SLAC-PUB-7961
October 1998

Internet Monitoring in the HEP Community

Warren Matthews, Les Cottrell

Stanford Linear Accelerator Center, Stanford University,
Stanford, CA 94309, USA

David Martin

HEP Network Resource Center, Fermi National Accelerator Laboratory,
Batavia, IL 60510, USA

Presented at International Conference on Computing in High Energy Physics 1998
(CHEP 98), 8/31/98—9/4/98, Chicago, IL, USA

Stanford Linear Accelerator Center, Stanford University, Stanford, CA 94309

Work supported by Department of Energy contract DE–AC03–76SF00515.

Internet Monitoring in the HEP Community. *

Warren Matthews, Les Cottrell
Stanford Linear Accelerator Center, Stanford, CA 94309, USA.
David Martin
HEP Network Resource Center, Fermi National Accelerator Laboratory, Batavia, IL 60510, USA.

October 6, 1998

Abstract

The HEP Internet Monitoring Project (known as *PingER* [1]) attempts to measure the performance of the Internet used by the High Energy Physics Research community and provide an accurate measurement of the end-to-end performance individuals may expect by monitoring the performance between a given monitoring node and remote node pair. This is achieved by monitoring the packet loss and round trip time (RTT) of ICMP *Ping* packets from 17 monitoring sites around the world to 373 nodes at 267 sites between 1084 monitoring-host-remote-site pairs, involving 27 countries.

This paper details the current work and status of this on-going project. The architecture, methodology and nature of the problem will be reviewed, some trends will be discussed, and the direction of further work will be outlined.

Keywords: Wide Area Networking; Network; Monitoring; End-to-end; Performance.

1 Introduction

The distributed nature of modern High Energy Physics research and the staggering processing and storage requirements of future accelerators along with the huge growth and commercialization of the Internet in recent years, have provided many physicists with cause for concern. HEP traffic must traverse the same congested paths and queue at the same congested routers as all other traffic. Monitoring end-to-end performance between monitoring sites and remote nodes provides vital information on the state of our connectivity and the feasibility of the HEP community's future plans for networking.

The *PingER* (Ping End-to-end Reporting) project measures round trip loss and delay by regular *pinging* of remote nodes by monitoring nodes. The monitoring involves sending *ping* packets from 17 monitoring sites in 10 countries across a total of 1084 end-to-end pairs of nodes to 373 remote nodes at 267 sites in 27 countries and logging the results for further analysis. The *PingER* project is thought to be the most extensive end-to-end Internet monitoring project in existence.

Ping is an ICMP (Internet Control Message Protocol) echo request [2], and it is part of the TCP/IP implementation and runs in the kernel. There is no need to install other applications and it is currently our best tool to indicate the true performance of the network, independent of the load on the machine that is running it.

The use of *ping* to measure End-to-end network performance of the network can be justified because the results from *ping* data correlate well with real network applications such as HTTP and FTP transfers [3].

Studies show a small amount of packets are lost

*Presented at Computing in High Energy Physics, August 1998.

Packets Lost (%)	Rank
0-1	Good
1-2.5	Acceptable
2.5-5	Poor
5-12	Very Poor
>12	Unusable

Table 1: Definitions of Quality in Packet Loss.

Response Time (ms)	Rank
0-62.5	Good
62.5-150	Acceptable
150-250	Poor
250-500	Very Poor
>500	Unusable

Table 2: Definitions of Quality in Response Time.

due to transmission errors, approximately 1 in 5000 packets arrive corrupted [4], but by far the main cause of packet loss is congestion in routers. Packets must queue to be processed, and if the queue is full, the packet is discarded[1]. Polite protocols such as TCP back-off if this happens, and slow down the rate of transmission. If the originating host does not receive an acknowledgment that the packet was received it is re-sent. Table 1 shows the thresholds for the quality of the packet loss defined by the authors for use in this analysis. The values are determined from examining the performance of applications. The thresholds are different from previous descriptions of this project [1], and are now stricter to provide a clearer guideline for interactive applications.

Ideally, traffic should traverse the Internet at the maximum speed for the medium (e.g. the speed of light in glass for fiber). However, connections very rarely do, the main reason being queuing at routers. The major effect of poor response time is on interactive sessions such as telnet, or packetized video or voice, where even fairly moderate delay can cause severe disruption, this is reflected in the times shown in Table 2 for the thresholds of the quality of the response time defined by the authors for use in this analysis.

There are 7 monitoring sites in the United States; The Stanford Linear Accelerator Center (SLAC), the HEP Network Resource Center (HEPNRC) at Fermi Lab, the Department of Energy (DOE) in Washington, Brookhaven National Laboratory (BNL), Atmospheric Radiation Measurement (ARM) Program, Carnegie Mellon University (CMU) and University of Maryland (UMD). There are 2 further sites in North America in Canada at the TRIUMF facility near Vancouver and at Carleton University in Ottawa. There are also 6 sites in Europe, at the European Center for Particle Physics (CERN), the Deutsches Elektronen Synchrotron (DESY), the Rutherford Appleton Laboratory (RAL) near Oxford in England, the Niels Bohr Institute (NBI) in Denmark, INFN's national Center for Telematics and Informatics (CNAF) in Italy, and the Research Institute for Particle and Nuclear Physics (KFKI) in Budapest, Hungary. Finally, there are 2 monitoring sites in Asia at Sinica in Taiwan and the KEK facility in Japan.

The monitoring sites are largely concentrated in North America and West Europe, but geographical location is largely irrelevent in networking, the important factor is network connectivity. The authors believe the distribution of monitoring sites and remote nodes and the large number of networks connecting them is representative of the Internet used by HEP.

The SLAC, HEPNRC, DOE, BNL and ARM monitoring sites are connected to the Energy Sciences network (ESnet), which connects all the Department of Energy funded laboratories in the United States. At the time of writing, over 70 US Universities are connected to the very-high-performance Backbone Network Service (vBNS). However, only two of the ¡em¿PingER¡/em¿ monitoring sites (CMU and UMD) are connected to the vBNS, and 27 of the remote (monitored) sites are connected via the

[1] The situation is complicated by router algorithms such as Random Early Drop (RED) which discards packets even if the buffer is not full. However, such algorithms are not widely deployed in routers and the above is sufficient to understand packet loss in general.

vBNS. Other US Universities are connected via local Internet Service Providers (ISPs), we shall refer to these Universities as "Others". In Canada TRIUMF is connected to BC-Net and Carleton University is connected via ON-Net, and in turn BC-Net and ON-Net are connected to the Canadian Canarie backbone. Networking in Europe is characterized by National Research Networks (NRNs) such as the Joint Academic Network (JANet) in the UK and RedIRIS in Spain. The NRNs are frequently interconnected with the TEN-34 (soon to be upgraded to TEN-155) backbone. The KEK facility in Japan is connected via NACSIS.

Another critical factor is how these networks are interconnected. Many networks have direct peering relationships. This will be discussed futher in the section on Results.

Previously, all monitoring sites have been free to *ping* sites of interest to themselves. Recently however, the concept of beacon sites has been introduced, and all monitoring sites have been requested to *ping* them. Beacon sites represent the various affinity groups monitored. All monitored sites, but especially beacon sites should be reliable (24-hours per day, 7-days per week), lightly loaded (or at least consistently loaded) and responsive to *pings*. Other factors determing the selection of a beacon site include its physical location, backbone connectivity and its importance to HEP in general. 50 beacon sites have been selected. Monitoring the beacons sites from all 17 monitoring sites gives better information for trouble shooting and understanding the network in general.

2 Data Gathering

Each monitoring site *pings* the remote sites on average every half an hour and the results of each sample are written to a file. Each day the archive site at HEPNRC retrieves the data and stores it in a SAS database. Currently, roughly 600 MBytes of data are stored per month.

3 Analysis Method and Tools

Analysis is done on the data by each monitoring site to show short term reports in the form of a table with the packet loss and response time of the latest data. Furthermore each site provides the raw data and a simple configurable graphing tool. Also, a collective analysis is done on all the data at the analysis site at SLAC, providing detailed hourly, daily and monthly reports.

All the data and analysis is available on-line via several WWW CGI front ends. Tables of daily reports with hourly ticks, monthly reports with daily ticks and summary reports with monthly ticks for all the pairs of monitoring-remote nodes are available with color-coded values as defined in table 1 and table 2. These tables can be sorted by any month, and the links can be filtered to display a smaller number selected by site, top-level domain, geographical location or a number of user-defined groups such as backbone provider, for example ESnet (which are pairs of monitoring-remote nodes involving sites on the ESnet backbone), or BABAR (which is a group of pairs of monitoring-remote nodes involving members of the BABAR collaboration, a collaboration at SLAC). Graphs are generated by SAS and *perl*, and the tables also provide output in tab-separated-variables (tsv) format to allow the user to import the data into a spreadsheet.

4 Results

Many pairs of monitoring-remote nodes improved in the summer of 1998. Sites or networks increased their bandwidth, changed their peering arrangements, or moved to a new ISP. In addition, it should be noted that the overall performance of the network improves dramatically during the main University holidays at Christmas, Easter and during the summer. Figure 1 shows packet loss between three sites on ESnet (SLAC, HEPNRC and BNL) to sites in the UK. Improvement in performance can be seen when the Bandwidth across the Atlantic was improved, but also significant improvement can be seen each year in December and August.

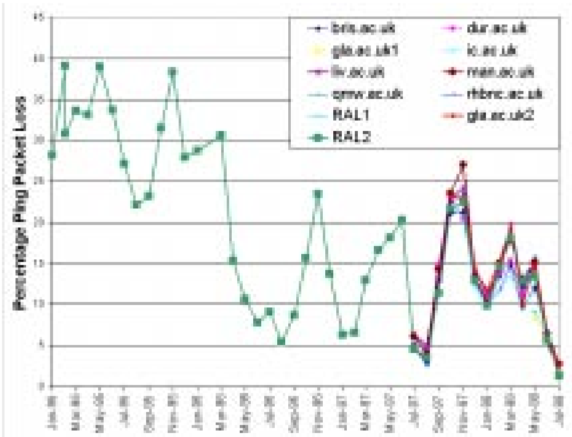

Figure 1: Performance between Monitoring Sites on ESnet and sites in the UK. Note the drop in packet loss (i.e. improved performance) in August'95, April'96 and July'97 corresponding to increased bandwidth on each occassion. Also note similar improvements in August and December corresponding to University holidays.

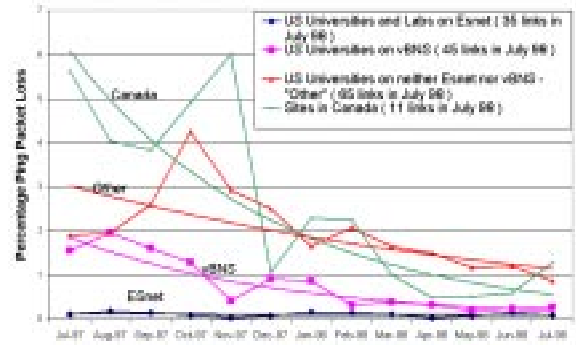

Figure 2: Performance on Links between Sites in North America with exponential fits. Note the steadily decreasing percentage of packets lost, i.e. improved performance, for all groups.

4.1 Performance within a Geographical Region

Figure 2 shows the median packet loss per month for sites in North America. Performance to sites on ESnet is good, with packet loss typically less than 0.5% and round trip times within a factor of 2 to that expected if the entire journey were to take place at the speed of light in glass. Sites on the vBNS are also good, with packet loss typically less than 1%. ESnet peers with vBNS in three locations, STARTAP in Chicago, Perryman in Washington and at the San Diego Supercomputing Center (SDSC). Traffic from SLAC to vBNS usually travels via Chicago. Performance to US Universities on neither ESnet nor vBNS and performance to sites in Canada are very variable. Connectivity to some sites, such as utk.edu (University of Kentucky) is good because it is connected to Oak Ridge National Laboratory which is an ESnet site and ESnet routing policy dictates traffic remains on ESnet for as long as possible, even if the ultimate destination is not on ESnet. Other sites are connected via commercial providers, and can suffer high packet loss and poor response times. The situation for all North American academic sites should improve dramatically over the next year or so, with the introduction of new high speed links as more Universities are connected to Internet II.

Performance within most of West Europe is good, although packet loss to Spain is significantly higher than other parts, and the trend is increasing, although recent data to the Spanish beacon sites indicate the situation may be improving. Links to East Europe are frequently saturated, leading to large packet loss and poor response time.

Performance inside Japan is good, but it is difficult to compare performance based in the geographic region of Asia because the connectivity is not geographically based. Traffic from KEK to Taiwan for example, travels via MCI's San Francisco hub.

4.2 Transoceanic Performance

Performance across the oceans is a common bottleneck for traffic. Performance between the U.S. and the U.K. has always been very variable, improving during University holidays and briefly when increased bandwidth is installed, but quickly returning to near saturation. This is shown quite clearly in Figure 1.

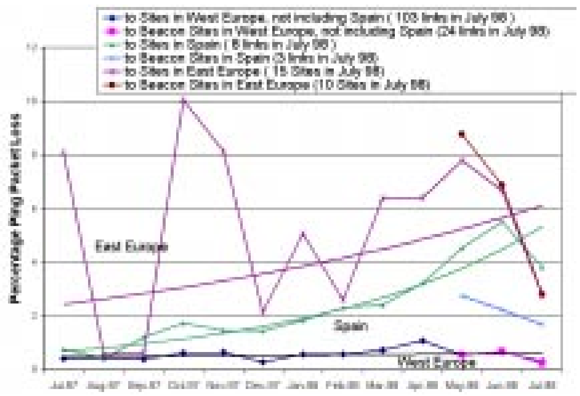

Figure 3: Performance on Links between sites in Europe with exponential fits. Performance within East Europe and Spain is Very Poor.

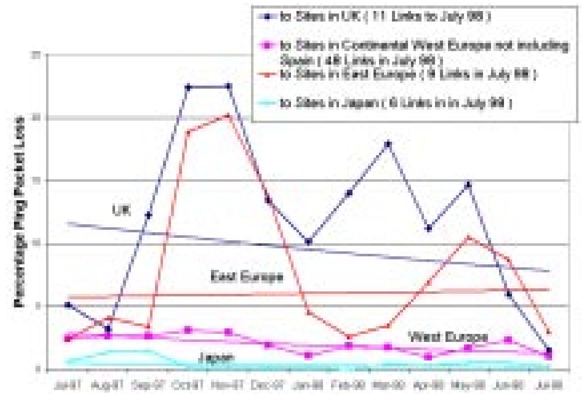

Figure 4: Transoceanic Performance from Sites on ESnet with exponential fits.

The U.K. Research Network (JANet) increased bandwidth again in May 1998 to 90Mb/s. Overall the trend is 1-2% improvement per month. Performance to the rest of West Europe (Spain was not included) is much better and improving. However the performance to East Europe is very variable, and the overall trend is deteriorating. The creation of TEN-155 is expected to improve the situation in both West and East Europe. Performance to Japan is good.

DFN, the German national Academic and Research (A&R) network, which provides connectivity for DESY, peers with ESnet at Perryman, a network exchange point in Washington, and CERN has a pipe to ESnet also peering in Washington. Hence the packet loss is low and the overall transoceanic connectivity to ESnet is good. DFN peers with vBNS at MAE-East, a heavily loaded public exchange point. CERN peers with vBNS at MCI's hub in Dallas, which is better than a public exchange point, but it is still the commercial Internet. Consequently, performance to vBNS is not as good as to ESnet.

Transoceanic performance seen from KEK has been somewhat variable, although lack of statistics makes it difficult to look at long term trends. KEK is connected via NACSIS and has a dedicated 512kbps line to ESnet. Performance to Europe has improved dramatically in recent months because of a new peer-

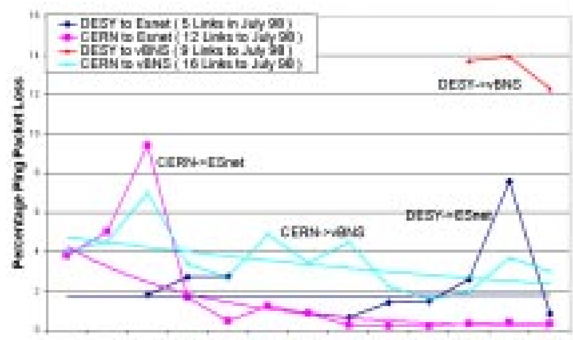

Figure 5: Transoceanic Performance from Sites in Europe.

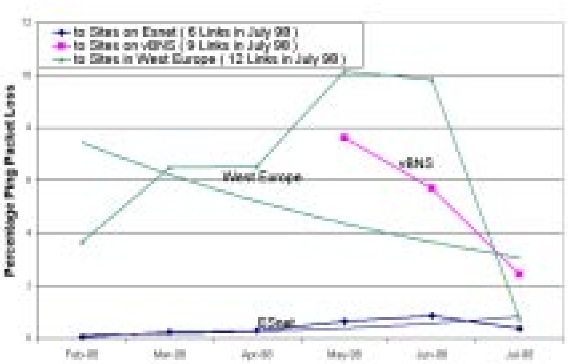

Figure 6: Transoceanic Performance from Sites in Asia.

ing arrangement with the European TEN-34 network.

5 Conclusion

Performance between nodes on the same backbone is usually good. Performance between U.S. laboratories connected via the ESnet backbone is excellent. U.S. Universities connected via the vBNS backbone performs equally as well, and performance between ESnet and vBNS is also very good. Performance between some U.S. sites on neither ESnet nor vBNS and between some U.S. and Canadian sites are poor or very poor. In West Europe and Asia, performance between most end-to-end pairs is good or acceptable, but in other regions some hops are often saturated resulting in poor performance. In general performance on Academic and Research networks is good, performance on commercial networks is bad.

Monitoring-site-remote-node pairs that involve routes that cross several networks, particularly those that involve transoceanic connectivity often perform less than good. Often resources are concentrated on links to the U.S. although recent changes have improved the situation but links to poorer countries and remote regions are understandably poor on many occassions.

Overall, performance between the nodes monitored is improving at the rate of a few percent per month.

6 Future Work

The nature of the Internet is changing. The advent of new applications and protocols, Quality of Service (QoS), traffic shaping, active networks and many other things inevitably affects the accuracy of network monitoring with *ping*. Even in the HEP community, the distributed nature of the work means *PingER* will evolve. Current plans include monitoring using packets other than *ping* and varying the distribution of the probes. Futhermore, SLAC is involved in a number of projects that have recently been developed, for example the *Surveyor* project. The *Surveyor* system measures one-way delay and one-way packet loss between nodes at over 25 U.S. Internet sites. The *Surveyor* machine at each site includes a global-positioning system (GPS) receiver. Packets are sent in both directions between each of the *Surveyor* machines, and the data is averaged over a one minute interval. Significant effort will be given to comparing results obtained from *Surveyor* with that obtained by *PingER*, and understanding any discrepencies.

Also, work on discovering network congestion points using a NetMap, that is a Network snapshot made using the traceroute tool, is under developement. The maps may indicate individual nodes, or just the Autonomous System (AS). It is expected that will help identify common bottlenecks afflicting several routes and resources can be allocated to resolve the issue.

7 Further Information

Please visit our page at http://www.slac.stanford.edu/comp/net/wan-mon.html for more information on the *PingER* project, links to related monitoring efforts and information about Internetworking in general.

8 Acknowledgements

This is a distributed development effort involving multiple sites and many people. The authors would like to thank the following developers and contributors who have been instrumental in the evolution of the *PingER* project: Bruce Anderson, John Halperin, Connie Logg, Lois White, Andrew Daviel, Franca Fiumana, Charles Granieri, John MacAllister, David Williams and Harvey Newman. Also, all the other members of the NETMON mailing list[2] and the contacts who have installed software, and are maintaining monitoring sites and providing feedback. Finally, all the members of the XIWT/IPWT, ICFA-NTF, ESSC, and the ESCC.

References

[1] Cottrell, R.L.A., Logg, C.A., and Martin, D.E. "What is the Internet Doing? Performance and Reliability Monitoring for the HEP Community" Computer Physics Communications, vol.110 p.142-148, May 1998.

[2] Postel, J., "Internet Control Message Protocol" RFC792, ftp://ftp.isi.edu/in-notes/rfc792.txt September 1981.

[3] Cottrell, R.L.A., and Halperin, J., "Effects of Internet Performance on Web Response Times" Unpublished, http://www.slac.stanford.edu/comp/net/wan-mon/ping/correlation.html January 1997.

[4] Paxson, V., "End-to-End Internet Packet Dynamics" Proceedings of SIGCOMM'97, September 1997.

[2] netmon@hepnrc.hep.net

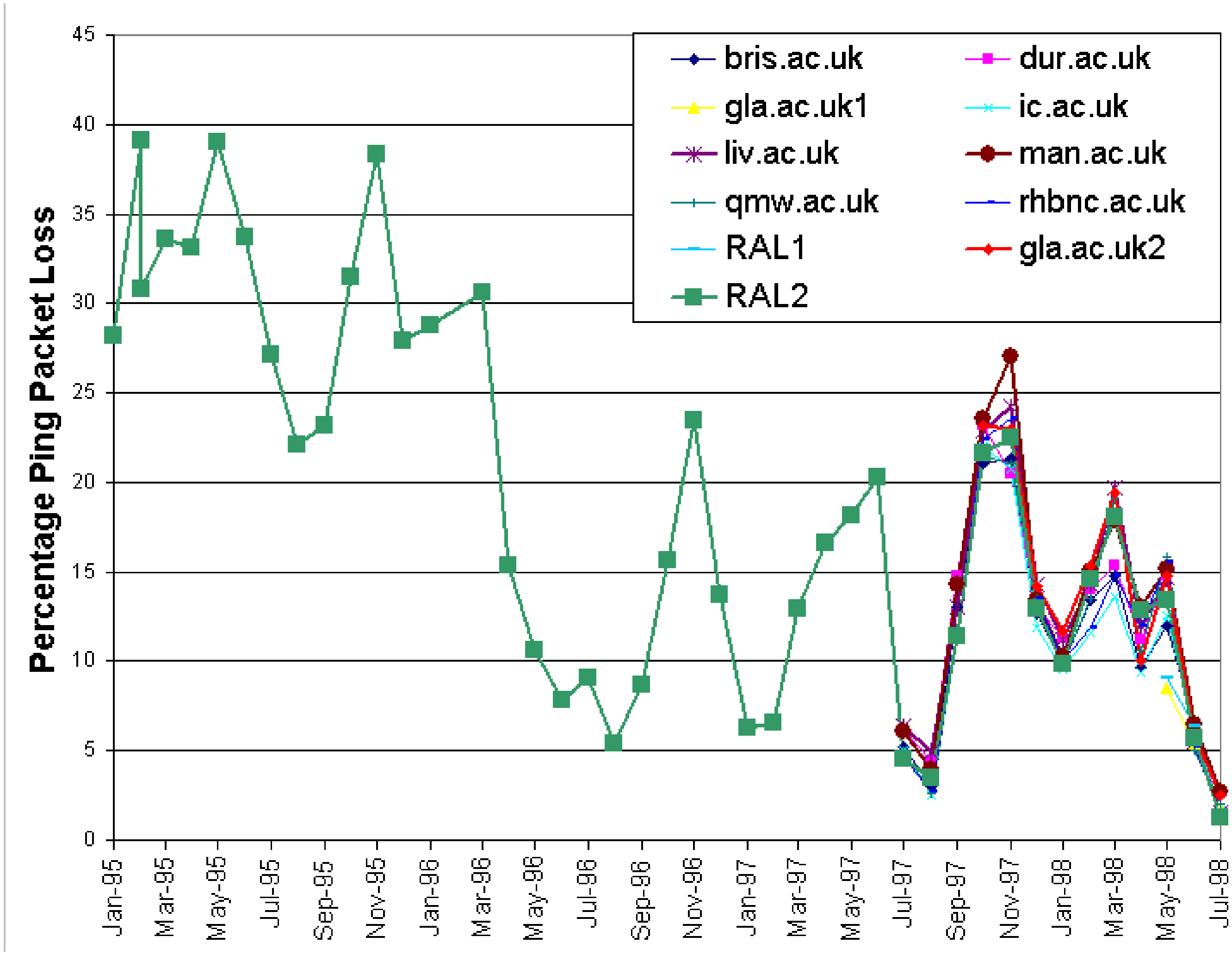

Figure 1: Performance between Monitoring Sites on ESnet and sites in the UK. Note the drop in packet loss (i.e., improved performance) in August '95, April '96 and July '97 corresponding to increased bandwidth on each occasion. Also note similar improvements in August and December corresponding to University holidays.

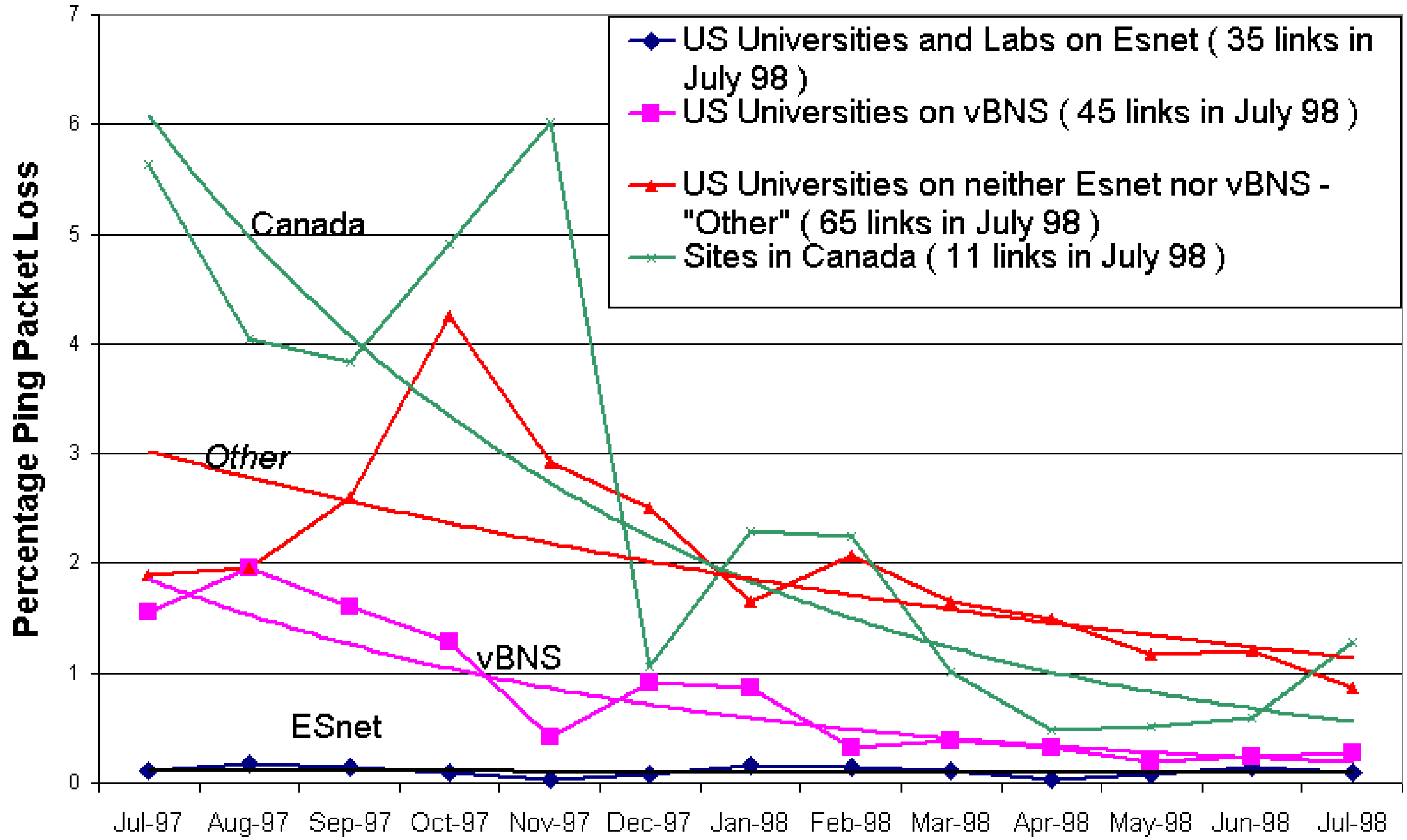

Figure 2: Performance on Links between Sites in North America with exponential fits. Note the steadily decreasing percentage of packets lost, i.e., improved performance, for all groups.

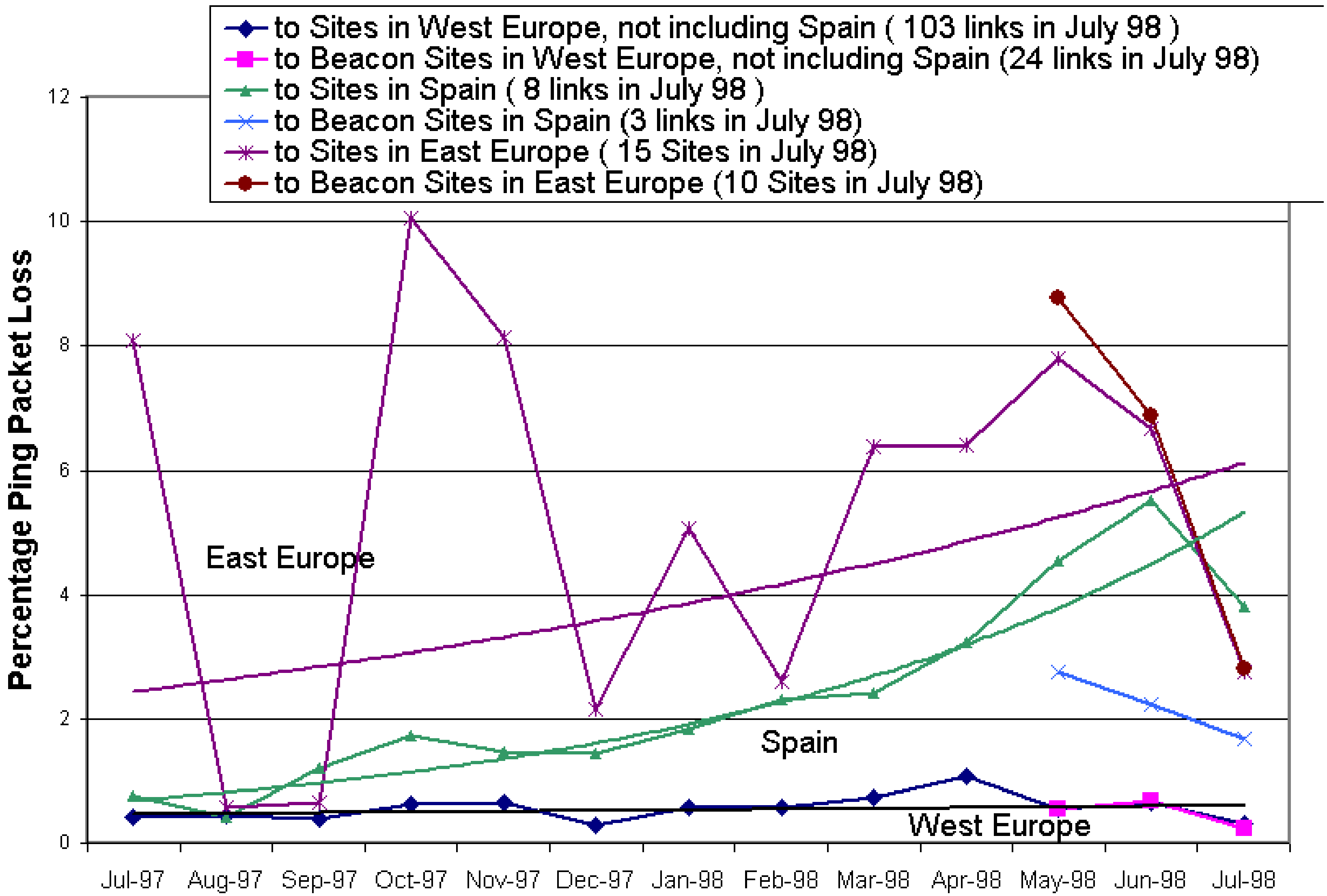

Figure 3: Performance on Links between sites in Europe with exponential fits. Performance within East Europe and Spain is Very Poor.

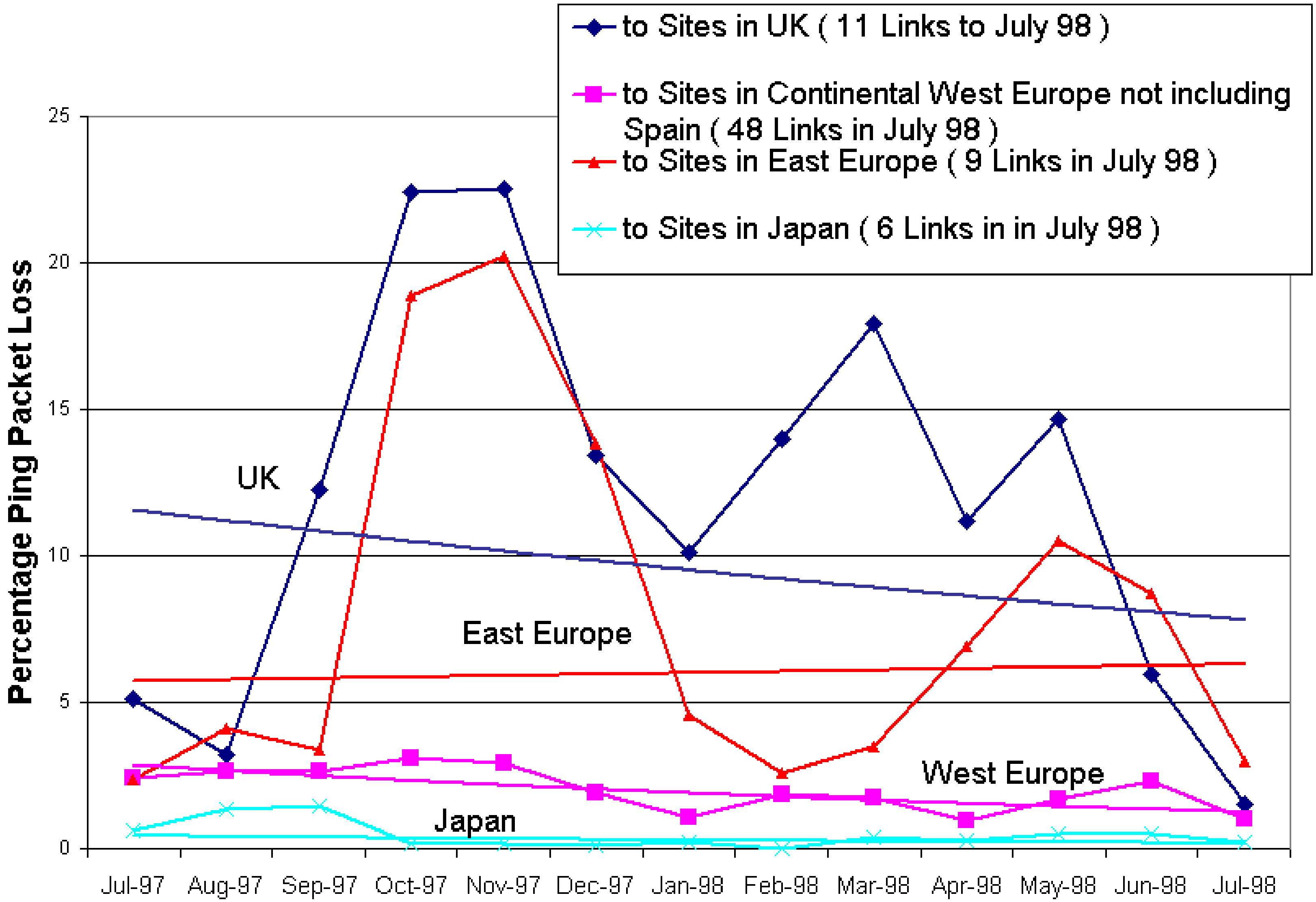

Figure 4: Transoceanic Performance from Sites on ESnet with exponential fits.

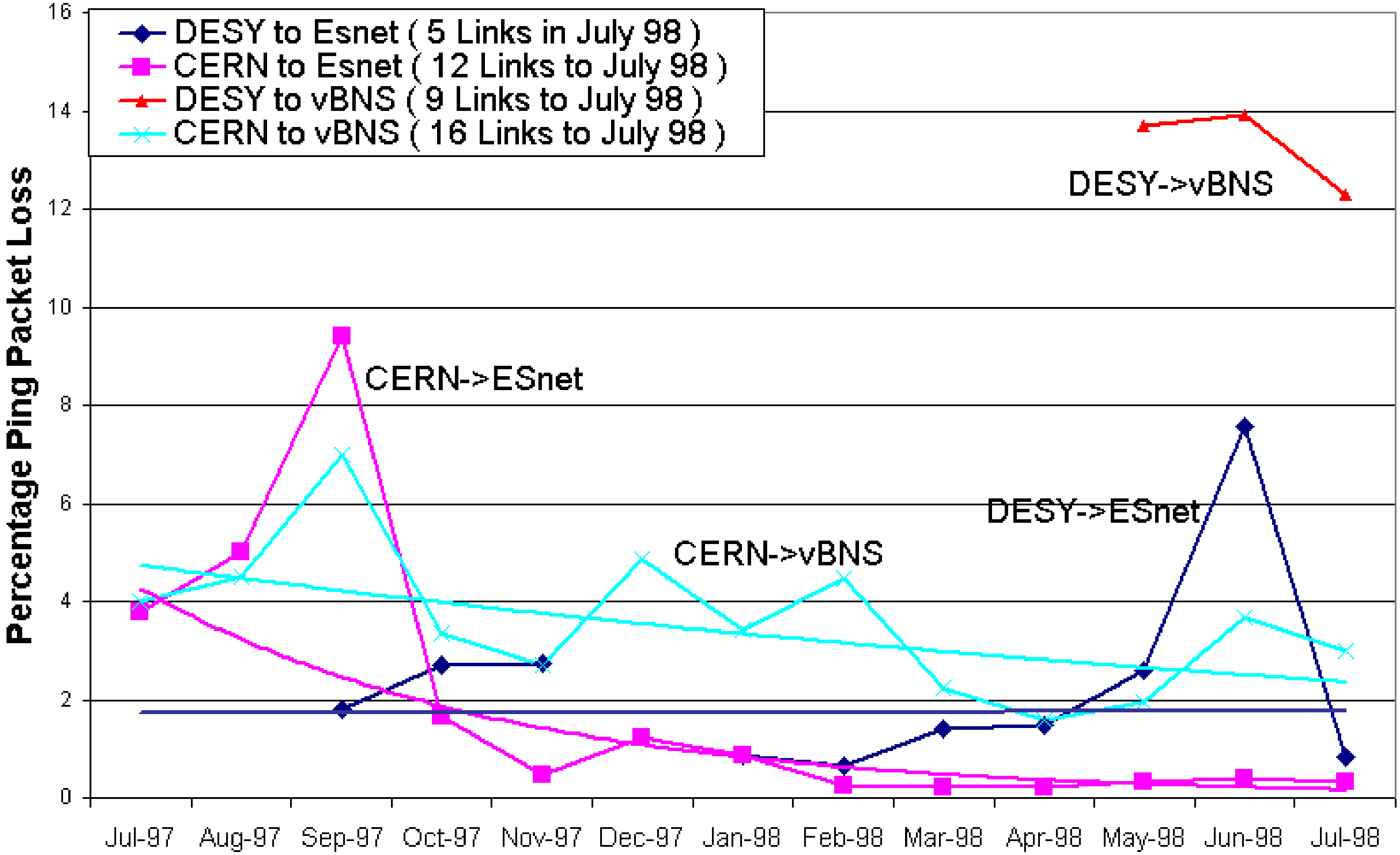

Figure 5: Transoceanic Performance from Sites in Europe.

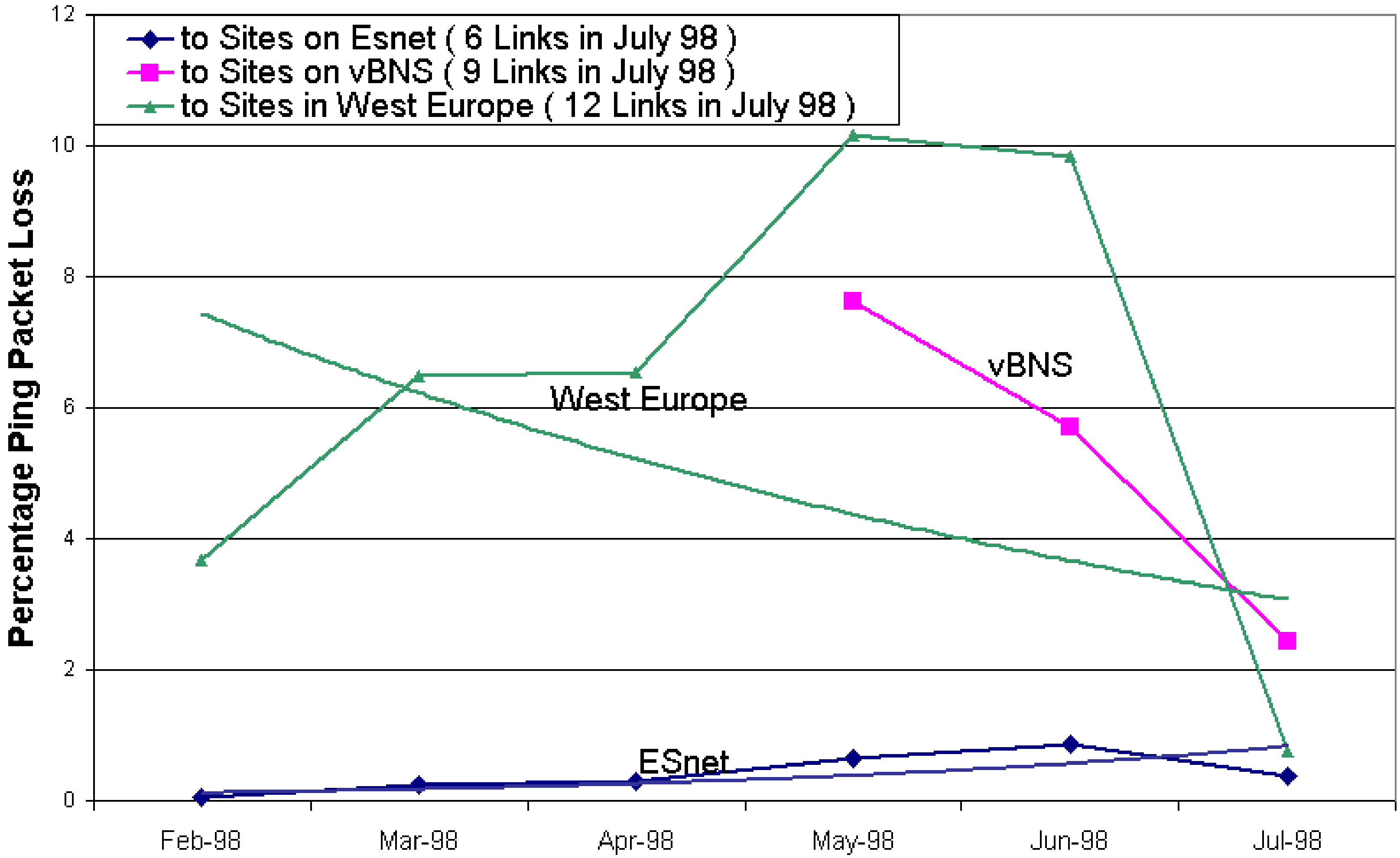

Figure 6: Transoceanic Performance from Sites in Asia.