                        Internet Growth (1981-1991)

Status of this Memo

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

Abstract

   This document illustrates the growth of the Internet by examination
   of entries in the Domain Name System (DNS) and pre-DNS host tables.
   DNS entries are collected by a program called ZONE, which searches
   the Internet and retrieves data from all known domains. Pre-DNS host
   table data were retrieved from system archive tapes. Various
   statistics are presented on the number of hosts and domains.

Table of Contents

Introduction

   This document provides statistics on the growth of the Internet by
   examining the number of Internet hosts and domains over a 10-year
   period. Before the Domain Name System was established, practically
   all hosts on the Internet were registered with the Network
   Information Center (SRI-NIC) and entries were placed in the Official
   Host Table for each one. Data on the number of hosts for pre-DNS

   years comes from copies of the host table at selected times. The DNS
   system was introduced around 1984 but took almost 4 years before it
   was fully implemented on the Internet. However, by this time many
   hosts were no longer registered in the Host Table.

   In 1986, the ZONE (Zealot Of Name Edification) program was written.
   ZONE was originally intended to be used during the host-table-to-DNS
   transition period. ZONE would "walk" the DNS tree and build a host
   table of all the information it collected. This host table could
   then be used by sites that had not yet made the DNS transition.
   However, ZONE was never used for this purpose. Instead, it was found
   to be useful for collecting statistics on the size of the domain
   system and the Internet.

   ZONE could not collect complete data on the DNS until around 1988,
   because early versions of BIND (the popular Unix DNS implementation)
   had major problems with the zone transfer function of the DNS
   protocol. ZONE has been used in varying ways ever since to collect
   this information. In the first few years, it was used to produce a
   wall-size chart of the domain tree. However, the number of domains
   quickly outgrew the size of the wall and the charts were abandoned.
   In later years, statistics on the number of hosts and domains were
   extracted from the resulting host table, sometimes categorizing data
   based on top-level domain names or on computer system type or
   manufacturer.

   The time to gather the data also grew from hours to a week, and the
   size of the host table produced soon reached 50 megabytes. In order
   to reduce the amount of data collected, ZONE is now run in a mode
   collecting only host names and IP addresses, ignoring protocol, host
   information and MX record data. The host table is then groveled over
   by some utilities (such as sort, uniq and grep) to produce the
   statistics required. ZONE is currently run every 3 months at SRI.

How ZONE Works

   ZONE maintains a list of domains and their servers and a flag
   indicating whether information for a domain has been successfully
   loaded from one of the servers. Because of another bug in BIND, ZONE
   must be primed with a list of all the top-level domains and their
   name servers. It then cycles through the domain list, attempting to
   contact one of the servers for each domain not yet transferred. When
   a server is contacted (via TCP), a Start of Authority (SOA) query is
   first sent to make sure the server is authoritative for the domain
   being requested. If so, then a zone transfer query (AXFR) is sent to
   request all the resource records for the domain to be retrieved.

   When a name server record (NS) is received, the referenced domain and

server are added to the list of domains to process. When host
records (A, CNAME, HINFO, MX) are received, they are added to an in-
core table of host information. The program ends when it has cycled
through the entire list of domains without receiving any new
information. It then dumps the table of host information to a
HOSTS.TXT format file.

Problems with Data Collection

   For various reasons, some Internet sites do not allow zone transfers
   of their domain servers. ZONE also eventually gives up trying to
   transfer a domain after too many failures. The number of domains
   that could not be zone transferred during the 1-Jan-92 ZONE run was
   around 800 out of 17,000. Additionally, it is assumed that not all
   hosts on the Internet are registered in a domain server. These
   problems cause the statistics gathered by ZONE to be lower than the
   actual amounts.

   Manual review of some of the data collected by ZONE also shows a lot
   of random entries in the DNS. Misformatted entries may cause bogus
   server or host records to appear. Many times a server is found to
   not be authoritative for the domain listed. Sometimes entire domains
   are renamed and their old entries left in place for a transition
   period, thus causing each host within that domain to be counted
   twice. These problems cause the results of ZONE to be higher than
   the actual amounts.

   Manual scanning of the data indicates that the additional entries are
   insignificant compared to the missing entries discussed earlier.
   ZONE data can thus be viewed as the minimum number of Internet hosts,
   and not the actual figures.

   A final problem with data collection is that of expense. Downloading
   domain information from every domain on the Internet generates a
   large amount of network traffic. It also puts an extra CPU load on
   each domain server it must contact. An organized effort might be
   considered to have only one such program doing this on the Internet
   at regularly scheduled intervals to keep the problem of multiple data
   collectors from occurring.

Scope of the Study

   A problem with counting hosts and domains on the Internet is defining
   what the Internet really is. Finding host entries in the DNS does
   not necessarily indicate that the host is reachable from the
   Internet. Many companies have mail gateways between the Internet and
   their local nets, thus disallowing direct access. However, some of
   these companies advertise all their hosts, and some advertise only

the gateway. Are these hosts on the Internet or not?

Furthermore, many domains in the DNS are just mail-forwarding (MX)
entries for off-Internet (such as Usenet) sites. Are these domains
really part of the Internet and should they be counted in an Internet
size study?

For the purposes of this study, a host has been defined as a
[name(s),IP-address(es)] grouping discovered from the DNS. This
prevents us from counting a host with multiple names or addresses
more than once. However, this does not consider whether the host is
directly accessible or not. When ZONE counts the number of domains
it includes all domains referenced by an NS record in the DNS, thus
including MX-only domain sites in the final results.

N. Results

This section presents data from archive tapes of SRI-NIC from 1981 to
1986, and statistics gathered by runs of ZONE from 1986 to 1992.

N.1 Number of Internet Hosts

The chart below shows the number of IP hosts on the Internet. These
are hosts with at least one IP address assigned. Data was collected
by ZONE except where noted. The following two sections are graphs of
the data in this chart.

```
        Date                Hosts

        08/81                 213       Host table #152
        05/82                 235       Host table #166
        08/83                 562       Host table #300
        10/84               1,024       Host table #392
        10/85               1,961       Host table #485
        02/86               2,308       Host table #515
        11/86               5,089
        12/87              28,174
        07/88              33,000
        10/88              56,000
        01/89              80,000
        07/89             130,000
        10/89             159,000
        10/90             313,000
        01/91             376,000
        07/91             535,000
        10/91             617,000
        01/92             727,000
```

                       Number of Internet Hosts (linear)
```
800|
780|
760|
740|                                                                   *
720|
700|
680|                                                                 .
660|
640|
620|
600| T                                                             *
580| h
560| o
540| u
520| s                                                           *
500| a
480| n                                                         .
460| d
440| s                                                       .
420|
400| o
380| f                                                     .
360|                                                     *
340| H                                                 .
320| o
300| s                                             *
280| t                                           .
260| s                                         .
240|                                          .
220|                                        .
200|                                      .
180|                                    .
160|
140|                                  *
120|                                *
100|                             ..
 80|                           *
 60|                         .
 40|                       *
 20|                 ..*...*
  0|...*....*......*.....*.....*.*....*...
   -------------------------------------------------------------------
     8     8     8     8     8     8     8     8     8     9     9     9
     1     2     3     4     5     6     7     8     9     0     1     2
                              Date
    "*"  = data point,  "." = estimate
This graph is a linear plot of the number of Internet hosts.
```

                    Number of Internet Hosts (logarithmic)


```
        | 1000000
        |                                                          *.*
        |                                            ..*.*..*
        |                                                 ...
        | 100000                                     ..**
        |                                       ...**
      H |                                  *.*
      o |                              ...*
      s | 10000                     .*
      t |                        ..
      s |                     ..
        |                  ....*
        |             ...*.*
    1000|          ...*..
        |        ...
        |     ...*
        | ..*....*...
     100|.
        -----------------------------------------------------------------
         8       8       8       8       8       8       8       8       8       9       9       9
         1       2       3       4       5       6       7       8       9       0       1       2
                                        Date
```
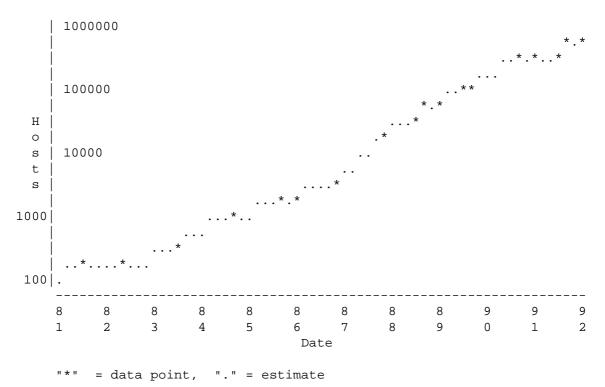
        "*" = data point, "." = estimate

This graph is a logarithmic plot of the number of Internet hosts.

N.2 Number of Domains

   This chart shows the number of domains existing in the Internet
   Domain Name System as collected by ZONE.

```
        Date            Domains

        07/88               900
        10/88             1,280
        01/89             2,600
        07/89             3,900
        10/89             4,800
        10/90             9,300
        01/91            11,200
        07/91            16,000
        10/91            18,000
        01/92            17,000
```

N.3 Distribution of IP Addresses per Host

   This chart shows how many hosts have how many IP addresses. This
   data was collected on 1-Jan-92 and only the first 10 entries are
   shown.

        Addresses   Hosts

            1   715143
            2   9015
            3   1027
            4   556
            5   314
            6   213
            7   100
            8   85
            9   58
           10   71

N.4 Distribution of Hosts by Top-level Domain

   This chart shows the number of hosts per top-level domain (top 40
   only) on 1-Jan-92. The percentage listed is the increase since 1-
   Oct-91. Large variations are probably due to problems and variations
   in the collection process; these figures are not meant to be
   authoritative, but serve as reasonable estimates.

   243020 edu 13%   13011 fr   4%   1791 dk   4%   357 be -5%
   181361 com 12%   12770 nl   21%   1662 es   15%   334 gr 14%
    46463 gov 13%   12647 ch   10%   1506 kr   9%   308 br 26%
    31622 au 19%   11994 fi   15%   1111 nz -16%   284 mx -5%
    31016 de 20%   10228 no   9%   1016 tw   n/a   207 is   0%
    27492 mil 26%   8579 jp   6%   929 za   n/a   146 pl 97%
    27052 ca 22%   4109 net -49%   784 pt   n/a   127 us 25%
    19117 org 10%   3324 at   19%   484 sg 251%   25 tn   0%
    18984 uk 139%   2719 it 197%   448 hk   78%   24 hu 71%
    18473 se 34%   2020 il   14%   374 ie   -7%   6 arpa 0%

N.5 Distribution of Hosts by Host Name

   This chart shows the distribution of hosts by their host name on 1-
   Jan-92. The host name is defined to be the first part of a fully
   qualified domain name. Only the top 100 names are shown.

384 venus	204 mac4	172 mac9	155 pollux	138 chaos
356 pluto	201 hobbes	172 mac11	155 frodo	136 bart
323 mars	201 hermes	170 mac8	153 helios	135 pc5
288 jupiter	198 thor	169 phoenix	152 mac17	135 larry
286 saturn	198 sirius	169 mac12	151 vega	135 cs
285 pc1	196 gw	169 hal	151 mac18	133 odin
282 zeus	195 calvin	168 snoopy	150 falcon	131 tiger
262 iris	194 mac5	168 mac13	150 bach	131 sparky
260 mercury	191 mac10	167 mac15	146 castor	131 ariel
259 mac1	190 fred	167 mac14	145 sol	130 sneezy
258 orion	189 titan	167 grumpy	145 dopey	128 mac
254 mac2	189 pc3	163 gandalf	144 mac20	127 sun1
240 newton	186 opus	162 pc4	144 mac19	127 rocky
234 neptune	186 mac6	160 uranus	142 spock	126 pc6
233 pc2	185 charon	159 mac16	142 euler	125 hydra
224 gauss	185 apollo	158 sleepy	141 mickey	125 homer
222 eagle	179 mac7	158 io	141 atlas	124 isis
213 mac3	179 athena	157 earth	140 maxwell	123 moe
209 merlin	177 alpha	156 europa	140 happy	123 delta
207 cisco	172 mozart	155 rigel	140 doc	122 pc10

Future Issues

   ZONE currently runs on a DECsystem-20 and is written in assembler.
   The amount of data is quickly reaching the limits of the DEC-20
   section address space, and the hardware's ability to survive gets
   slimmer each day. ZONE assembles all its data in core before dumping
   it to disk. The implementation does this in order to be able to
   match host nicknames with official names before dumping complete host
   records. Sometimes a nickname can be in a different domain than the
   official name, complicating simpler methods.

   A new version of ZONE needs to be written to run on a modern computer
   system. A completely new architecture should be designed to handle
   the enormous amount of data collected and expected in the future.
   Data should be kept on disk so that a system crash will not wipe out
   days of collection. Multiple zone transfers could be occurring in
   parallel to reduce the time needed for data gathering. A new ZONE
   might run continuously, cycling through the domain system on a cycle
   lasting weeks to a month, updating a local database with statistics
   collected for each domain. In this way, current statistics on the
   size of the Internet would always be known. The resulting database

   may also be useful for other network information services.

RFC References

   Libes, D., "Choosing a Name for Your Computer", RFC 1178, Integrated
   Systems Group/NIST, August 1990. (Also FYI 5.)

   Mockapetris, P., "Domain Names - Implementation and Specification",
   RFC 1035, USC/Information Sciences Institute, November 1987.

   Mockapetris, P., "Domain names - Concepts and Facilities", RFC 1034,
   USC/Information Sciences Institute, November 1987.

   Lazear, W., "MILNET Name Domain Transition", RFC 1031, Mitre,
   November 1987.

   Harrenstien, K. Stahl, M., and J. Feinler, "DoD Internet Host Table
   Specification", SRI, October 1985.

   Postel, J., "Domain Name System Implementation Schedule - Revised",
   RFC 921, USC/Information Sciences Institute, October 1984.

Security Considerations

   Security issues are not discussed in this memo.

Author's Address

   Mark K. Lottor
   SRI International
   Network Information Systems Center
   333 Ravenswood Avenue, EJ282
   Menlo Park, CA 94025

   EMail: mkl@nisc.sri.com