

How Big is the Internet?
Michael F. Schwartz
University of Colorado - Boulder
Published in Internet Society News 1(2), Spring 1992

The question often arises, "How big is the Internet?" To answer this question, 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). This definition is counterintuitive, since it elides differences between various types of physical connectivity. In particular, it does not distinguish the parts of the network that can support interactive applications (like remote login) from dialup-based, mail-only connections. Given the advantages of interactive connectivity and the growing popularity of IP, in this article I consider only the interconnected IP Internet.

Lottor recently published results of a ten year study that counted the number of hosts in domains that have IP addresses registered in the DNS (as opposed to domains that register only "mail exchange" (MX) records that allow mail to be forwarded to through an intermediary host) [Lottor 1992]. In the early years the data were extracted from host tables maintained by the DDN Network Information Center. Later, measurements were taken by a program that recursively descends the Domain Naming tree, retrieving information about all domains that allow "zone transfers".

Many of the hosts counted by Lottor's study are hidden behind secure gateways or otherwise not directly 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. I believe a more meaningful measure of Internet size is the number of domains at which common network services can be contacted, since it is through such services that a site gains the advantages of connectivity.

I am currently performing such a study. Specifically, this study tracks changes in service-level reachability in the Internet [Schwartz 1991]. While the measurements will not be complete until the end of 1992, the first set of measurements that have been collected can be used to characterize the current size of the interconnected IP Internet. The final study will provide much more information than just Internet size. It will indicate relative growth rates among different countries, trends in the types of services to which sites limit access, how sites limit access to these services, and the types and geographical distribution of sites that distance themselves from the Internet.

Starting with a large list of domains, my study attempts to connect to the following TCP/IP services at each domain:

Port Number	Service	Port Number	Service
13	daytime	111	Sun portmap
15	netstat	513	rlogin
21	FTP	514	rsh
23	telnet	540	UUCP
25	SMTP	543	klogin
53	Domain Naming System	544	krcmd, kshell
79	finger		

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 is meaningful). The one exception is the Domain Naming System, for which the machines to probe are selected from information obtained from the Domain system itself. Only TCP services are tested, since the TCP connection mechanism allows one to determine if a server is running in an application-independent fashion.

From a list of approximately 12,700 Internet domains worldwide (generated from Lottor's January 1991 data plus a number of other sources), successful connections were recorded to at least one of the above services in 4,455 domains, broken down by top-level domain as follows:

Top-level Domain Name	Description	Number of Domains Reachable by Measured Internet Services
edu	U.S. Educational	2048
com	U.S. Commercial	494
ca	Canadian	299
au	Australian	278
de	German	174
se	Swedish	167
gov	U.S. Government	128
mil	U.S. Military	115
jp	Japanese	106
net	Named by network	96
nl	Dutch	84
org	Non-profit	56
fr	French	55
no	Norwegian	55
fi	Finnish	45
uk	British	44
it	Italian	39
dk	Danish	38
at	Austrian	21
nz	New Zealand	21
ch	Swiss	20
il	Israeli	16
is	Icelandic	8
es	Spanish	8
kr	Korean	5
be	Belgian	4
gr	Greek	4
za	South African	4
br	Brazil	3
ie	Irish	3
tw	Taiwanese	3
us	Other U.S.	3
arpa	ARPANET names	2
mx	Mexican	2
sg	Singapore	2
hk	Honk Kong	1
in	Indian	1
int	International	1
pt	Portuguese	1
tn	Tunisian	1

This list is a lower bound, since it depends on the span of the initial list of domains, and sites in other countries have connected to the Internet since this list was compiled. Nonetheless, the measurements provide an interesting point of comparison. For example, it is clear that the number of U.S. sites is much larger than the number of sites in any other country in the world. In fact, there are nearly twice as many U.S. sites as sites in all other countries combined. However, given the rapid growth rate of IP connectivity in other countries, within one to two years I expect there to be more sites internationally than in the U.S.

To help underscore the distinction between service-level connectivity and IP host count at Internet sites, I found that 7,242 domains in Lottor's January 1991 list (out of 11,194 in that list) were not reachable by the above Internet services. The ratio of service reachable to all IP domains may continue to decrease, as security problems garner increasing concern. The results of my study will help uncover the trend here.

The services reached by my measurement software were as follows:

Service	Number of Domains
telnet	4170
FTP	4027
SMTP	3952
rlogin	3811
rsh	3777
finger	3637
daytime	3492
Sun portmap	3421
UUCP	2217
Domain	1803
netstat	294
klogin	95
krcmd, kshell	93

From this list it is clear that the "Big Three" applications (remote login, file transfer, and mail) are the main services in use. Interestingly, UUCP appears in more domains than DNS, even though TCP based UUCP (as opposed to dialup UUCP) is being phased out of existence, as NNTP gains popularity. The reason for this is probably two fold. First, most domains contract DNS service from other domains, to avoid the administrative effort required to run a Domain server. Second, many computers probably come with UUCP configured in by the manufacturer.

For a discussion of the size of the set of computer networks interconnected for at least mail or news service (referred to as "The Matrix"), see [Quarterman 1992]. For a measure of the diameter of the interpersonal communication graph enabled by electronic mail, see [Schwartz & Wood 1992]. Anyone who is considering performing measurement studies of the Internet is urged to read [Cerf 1991].

References

[Cerf 1991]

V. G. Cerf, editor. Guidelines for Internet Measurement Activities. Req. For Com. 1262, Internet Activities Board, Oct. 1991.

[Lottor 1992]

M. Lottor. Internet Growth (1981-1991). Req. For Com. 1296, Network Information Systems Center, SRI Int., Jan. 1992.

[Quarterman 1992]

J. S. Quarterman. How Big is the Matrix? Matrix News, 2(2), Matrix Information and Directory Services, Austin, TX, mids@tic.com, Feb. 1992.

[Schwartz 1991]

M. F. Schwartz. A Measurement Study of Changes in Service-Level Reachability in the Global TCP/IP Internet: Goals, Experimental Design, Implementation, and Policy Considerations. Req. For Com. 1273, Dept. Comput. Sci., Univ. Colorado, Boulder, CO, Nov. 1991.

[Schwartz & Wood 1992]

M. F. Schwartz and D. C. M. Wood. Discovering Shared Interests Among People Using Graph Analysis of Global Electronic Mail Traffic. Dept. Comput. Sci., Univ. Colorado, Boulder, CO, Feb. 1992. Submitted for publication.