

Experiences Using Web100 for End-to-End Network Performance Tuning for Visible Human Testbeds

Thomas J. Hacker
Center for Advanced
Computing
University of Michigan

Brian D. Athey
Michigan Center for
Biological Information
University of Michigan

Jason Sommerfield
Pittsburgh
Supercomputing Center
Carnegie Mellon
University

Deborah S. Walker
School of Nursing
University of Michigan

Email: {hacker, bleu, dswalker}@umich.edu, jasons@psc.edu

Abstract

A major source of performance degradation in high-performance distributed applications has been attributed to poor end-to-end Transport Control Protocol (TCP) performance. The root causes of poor TCP performance are difficult to isolate and diagnose, and the efficacy of tuning efforts are often difficult to gauge. This paper describes sources of poor TCP performance, and suggests a method to diagnose some of these problems based on a combination of existing performance tools and the Web100 tuning package. Using this methodology, the TCP performance of an application developed for the Visible Human Project is shown to significantly improve, making it practical for full resolution Visible Human data to be used interactively by multiple users in applications testbeds.

1.0 Introduction

Distributed application performance problems traceable to poor TCP performance has been identified as a major source of performance degradation in high-performance applications [1]. Appropriately provisioned network infrastructure is essential for providing support for high-performance networking. However, lack of proper host tuning and unexpected levels of packet loss can adversely affect actual end-to-end network performance to such an extent that it can nullify the benefits of network infrastructure investments. Applications can greatly benefit from application and host tuning efforts targeted at improving aggregate network performance.

Measurements of the Edgewarp application [2] written for the Visible Human Project (VHP) at the University of Michigan (UM) [3] have shown that the effects of poor host and application tuning can seriously degrade bulk transport performance necessary for delivering images by at least a factor of four. One of the project's goals is to support the simultaneous access of content through the Internet by at least 40 teaching stations for each classroom session.

1.1 Actual TCP Bandwidth Delivered to the Application

1.2 Uncooperative Network Application Behaviors

1.3 Tuning Methodology and Other Sources of Poor Network Performance

These sections are in the full version of this paper.

1.4 Web100

Web100 [12] has been used with great success to identify and diagnose the symptoms and causes of network performance problems and for immediately measuring the effects of performance tuning. It is hoped that the work described in this paper will be useful to application developers and system administrators for tuning their host systems and improving network and application performance.

2.0 Related Work

2.1 General Tools

2.2 Network Specialist's Tools

These sections are in the full version of this paper.

3.0 Web100 and the Visible Human Project

The UM VHP is a data and visualization intensive grid-computing project designed to deliver pedagogical content to students in the health professions at universities and medical centers. The Edgewarp [2, 3, 24, 32] application was developed in conjunction with the VHP to distribute volumetric three-dimensional rendered human anatomy images and image sequences. The anatomical content is delivered as a sequence of landmark-based image voxels from an Edgewarp server to a set of clients. To improve the performance of Edgewarp, along with other applications developed to support the VHP, Web100 was used with a toolset consisting of ping, traceroute, pipechar, pchar, and lperf to improve the network performance of VHP data servers and applications [25]. The features of Web100 that proved to be most useful for tuning were the real-time measurements of data bytes transmitted, packet retransmission, receiver TCP window size, and the display of TCP options. As the result of tuning, pedagogically valuable data that was previously unusable due to poor performance (less than 1 Mb/sec) can now be used to the fullest extent by students in the health professions.

4.0 Using Web100 to Tune End-to-End Performance

To test the Edgewarp server’s ability to deliver voxels to a client application, a test rig that connects to the Pittsburgh Supercomputing Center (PSC) Visible Human voxel server was developed by the VHP development group to simulate the retrieval characteristics of the Edgewarp browser. The results from the test rig were analyzed in combination with the use of ping, traceroute [16], pipechar [14], pchar [13], and Iperf [15] to determine if there were any performance bottlenecks in the network path between the voxel server at PSC and the Edgewarp client at U-M. Once the structural properties of the network path were determined, the ability of the Edgewarp server to deliver data from PSC to U-M was tested. Web100 was used on the server at PSC to monitor in real-time the characteristics of the TCP transfer.

4.1 Characteristics of a Mistuned TCP Receiver

Web100 was used to investigate the TCP connection properties for the session. Examination revealed that several TCP options were disabled. First, Selective Acknowledgement (SACK) [26], which is critical for good TCP performance in networks with packet loss, was disabled. Second, the Maximum Segment Size (MSS) was very small. Finally, timestamps and window scaling options were not used. These options are described in RFC1323 [27] and are critical for high-speed TCP connections. Further examination revealed that there was no packet loss, and that the congestion window size (CurrentCwnd) was very large when compared with the receiver window size. Thus, the number of outstanding segments was limited by the TCP receiver window size, not the maximum capacity of the network. Consequently, TCP performance was limited by an inappropriately sized TCP receiver window. The full version of this paper gives a complete description of the diagnosis methodology.

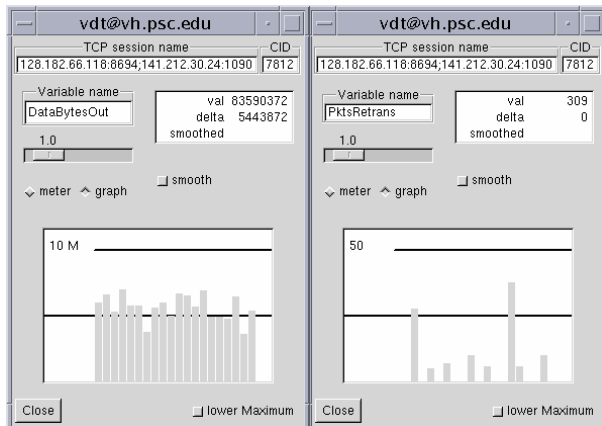


Figure 1. TCP Transfer Rate and Packet Loss Rate on Tuned Host as Measured by Web100

4.2 Tuning the Network and Hosts

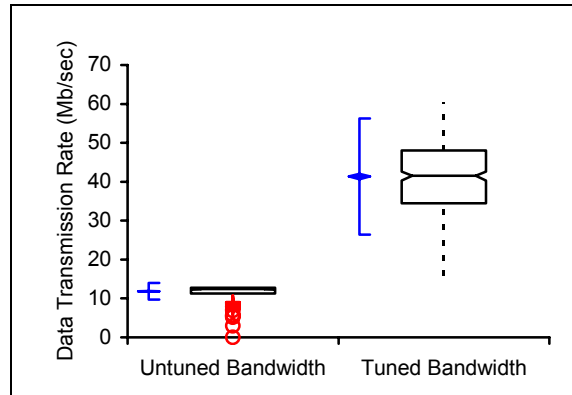
After the Physical and Data Link characteristics of the local network infrastructure were validated, the host tuning problems were addressed. The client host was tuned to support SACK, MTU discovery, Timestamps, and Window Scaling.

The TCP maximum and default send and receive socket buffer, which is used by TCP to determine the receiver window size, was set to 2 MB.

Figure 1 shows the effects of host tuning on the data transmission rate and on packet loss. The data transmission rate increased from 1.5 MB/sec to 5 MB/sec. Packet loss also occurred, which indicates that the network experienced either congestion or systemic packet loss.

If the full version of this paper, we found that the structural bottleneck in the network path between the server and client was a 100 Mb/sec link. The reasonable maximum bandwidth that could be achieved on this link was approximately 80 Mb/sec.

Figure 2 shows a series of 601 measurements of the Edgewarp test rig prior to and after host tuning. The results show that changing a few TCP tuning parameters, increased performance by a factor of four.



	<i>n</i>	SD	95% CI of Mean	Median
Mistuned Bandwidth	601	1.3121	11.728 to 11.938	12.395
Tuned Bandwidth	601	9.0751	40.613 to 42.067	41.578

Figure 2. Effects of Host Tuning on Edgewarp Test Rig Performance

5.0 Conclusions and Future Work

This paper provides evidence that Web100 can be effectively used in combination with network tuning and that the network performance tools currently available to identify structural and host tuning problems can adversely effect end-to-end TCP performance. The use of Web100 along with the other tools mentioned

Extended Abstract

will be critical in assessing the impact of these tuning efforts, and ultimately will increase the pedagogical value of the database for anatomy education.

Acknowledgements

The authors thank Anjana Kar at PSC and the Web100 development team who provided help in the building and installation of the Web100 software; Bill Green and Art Wetzel of PSC who developed the Edgewarp client, voxel server, and test rig; and finally, Matt Mathis at PSC who provided tutoring on TCP network tuning and the use of Web100. This work was supported in part by the National Institutes of Health Visible Human Grant NO1-LM-0-3511.

REFERENCES

- [1] R. Hobby, Internet2 End-to-End Performance Initiative. <http://www.internet2.edu/e2eperf/papers/End-to-End-Perf-Design-Paper.pdf>.
- [2] F. L. Bookstein, W. D. K. Green, "Edgewarp 3D: A Preliminary Manual", Posted to the Internet as <ftp://brainmap.med.umich.edu/pub/edgewarp3.1/manual.html>, 1998.
- [3] University of Michigan Visible Human Project. <http://vhp.med.umich.edu>.
- [4] V. Paxson, M. Allman, S. Dawson, W. Fenner, J. Griner, I. Heavens, K. Lahey, J. Semke, B. Volz, Known TCP Implementation Problems, RFC 2525, Informational, March 1999. URL <ftp://ftp.isi.edu/in-notes/rfc2525.txt>.
- [5] B. Tierney, "Information on critical Linux TCP bug for high-speed WAN applications." <http://www.didc.lbl.gov/Linux-tcp-bug.html>. December 2000.
- [6] V. Jacobson, "Congestion Avoidance and Control." Proceedings of ACM SIGCOMM '88. <ftp://ftp.ee.lbl.gov/papers/congavoid.ps.Z>.
- [7] M. Mathis, J. Semke, J. Mahdavi, T. Ott, "The Macroscopic Behavior of the TCP Congestion Avoidance Algorithm." Computer Communication Review, volume 27, number3, July 1997.
- [8] Marix.net Internet Ratings. <http://ratings.miq.net/>.
- [9] J.F. Kurose, K.W. Ross: Computer Networking: A Top-Down Approach Featuring the Internet, Addison-Wesley, 2001.
- [10] H. Sivakumar, S. Bailey, R. L. Grossman, "PSockets: The Case for Application-level Network Striping for Data Intensive Applications using High Speed Wide Area Networks," Proceedings of Supercomputing 2000, IEEE.
- [11] J. Sorensen, "Alteon AceNIC / 3Com 3C985 / NetGear GA620 Gigabit Ethernet Adapter," <http://jes.home.cern.ch/jes/gige/acenic.html>.
- [12] Web100 Project. <http://www.web100.org>.
- [13] B. Mah, "pchar: A tool for measuring internet path characteristics," <http://www.employees.org/bmah/Software/pchar/>.
- [14] J. Goujun. "Methods for Network Analysis and Troubleshooting," <http://www-didc.lbl.gov/~jin/network/net-tools.html>.
- [15] M. Gates, A. Warshavsky, Iperf version 1.1.1, Bandwidth Testing Tool, NLANR Applications, February 2000.
- [16] V. Jacobson, "Traceroute: A tool for printing the route packets take to a network host," available from <ftp://ftp.ee.lbl.gov/nrg.html>.
- [17] V. Jacobson, C. Leres, S. McCanne, tcpdump, available at <ftp://ftp.ee.lbl.gov/tcpdump.tar.Z>.
- [18] M. Faerman, A. Su, R. Wolski, and F. Berman. "Adaptive Performance Prediction for Distributed Data-Intensive Applications." In Proceedings of Supercomputing 1999. IEEE Computer Society Press, 1999.
- [19] B. Tierney, W. Johnston, B. Crowley, G. Hoo, C. Brooks, D. Gunter, "The NetLogger Methodology for High Performance Distributed Systems Performance Analysis", Proceeding of IEEE High Performance Distributed Computing conference, July 1998, LBNL-42611. <http://www-didc.lbl.gov/NetLogger/>.
- [20] C. Lee, J. Stepanek, R. Wolski, C. Kesselman, I. Foster, "A Network Performance Tool for Grid Environments," in Proceedings of 7th IEEE International Symposium on High Performance Distributed Computing, pp. 260--267, 1998.
- [21] Pittsburgh Supercomputer Center. "About the PSC Treno Server," Available at [http://www.psc.edu/psnoc/treno info.html](http://www.psc.edu/psnoc/treno%20info.html), November 1995.
- [22] NLANR Engineering Services. "A Preconfigured TCP test rig" <http://www.ncne.nlanr.net/research/tcp/testrig/>.
- [23] NLANR Engineering Services. "TCP Trace Based Performance Diagnosis Flowchart" <http://www.ncne.nlanr.net/research/tcp/debugging/>.
- [24] M. J. Ackerman, "The Visible Human Project," J. Biocomm., vol. 18, p 14, 1991.
- [25] PSC News Release. "Web100 Takes First Step Towards Improving Network Performance" March, 2001. <http://www.psc.edu/publicinfo/news/2001/web100-03-19-01.html>.
- [26] M. Mathis, J. Mahdavi, S. Floyd, A. Romanow, "TCP Selective Acknowledgement Options. RFC 2018, Proposed Standard, April 1996," URL <ftp://ftp.isi.edu/in-notes/rfc2018.txt>.
- [27] V. Jacobson, R. Braden, D. Borman, "RFC1323: TCP Extensions for High Performance," May 1992.
- [28] M. S. Borella, D. Swider, S. Uludag, G. Brewster, "Internet Packet Loss: Measurement and Implications for End-to-End QoS," Proceedings,

Extended Abstract

- International Conference on Parallel Processing,
Aug. 1998.
- [29] Braden, B., Clark, D., Crowcroft, J., Davie, B.,
Deering, S., Estrin, D., Floyd, S., Jacobson, V.,
Minshall, G., Partridge, C., Peterson, L.,
Ramakrishnan, K., Shenker, S., Wroclawski, J., and
L. Zhang, "Recommendations on Queue
Management and Congestion Avoidance in the
Internet," RFC 2309, April 1998.
- [30] Intel Corporation. Intel L440GX+ Server Board
Product Guide.
<http://support.intel.com/support/motherboards/server/l440gx/pg.htm>.
- [31] W.R. Stevens. Unix Network Programming
Volume 1: Networking APIs: Sockets and XTI, 2nd
Edition. Englewood Cliffs, New Jersey. Prentice-
Hall 1997.
- [32] B. D. Athey, A. W. Wetzel, and W. D. K. Green.
"Navigating solid medical images by pencils of
sectioning planes," Pp. 63--76 in Mathematical
Modeling, Estimation, and Imaging, eds. D. Wilson,
H. Tagare, F. Bookstein, F. Preteaux, and E.
Dougherty, Proc. SPIE, vol. 4121, 2000.