodology section, once the software successfully connected to a particular service at a site, it never attempted to connect to that service on any machine in that site again, for the duration of the measurement cycle. Once it recorded 3 connection refusals at any machines in that site for a service, it did not try that service at that site again during that measurement run. If it experienced 3 timeouts on any machine in a site, it gave up on the site, possibly to be retried again a day later (to overcome transient network problems). Thus, in the worst case there would be 3 connection failures for each service at 3 different machines, which amounts to 37 connection requests per site (3 for each of the 12 services other than the Domain Naming System, and one for the Domain Naming System). However, the average was much less than this.

To quantify the actual Internet load, we now present some measurements from test runs of the measurement software that were performed in August 1991. In total, 50,549 Domain Naming System lookups were performed, and 73,760 connections were attempted. This measurement run completed in

Figure 1: Reachable Domains vs. Host Counts

approximately 10 hours, never initiating more than 20 network operations (name lookups or connection attempts) concurrently. The total NSFNET backbone load from all traffic sources that month was approximately 5 billion packets. Therefore, the traffic from our measurement study amounted to less than .5% of this volume on the day that the measurements were collected. Since the Internet contains several other backbones besides NSFNET, the proportionate increase in total Internet traffic was significantly less than .5%.

The cost to a remote site being measured was effectively zero. From the above measurements, on average we attempted 5.7 connections per remote site. The cost of a connection open/close sequence is quite small, particularly when compared to the cost of the many electronic mail and news transmissions that most sites experience on a given day.

Control Over Data Collection Process

The measurement software evolved from an earlier set of experiments used to measure the reach of the

Netfind Internet user directory service (Schwartz & Tsirigotis, 1991a). We evolved and tested this software extensively over a period of two years prior to starting the current measurement study. During that time it was used in a number of experiments of increasing scale.

The measurement software used several redundant checks and other mechanisms to ensure that careful control was maintained over the network operations that were performed (Schwartz & Tsirigotis, 1991b). In addition, we monitored the progress and network load of the measurements during the measurement runs, observing the log of connection requests in progress as well as physical and transport level network status (which indicate the amount of concurrent network activity in progress). Finally, because the measurements were controlled from a single centralized location, it was easy to stop the measurements at any time that need might arise.

Network Appropriate Use and Privacy

When we performed our initial test runs of this study, we attempted to inform site administrators at each study site about the study, by posting a message on the USENET newsgroup ‘‘alt.security’’, and by sending individual electronic mail messages to site administrators. We also informed the Computer Emergency Response Team of the study.

As a practical matter, informing all sites turned out to be quite difficult. Part of the problem was that no channels exist to allow such information to be easily disseminated. Approximately half of the messages we sent to site administrators were returned by remote mail systems as undeliverable. Moreover, the network traffic and remote site administrative load caused by the study announcement messages far outstripped the network and administrative load required by the study itself. Some sites felt that the announcement was an unnecessary imposition of their time.

In addition to these practical problems, a broad announcement of this study could affect the measurements it attempts to gather. Some sites would likely react to the announcement by changing the reachability of their services. Asking for explicit permission from sites would yield even worse methodological problems, as this would have provided a self-selected study group consisting of sites that are less likely to distance themselves from the Internet.

In contrast with our attempts to announce the study, running the study without announcing it caused only a small number of site administrators to notice the traffic and inquire about it to either the CERT or to one of the network points of contact at the University of Colorado. The remote site administrator and network overhead of announcing the study, coupled with the practical and methodological problems of announcing the study, led us to prefer to run the study without further broad announcements. Yet, to avoid causing alarm at a site detecting our network measurement activity, we felt obligated to announce the study.

To resolve this problem, we discussed the study with the Internet Architecture Board, Internet Engineering Steering Group, National Science Foundation, representatives of several U.S. regional networks, and a number of individuals involved with network security, including the Computer Emergency Response Team, members of the Internet Engineering Task Force Security and Advisory Group, and a member of the Lawrence Livermore National Laboratory Computer Incident Advisory Capability. The first part of our efforts resulted in the production of Internet Request For Comments (RFC) number 1262 (Cerf, 1991). Beyond this, we agreed that the appropriate action was to announce the study well ahead of running it via an Internet Request For Comments document (RFC) (Schwartz, 1991a), augmented by electronic bulletin board postings that briefly describe the study goals and methodology and points to this RFC. Moreover, to help sites that missed these announcements, we set up the measurement software in a fashion intended to minimize the effort a site administrator might expend to determine the nature of the activity after detecting it. In particular, we ran the program from an account called ‘‘testnet’’ on a machine with few other users logged in. ‘‘Fingering’’ (Zimmerman, 1990) this machine indicated the testnet login. ‘‘Fingering’’ or sending mail to the testnet login returned information about the study.

The data collected by the study are somewhat sensitive to privacy and security concerns, in the sense that they might be used as a ‘‘road map’’ of accessible network services. Therefore, we treat the raw data as private information, providing measurements only in global statistical terms, divorced from the actual sites that make up the underlying data points.

Growth in Sites Connecting to the Internet

In this section we present measurements of growth in reachable Internet sites. We will use these measurements later in this paper, when we try to extrapolate growth and disconnection trends in the Internet. The measurements also demonstrate some surprising characteristics in themselves, which we use as a basis for speculating about the nature of global network connectivity patterns.

Figure 2 plots the count of reachable sites over time. Given that the Internet is currently experiencing exponential growth (Merit, Inc., 1992), it was surprising that the curve appears to be leveling off. A likely explanation is that by the final measurement period we had reached nearly all sites in the site list that were going to connect to the Internet any time soon (e.g., until even very small companies routinely establish continuous Internet connections). This hints that on average, the time from a site acquiring a

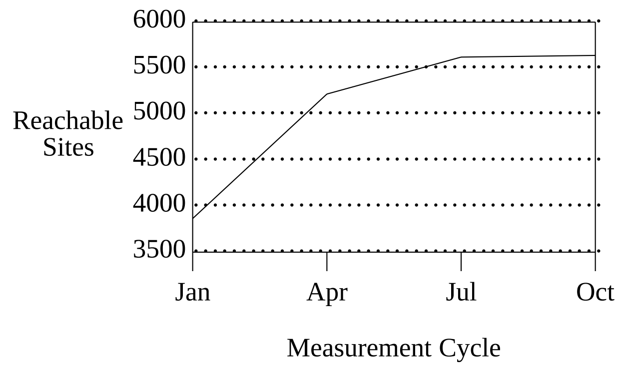

Figure 2: Total Reachable Sites

Domain name to gaining Internet connectivity is less than one year. In turn this indicates the coming ubiquity of Internet connectivity. In the past a common mode was having a Domain name with just a periodic dialup (often UUCP protocol-based) connection. Figure 2 indicates that sites now tend to acquire both a name and Internet connectivity within a small time frame. While dialup UUCP is less expensive, the Internet is more capable, and tends to draw organizations into connectivity.

Figure 3 plots the site counts for each top-level domains that showed up as one of the ten with highest count during each of the measurement runs.

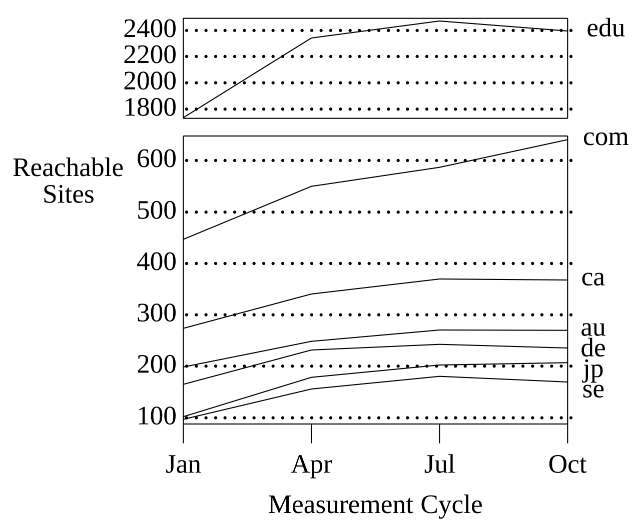

Figure 3: Most Reachable Top-Level Domains

The fact that connectivity counts decreased a bit for some of the sites in later measurement cycles indicates a combination of service disconnections and network availability problems - both of which count as disconnections, based on our definition of Internet connectivity. Interestingly, the plots do not cross, indicating that growth is fairly consistent across these top-level domains. It is also interesting to note that we saw approximately the same order of the first few top-level domains in other network connectivity data for completely different networks, such as UUCP, FidoNet, and BITNET (Quarterman, 1993b). This hints that network connectivity is a sociological phenomenon, with certain trends that transcend particular networking technologies. For more discussion of this point, see (Quarterman, 1993a).

Figure 4 plots reachable site counts for each country.[5]

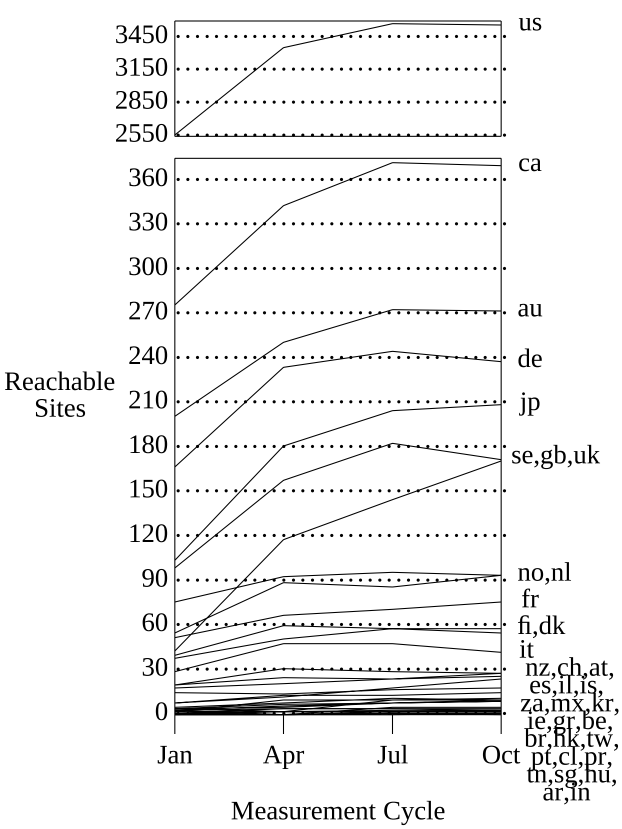

Figure 4: Reachable Sites by Country

The U.S. had more sites than all other countries combined, because the Internet originated in the U.S. (as the ARPANET). Yet, Figure 5 shows that several less-populated countries lead the U.S. in connected sites per unit population. These sites are often in colder or more geographically isolated regions of the world, where Internet connectivity presumably offers even more enticing advantages than it does to U.S. sites. It is interesting to note the similar growth patterns in this plot for the more highly networked countries that are culturally and geographically close, such as Norway and Sweden, the U.S. and Canada, and Finland and Denmark.

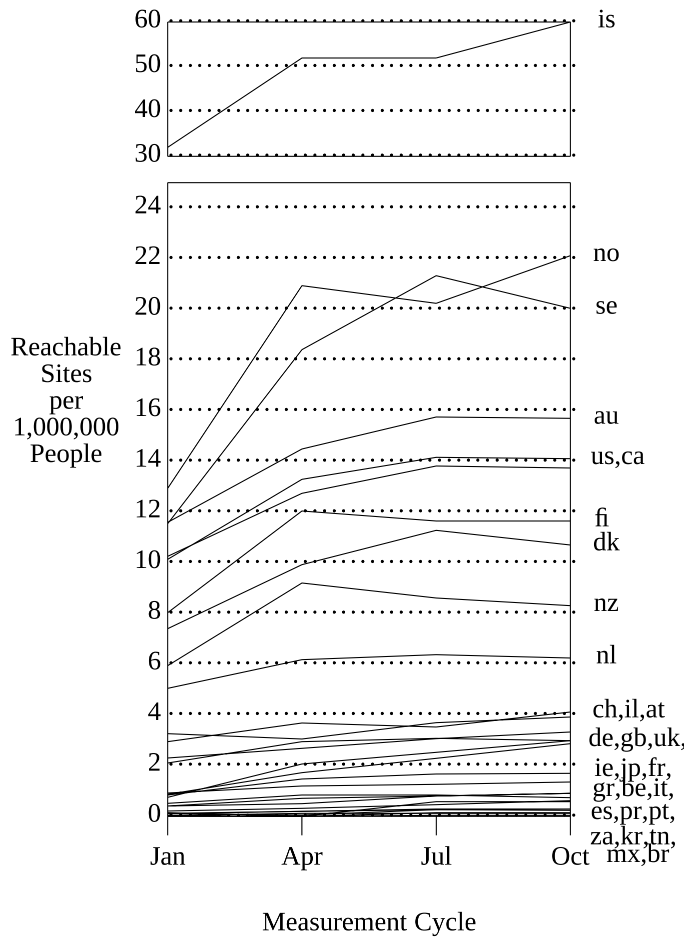

Figure 5: Reachable Sites by Population

Site Distancing Measurements

Figure 6 plots the percentage of top-level domains that were reachable at some earlier time and then became *isolated* for the remainder of the measurement runs (through October 1992), for each measurement cycle. By isolated we mean that a host in a higher level domain than that site could be reached, but no hosts in the site in question could be reached. For example, if hosts within the colorado.edu domain could be reached but none within cs.colorado.edu could be reached, cs.colorado.edu would be an isolated domain. This definition highlights sites that use firewall gateways that only allow mail and news to be exchanged through forwarding.

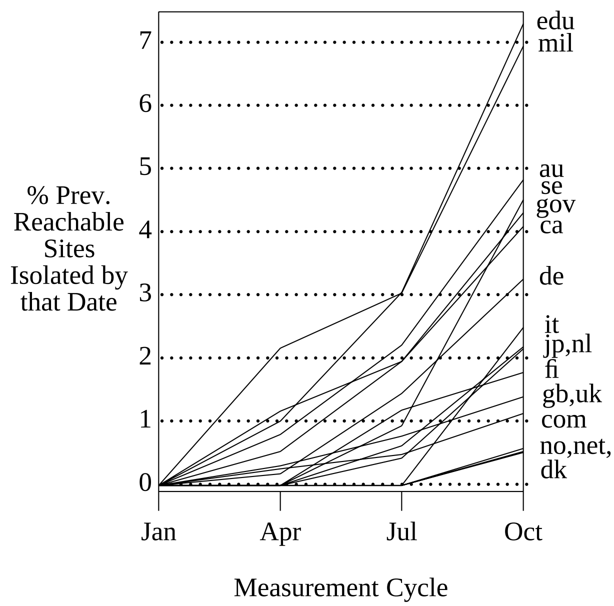

Figure 6: Sites Becoming Isolated

A top-level domain's being low in Figure 6 indicates either that its subdomains do not significantly distance themselves from the Internet, or that they had already distanced themselves before our first measurement cycle. For example, sites in the ''com'' domain often connect to the Internet with firewall gateways installed from the outset.

Figure 7 plots changes in service reachability as a function of type of institution. For this analysis we divided the top-level domains based on naming conventions where available - such as ''com'' in the U.S. and ''co.kr'' in Korea. This figure indicates that reachability does not change much within a type of institution, with the surprising exception of educational sites. We had expected the largest changes proportionately to be in the commercial domains. A likely explanation is that educational sites tend to start off directly connected to the Internet, while other types of institutions typically take a more cautious approach.

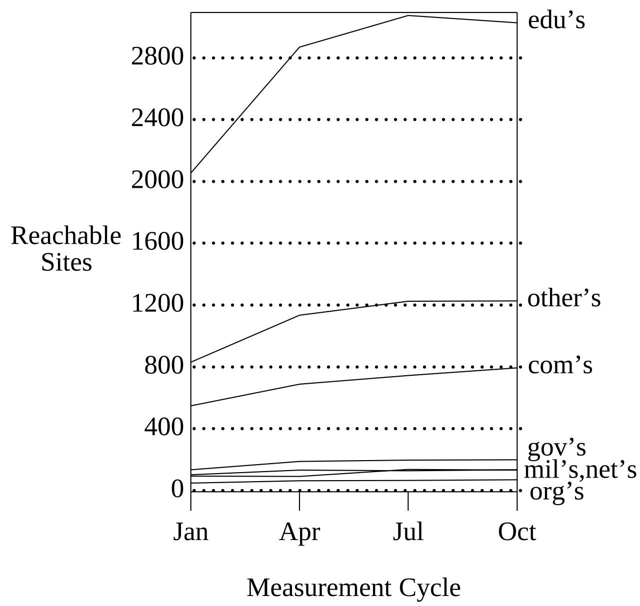

Figure 7: Reachability Changes by Type of Site

Figure 8 plots changes in reachability of the tested services. As expected, telnet, FTP, and SMTP remain the most commonly available services, although there was a slight decrease in the final measurement cycle. Somewhat surprisingly, potentially invasive services like telnet and FTP exhibit very similar curves to SMTP, even though the latter is an ''at arms length'' service. Apparently, sites do not tend to turn off just the most security sensitive services. Also somewhat surprising was the fact that who, finger, and daytime were reachable nearly as

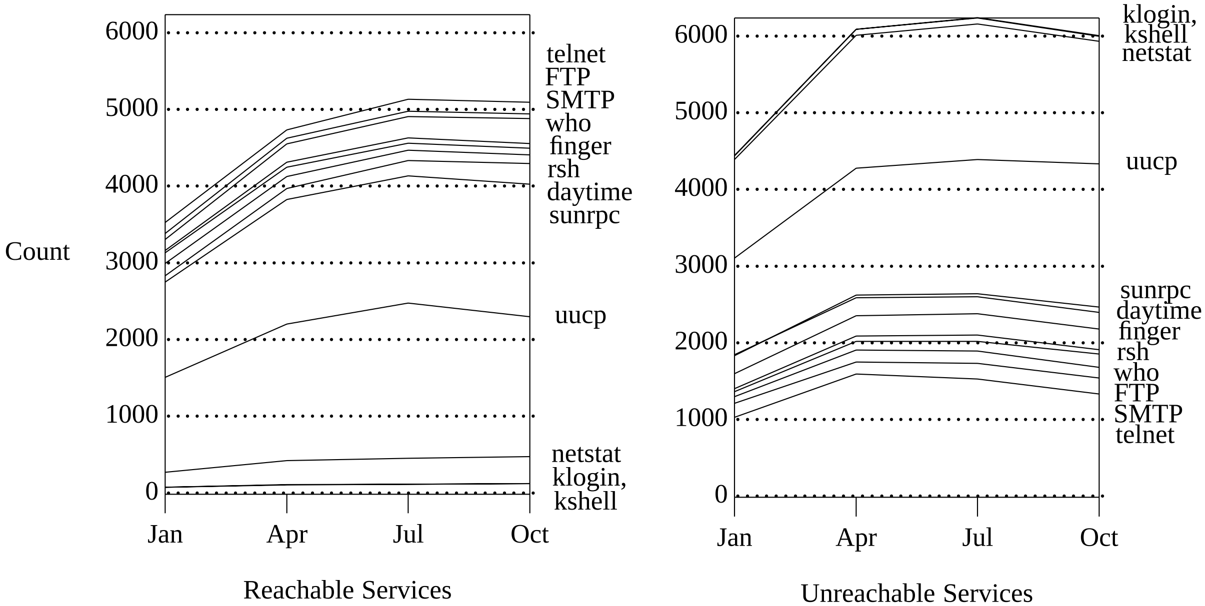

Figure 8: Reachability by Service

often, even though they support less general-purpose applications and are used less frequently (as indicated by NSFNET network traffic measurements). Still, based on the plots, there appears to be a trend to turn off these services.

Sun RPC and UUCP-over-TCP show the most marked drops. The former could indicate the increasing heterogeneity of Internet-connected systems (reaching beyond just UNIX systems), plus the increasing use of libraries and toolkits that hide the RPC layer (such as the X window system libraries). We speculate that people are turning UUCP-over-TCP off entirely, because the number of connections refused for UUCP-over-TCP is not decreasing as quickly as for the other services. The only services that seem to continue gaining in reachability are netstat, kshell, and klogin. The fact that refused connections dropped off more precipitously for kshell and klogin in the final measurement run may indicate a move to more authenticated services in the Internet. Possibly netstat shows increases because people are using it to set ‘‘tcpd’’ traps. If so, that would make all three of netstat, kshell, and klogin security features, indicating that people are probably changing them all at once for improved security.

Putting Together the Trends

In this section we estimate overall Internet growth rates, taking into account the rates of creation of new domains, domains connecting to the Internet, and domains disconnecting from the Internet. To do this, we created regression models for each of the top-level domains from Lottor’s host count data (1992a; 1992b), and scaled these functions by domain connection probabilities computed from the current study data.

For the host creation rate regression models we first dropped measurement values that were obviously the result of network problems. For example, Lottor’s April 1992 run could not measure any hosts in Korea, even though 1,506 and 2,902 were measured in January and July, respectively. Note that these models are imperfect because some sites do not allow ‘‘zone transfers’’, which could mean entire subtrees are excluded from consideration.

We tried both linear and curvilinear regression models on each of the top-level domains, and selected the model with the largest coefficient of determination (except in cases where both were quite high, in which case we preferred the exponential functions). Where linear models fit better, the explanation was probably that only one year's worth of data was available that gave breakdowns by top-level domain (Lottor, 1992b), and the growth functions had all passed the knees of their presumably exponential growth curves by that time. It is also possible that those growth rates truly are linear, although that seems less likely.

We computed connection probabilities by plotting the site reachability counts for each measurement cycle, approximating the asymptotes the curves were approaching, and dividing these values by the total site counts for each. For example, we tested 230 ‘‘mil’’ domains, and found an approximate asymptote at 160, leading to a .70 connection probability.

Table 3 shows the computed regression curves and domain connection probabilities for top-level domains that were measured by both studies, and for which the current study tested at least 10 domains. Looking at the regression models, the ‘‘com’’ domain is clearly the fastest growing: all of the top-level domains with exponential growth functions that have larger bases have much smaller multiplicative constants. These other models primarily reflect networked populations that, because of their small size, have rapid growth rates that will likely slow in the near future.

We also computed overall growth rates (listed as ‘‘ALL’’ in the table). These rates are much more accurate, because 10 years worth of that data was available (Lottor, 1992a). Note that while the DNS data show no leveling off in the domain creation rate, at some point growth must level out. Hence, our models only have predictive value during perhaps the next two to three years, during which time exponential growth will likely continue to occur.

Internet connectivity growth functions can be obtain by multiplying each regression function in Table 3 by the corresponding connection probability. Figure 9 plots these functions for each top-level domain (plus all domains combined) whose coefficient of determination exceeded .9, and for which we tested at least 10 subdomains and were also able to compute a domain connection probability. Note that the X-intercept of the various curves represent the modeled time when that domain joined the Internet, based on Lottor's DNS data.

While it would be interesting to project when counts will cross each other, for the sake of legibility we only plotted through January 1994. Moreover, we could not predict the time when growth will level out, which would be a critical aspect of such projections.

Discussion

In this section we discuss two broad sets of implications of this study: trends in service development and international access concerns.

Service Development

Our data indicate a number of implications on Internet service development.

First, while sites typically established only a domain name and periodic dialup mail/news connections in the past, sites now tend to acquire Internet connectivity within a year of establishing a name. Primitive dialup services (such as dialup UUCP) are decreasingly prevalent, being offset by more powerful interactive services (such as telnet). In time newer types of networked information discovery and retrieval services (such as WAIS and Gopher) may reach similarly ubiquitous levels.

Second, the rate of disconnection is far outstripped by the growth rate of new sites being

Top Level Domain	Regression Model	Coeff. of Determ.	Conn. Prob.	Top Level Domain	Regression Model	Coeff. of Determ.	Conn. Prob.
aq	$Y = 3.45e+04 * 0.93^{X}$	1.00	N	in	$Y = 1.73e-05 * 1.10^{X}$	0.95	F
ar	$Y = 2.55e-37 * 1.93^{X}$	0.75	0.14	int	$Y = 9.83e-03 * 1.07^{X}$	0.86	F
arpa	$Y = 3.74e-05 * 1.10^{X}$	1.00	0.00	is	$Y = 2.03e-05 * 1.14^{X}$	0.99	0.70
at	$Y = 3.76e-02 * 1.10^{X}$	0.97	N	it	$Y = 2.54e-02 * 1.10^{X}$	0.98	0.38
au	$Y = 2821.50X + 3.21e+05$	1.00	0.55	jp	$Y = 6.71e-02 * 1.10^{X}$	0.99	0.34
be	$Y = 151.00X + 1.84e+04$	0.94	N	kr	$Y = 1.75e-03 * 1.12^{X}$	1.00	0.41
br	$Y = 144.57X + 1.80e+04$	0.88	N	lu	$Y = 3.49e+01 * 1.01^{X}$	0.96	F
ca	$Y = 1996.83X + 2.23e+05$	1.00	0.67	mil	$Y = 7.09e-01 * 1.09^{X}$	0.98	0.70
ch	$Y = 1.73e+01 * 1.05^{X}$	1.00	0.34	mx	$Y = 3.31e-07 * 1.18^{X}$	0.99	0.71
cl	$Y = 3.32e-10 * 1.23^{X}$	1.00	F	net	$Y = 6.63e-04 * 1.13^{X}$	0.93	0.72
com	$Y = 1.04e+02 * 1.06^{X}$	0.97	N	nl	$Y = 1104.73X + 1.24e+05$	0.93	0.42
cs	$Y = 98.47X + 1.24e+04$	0.98	F	no	$Y = 736.87X + 8.18e+04$	0.99	N
de	$Y = 2148.50X + 2.37e+05$	0.97	0.45	nz	$Y = 97.30X + 1.10e+04$	0.96	0.39
dk	$Y = 6.25e-03 * 1.11^{X}$	0.94	0.31	org	$Y = 1037.30X + 1.10e+05$	0.99	N
ec	$Y = 2.80X + 3.56e+02$	0.60	F	pl	$Y = 88.53X + 1.09e+04$	0.99	F
edu	$Y = 7.03e+02 * 1.05^{X}$	1.00	0.66	pt	$Y = 81.20X + 9.32e+03$	0.98	N
ee	$Y = 6.80e-10 * 1.21^{X}$	1.00	F	se	$Y = 5.23e+02 * 1.03^{X}$	0.92	0.34
es	$Y = 348.97X + 4.20e+04$	1.00	N	sg	$Y = 72.27X + 8.51e+03$	0.97	F
fi	$Y = 693.13X + 7.46e+04$	0.98	0.35	su	$Y = 4.50X + 5.72e+02$	0.60	F
fr	$Y = 960.17X + 1.07e+05$	1.00	N	th	$Y = 7.16e-14 * 1.26^{X}$	1.00	F
gb,uk	$Y = 3220.23X + 3.84e+05$	1.00	N	tn	$Y = -1.60X + 2.20e+02$	0.55	N
gov	$Y = 1.04e+02 * 1.05^{X}$	1.00	0.64	tw	$Y = 6.14e-05 * 1.14^{X}$	0.96	F
gr	$Y = 36.57X + 4.20e+03$	0.91	N	us	$Y = 3.54e-06 * 1.15^{X}$	0.88	0.03
hk	$Y = 267.93X + 3.30e+04$	0.97	1.00	ve	$Y = 9.92e-02 * 1.03^{X}$	0.75	F
hu	$Y = 3.14e-11 * 1.24^{X}$	0.99	F	yu	$Y = 5.50e-01 * 1.03^{X}$	1.00	F
ie	$Y = 1.44e-03 * 1.10^{X}$	0.98	N	za	$Y = 209.07X + 2.52e+04$	0.99	N
il	$Y = 99.10X + 1.03e+04$	0.91	N	**ALL**	$Y = 9.77e+01 * 1.07^{X}$	0.98	0.41

Table 3: Overall Reachable Internet Site Growth Regression Models and Connection Probabilities

Incorporates DNS growth rates, rates of connection to the Internet, and rates of distancing from the Internet
X = months since August 1981; *Y* = modeled host count
''F'' for connection prob. = collected data about fewer than 10 domains
''N'' for connection prob. = was not approaching an asymptote during measurements

created, with commercial institutions comprising the fastest growing segment[6]. In the short to medium term, this means there will be rapid overall increases in the number of reachable Internet services. In the longer term, new domain creation growth rates must inevitably slow, at which point network distancing mechanisms will have a bigger impact on the Internet service infrastructure.

There are two conflicting trends in the development of Internet service infrastructure. Because commercial institutions are growing rapidly and tend to make significant use of distancing mechanisms, the Internet service infrastructure as we currently know it (i.e., free, publically accessible network services) will likely be supported by a decreasing proportion of Internet sites, comprised primarily of non-profit, government, and academic institutions. At the same time, once the technology and market for commercial, for-fee services is firmly established in the Internet, an explosion in new types of services will likely take place.

At least three types of changes must happen before commercial services can become widespread on the Internet. First, there must be support for commerce-grade (authenticated and private) communication. Our data indicate a gradual move towards authenticated services (such as klogin), but for example, Privacy Enhanced Mail (Linn, 1987) has reached very limited deployment to date. Second, the various commercial entities must become convinced of the readiness of the marketplace for this new type of service. While there is a great deal of discussion taking place about commercial network services (e.g., on the com-priv@psi.com mailing list), at present there are very few examples of commercial services available over the Internet. Third, there

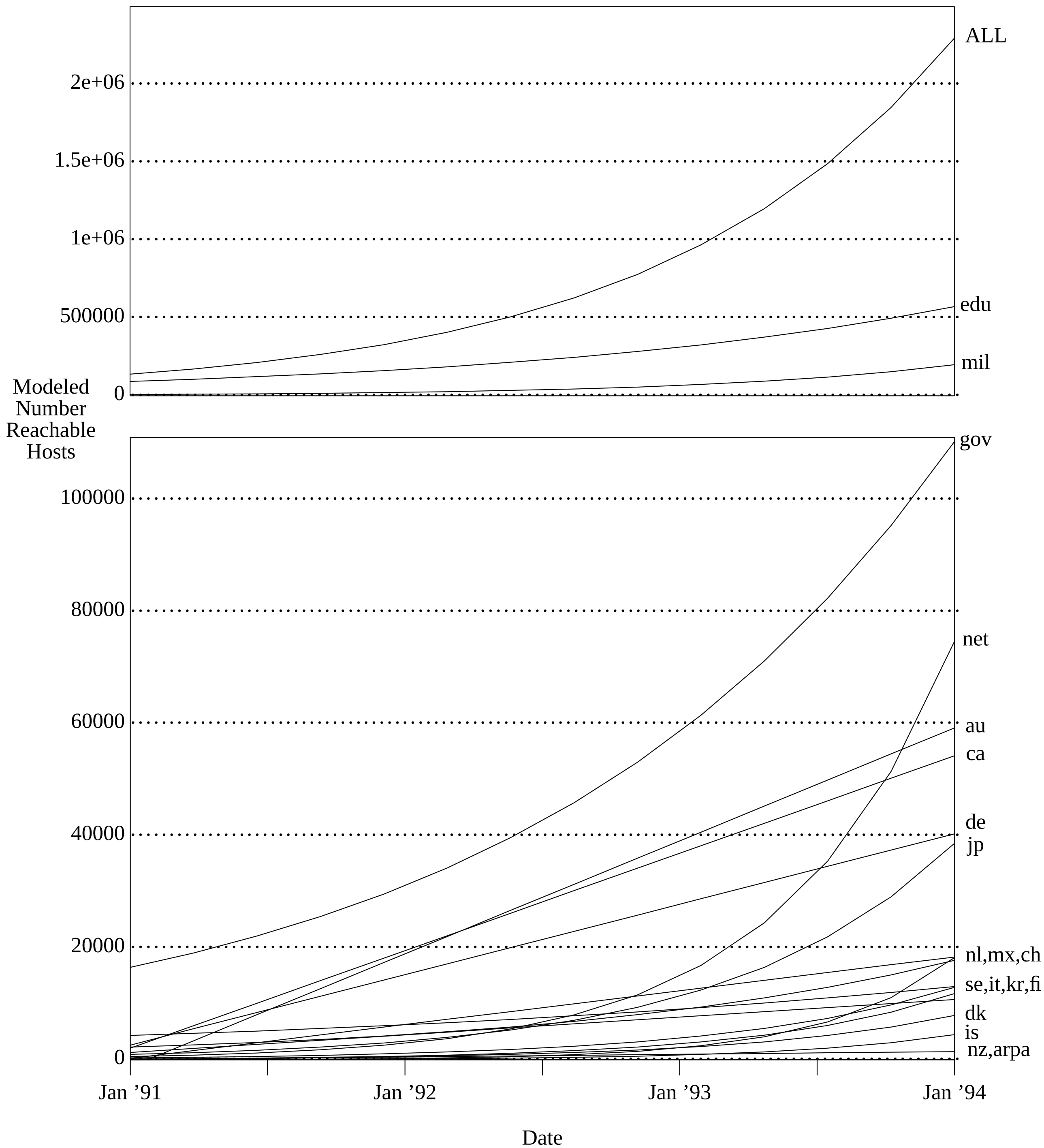

Figure 9: Modeled Overall Growth Rates

must be electronic banking services, so that network service retailers are not forced to do billing directly.

For all types of sites, security is an issue of increasing concern. While commercial sites tend to start off wary (e.g., using firewall gateways when they first connect to the Internet), non-profit institutions tend to start off with a more trusting approach, and over time install more secure networking mechanisms. These mechanisms tend to limit access below the level of network services. For example, we found that potentially invasive services (like telnet and FTP) exhibit very similar growth curves as SMTP, even though the latter is typically regarded as less security sensitive. Moreover, even less general-purpose services (such as who and finger) were reachable nearly as often, indicating that sites tend to leave network services in place, and instead limit access to the site itself (for example, through the use of a firewall gateway), or in some cases use selective service filtering mechanisms.

International Access Concerns

The Internet originated in the United States, and at present there are more service-reachable sites in the U.S. than in all other countries combined. In time the Internet connectivity of other countries will catch up to that of the U.S., since the U.S. has a lower Internet growth rate than many other countries. Yet, from our data this will not happen for at least a few years. During this time, the predicted number of service-reachable U.S. commercial, educational, military, and government sites will each remain ahead of those of any other country.

If, on the other hand, one counts service-reachable sites per unit population, several less-populated countries lead the U.S. These sites are often in colder or more geographically isolated regions of the world, where Internet connectivity presumably offers even more enticing advantages than it does to U.S. sites. Some of these countries (such as Norway) also have other practical advantages that lead to high Internet connectivity growth rates, including wealth, government funding for networking, and the existence of multiple commercial IP providers.

Beyond providing access to more isolated regions of the world, Internet connectivity seems a sociological phenomenon, with trends that transcend particular networking technologies: we saw approximately the same order of most-reachable top-level domains in other network connectivity data for completely different networks, such as UUCP, FidoNet, and BITNET. We also note that once countries reach a certain level of ‘‘Internet connectivity maturity’’, growth patterns tend to be similar in countries that are culturally and geographically close, such as Norway and Sweden, the U.S. and Canada, and Finland and Denmark. Before countries reach this state of service-reachablability, their growth rates are correlated most strongly with wealth (Quarterman, 1993a), and less with other indications, such as population and technical sophistication. Moreover, the least developed countries have the highest host growth rates (Quarterman & Phillips, 1993). These observations might be important to potential commercial service providers, as market predictability is an important commodity.

Conclusions

When faced with the question of how fast the Internet is growing and what that growth will mean to commerce, research and education, and the society at large, people typically turn to numbers provided by network registration tables or traffic counts. These measurements are difficult to interpret, because of the range of different ways that sites connect to and use the Internet. In this paper we presented measurements of Internet growth based on tests of what services could be reached at over 13,000 sites, in four different measurement cycles over the course of 1992. Our analysis uncovers a number of issues.

While some sites are clearly distancing themselves from the Internet, the rate of disconnection is far outstripped by the growth rate of new domains being created. In the short to medium term, this means there will be rapid overall increases in the number of reachable Internet services. In the longer term, once most of the existing departments and divisions of the world’s sites have established Internet connectivity, growth will be dominated by newly created sites. At this time, the current exponential growth rate of the Internet will slow to the growth rate of new sites. This slow down will mean a bigger impact of network distancing mechanisms on the Internet service infrastructure.

The use of distancing mechanisms appears most closely tied to type of institution, rather than to attitudinal changes developed over time. At present, commercial institutions are the most rapidly growing sector, and these institutions tend to make significant use of distancing mechanisms. Thus, over time the current set of free Internet services will likely comprise a decreasingly small proportion of the Internet, to be displaced by commercial IP services (charging by bandwidth, not per-byte or per-message metered service), once the technology and market for them is firmly established. For non-profit as well as commercial services, a key issue will be resolving the tension between connectivity and security.

On average, sites tend to acquire Internet connectivity and settle on a comfortable level of security/distancing within a year after connecting to the Internet (although in some cases they change their methods later). While the probability that sites directly connect to the Internet varies significantly as a function of geography and type of institution, on average 41% of the time sites acquire direct Internet connectivity. Based on this and similar asymptotic computations, we have constructed regression models of overall Internet growth rates, taking into account the rates of creation of new domains, domains connecting to the Internet, and domains disconnecting from the Internet. These models can be used to analyze a number of different questions about Internet connectivity.

At present, the Internet is heavily U.S.-centric, with more service-reachable sites in the U.S. than in all other countries combined. Other countries will catch up in time, since the U.S. Internet connectivity growth rates are slower than those in many other countries. However, U.S. connectivity will exceed the connectivity of other countries for at least the next few years.

Future Work

While this study offers a new set of metrics for the growth of the Internet, it leaves a number of questions unanswered. Perhaps most important is limitations in the the types of conclusions that can be drawn based on measurements of reachable services. For example, it would be interesting to measure the growth rates of client-only sites, i.e., sites that make use of Internet services without exporting any services of their own. Measurements of client use of the Internet might provide data about potential market sizes. Collecting these measurements would require monitoring network traffic to detect the existence of clients accessing remote services. Clearly, such measurements would need to be done with attention to privacy and network security concerns.

It would also be interesting to measure growth rates in the new generation of networked information discovery and retrieval tools, such as WAIS and Gopher. We were not able to collect such measurements for the current study because of our random sampling methodology. It would be useful to know the relative growth rates of these new types of services, compared with their more ‘‘mundane’’ predecessors (remote login, file transfer, etc.), and to use this information to predict trends in what types of service are most popular/important. Collecting these measurements could be done by periodically collecting lists of servers where available (e.g., from the WAIS directory of servers, or by traversing pointers between Gopher servers). These servers might also be discovered by monitoring network traffic, but doing so again raises privacy and security issues.

It might also be worthwhile to conduct a study similar to the current study, but instead attempting connections from several different types of sites concurrently. Doing so could uncover non-uniformities in network accessibility of Internet services. As an example of non-uniform network accessibility, in March of 1992 MILNET began restricting their advertising of selected foreign commercial and educational networks, because their routers could not handle the rapid growth of network routes.

While the above measurements cannot be determined from the data we collected, there are a variety of other characteristics that could be uncovered from our data. As one example, we could analyze trends in the types of mechanisms being used for distancing (e.g., firewall gateways vs. turning off services). We could also analyze Internet reliability and distribution of message transit times, based on logs of our measurement traffic. Finally, we could compare the rates of Internet service reachability changes with network traffic rate changes, to indicate changing uses of the Internet (e.g., to uncover increasing use of bandwidth intensive applications like audio and video, which could cause traffic to increase faster than sites are being connected to the Internet).

Acknowledgements

We thank Dave Clark of MIT for mentioning the idea of ‘‘The Great Disconnection’’ at the Fall 1990 Interop conference. We thank Mark Lottor of SRI International for his measurements of Domain Naming System growth. We thank the University of Colorado systems administration staff and the Computer Emergency Response Team for their helpful support in responding to inquiries raised by our measurement runs. We thank Panagiotis Tsirigotis for correcting a deficiency in the SunOS kernel that affected our measurement software.

This paper is based in part on an earlier conference publication (Schwartz & Quarterman, 1993).

Schwartz was supported for this work in part by the National Science Foundation under grants DCR-8420944 and NCR-9105372, a grant from AT&T Bell Laboratories, and a grant from Sun Microsystems’ Collaborative Research Program. Quarterman’s participation was part of an ongoing private project analyzing the growth and extent of several networks, conducted by Matrix Information and Directory Services, Inc.

Notes

1. Until Spring 1993 there was a single NIC for the U.S. After this time the NIC split into military and civilian branches. The related work discussed in this paper are all based on the original single NIC, which we refer to as "the U.S. NIC".

2. UUCP is a network protocol (Nowitz & Lesk, 1978) that can run over dialup lines or over TCP/IP. It also refers to a worldwide dialup network based on the UUCP protocol (Quarterman, 1990). To distinguish between these different meanings, throughout this paper we refer to either the UUCP protocol, UUCP-over-TCP, UUCP links (which use the UUCP protocol), or the UUCP network.

3. The correspondence between top-level domains and geography is actually not precise. For example, there are Canadian, Swiss, and other nationalities of companies in the ‘‘com’’ domain. For more information, see (Quarterman, 1992).

4. Note in particular this means that the security mechanism behind individual network services was not tested.

5. In this and the following country plot, the U.S. entries included domains in the ‘‘com’’, ‘‘edu’’, ‘‘gov’’, ‘‘mil’’, and ‘‘us’’ domains. In all other plots, ‘‘us’’ refers explicitly to the top-level ‘‘us’’ domain.

6. In fact, Rutkowski found that commercial institutions now comprise the largest number of re-

gistered domains (1993).

References

Blokzijl, Rob. (1990). RIPE terms of reference. Request For Comments 1181, RIPE/NIKHEF.

Braun, Hans-Werner, Chinoy, Bilal, Claffy, Kimberly C. and Polyzos, George C. (1993). Analysis and modeling of wide-area networks: Annual status report.

CCITT/ISO. (1988). *The directory, Part 1: Overview of concepts, models and services.* CCITT/ISO, Gloucester, England. CCITT Draft Recommendation X.500/ISO DIS 9594-1.

Caceres, Raymond, Danzig, Peter B., Jamin, Sugih and Mitzel, Daniel J. (1991). Characteristics of individual application conversations in TCP/IP wide-area internetworks. *Proceedings of the SIGCOMM Conference*, pp. 101-112, Zurich, Switzerland.

Carl-Mitchell, Smoot and Quarterman, John S. (1992). Internet firewalls. *UNIX/World*, 9(2), pp. 93-102, Tech Valley Publishing, Mountain View, California.

Carl-Mitchell, Smoot and Quarterman, John S. (1993). Local protection for networked systems. *UNIX/World*, 10(7), Tech Valley Publishing, Mountain View, California.

Cerf, Vinton G., editor. (1991). Guidelines for Internet measurement activities. Request For Comments 1262, Internet Activities Board.

Claffy, Kimberly. (1993). Network analysis issues for a public Internet. *Proceedings of the Public Access to the Internet Workshop*, John F. Kennedy School of Government, Harvard University, Cambridge, Massachusetts.

Emtage, Alan and Deutsch, Peter. (1992). Archie - an electronic directory service for the Internet. *Proceedings of the USENIX Conference*, pp. 93-110, San Francisco, California.

Feinler, Elizabeth, Harrenstien, Ken, Su, Zaw-Sing and White, Vic. (1982). DoD Internet host table specification. Request For Comments 810, Network Information Center, SRI International.

Ganatra, Nitin K. (1992). Census: Collecting host information on a wide-area network. Technical Report Number 92-34, University of California, Santa Cruz.

Heimlich, Steven A. (1990). Traffic characterization of the NSFNET national backbone. *Proceedings of the SIGMETRICS Conference on Measurement and Modeling of Computer Systems*, pp. 257-258, Boulder, Colorado.

Kahle, Brewster and Medlar, Art. (1991). An information system for corporate users: Wide area information servers. *ConneXions - The Interoperability Report*, 5(11), pp. 2-9, Interop, Inc.

Linn, John. (1987). Privacy enhancement for internet electronic mail: Part I: message encipherment and authentication procedures. Request For Comments 989.

Lottor, Mark. (1990). Personal communication. Discussion of Internet measurements to count number of domains and machines in the Internet.

Lottor, Mark. (1992a). Internet growth (1981-1991). Request For Comments 1296, Network Information Systems Center, SRI International.

Lottor, Mark. (1992b). *Internet Domain Survey Results.* Network Information Systems Center, SRI International. Anonymous FTP from ftp.nisc.sri.com: pub/zone/.

MIDS. (1993). The Matrix mapped and measured. *Matrix News*, 3(3), pp. 1-2,7, Matrix Information and Directory Services, Austin.

McCahill, Mark. (1992). The Internet gopher: A distributed server information system. *ConneXions - The Interoperability Report*, 6(7), pp. 10-14, Interop, Inc.

Merit, Inc. (1992). NSFNET Statistics. Anonymous FTP from nis.nsf.net:statistics/nsfnet/*.

Mockapetris, Paul. (1987). Domain names - Concepts and facilities. Request For Comments 1034, USC Information Sciences Institute.

Muuss, Michael. (1983). *Ping software.* U. S. Army Ballistic Research Laboratory. Avaliable by anonymous FTP from uunet.uu.net: /bsd_sources/src/ping.

NNSC. (1989). *NSF Network News.* NSF Network Service Center. Anonymous FTP from nis.nsf.net:statistics/nsfnet/*.

National Research Council. (1991). *Computers at Risk: Safe Computing in the Information Age.* National Academy Press, Washington, D.C.

Nowitz, David A. and Lesk, Michael E. (1978). *A dial-up network of UNIX systems.* Bell Laboratories, Murray Hill, New Jersey.

Paxson, Vern. (1991). Measurements and models of wide area TCP conversations. LBL technical report #30840, Lawrence Berkeley Laboratory Computer Systems Engineering Group.

Postel, Jon and Reynolds, Joyce. (1985). File Transfer Protocol (FTP). Request For Comments 959, USC Information Sciences Institute.

Quarterman, John. (1990). *The Matrix: Computer Networks and Conferencing Systems Worldwide.* Digital Press.

Quarterman, John S. (1992). Where is the Internet? *Matrix News*, 2(8), pp. 9-15, Matrix Information and Directory Services, Inc. (MIDS), Austin.

Quarterman, John S. (1993a). Hosts by wealth. *Matrix News*, 3(5), pp. 1-6, Matrix Information and Directory Services, Inc. (MIDS), Austin.

Quarterman, John S. (1993b). The most networked countries. *Matrix News*, 3(3), pp. 4-7, Matrix Information and Directory Services, Inc. (MIDS), Austin.

Quarterman, John and Phillips, Gretchen. (1993). Network host growth rates. *Matrix News*, 3(7), Matrix Information and Directory Services, Inc, Austin, Texas.

Romkey, John. (l988). A nonstandard for transmission of IP datagrams over serial lines: SLIP. Request For Comments 1055.

Rutkowski, Anthony M. (1993). Internet Metrics. *Internet Society News*, 2(1), pp. 33-38.

Schwartz, Michael F. (1991a). A measurement study of changes in service-level reachability in the global TCP/IP Internet: Goals, experimental design, implementation, and policy considerations. Request For Comments 1273. Department of Computer Science, University of Colorado, Boulder.

Schwartz, Michael F. and Tsirigotis, Panagiotis G. (1991a). Experience with a semantically cognizant Internet white pages directory tool. *Journal of Internetworking: Research and Experience*, 2(1), pp. 23-50.

Schwartz, Michael F. (1991b). The Great Disconnection? Technical Report CU-CS-521-91, Department of Computer Science, University of Colorado, Boulder.

Schwartz, Michael F. and Tsirigotis, Panagiotis G. (1991b). Techniques for supporting wide area distributed applications. Technical Report CU-CS-519-91, Department of Computer Science, University of Colorado, Boulder.

Schwartz, Michael F., Emtage, Alan, Kahle, Brewster and Neuman, B. Clifford. (1992). A comparison of Internet resource discovery approaches. *Computing Systems*, 5(4), pp. 461-493.

Schwartz, Michael F. (1992). How big is the Internet? *Internet Society News*, 1(2), pp. 3-5.

Schwartz, Michael F. and Pu, Calton. (1993). Application-specific resource discovery in a changing and growing internet. Technical Report, Department of Computer Science, University of Colorado, Boulder. In preparation.

Schwartz, Michael F. and Wood, David C. M. (1993). *Discovering shared interests using graph analysis.* Department of Computer Science, University of Colorado, Boulder. To appear, Communications of the ACM.

Schwartz, Michael F. and Quarterman, John S. (1993). A measurement study of changes in service-level reachability in the global Internet. *Proceedings of INET '93*, San Francisco, California.

Spafford, Eugene H. (1989). The Internet Worm: Crisis and aftermath. *Communications of the ACM*, 32(6), pp. 678-687.

Stoll, Clifford. (1988). Stalking the wiley hacker. *Communications of the ACM*, 31(5), pp. 484-497.

Wood, David C. M., Coleman, Sean S. and Schwartz, Michael F. (1993). Fremont: A system for discovering network characteristics and problems. *Proceedings of the USENIX Conference*, pp. 335-348, San Diego, California.

Zimmerman, David. (1990). The finger user information protocol. Request For Comments 1288, Center for Discrete Mathematics and Theoretical Computer Science.