Guest Editorial
Internet Proxy Services

I. Internet Proxy Services

PROXIES have become a crucial part of the Internet infrastructure during the past few years. In the traditional client/server model, two applications communicate directly with a clear separation of roles between the client and the server. A proxy extends this model by acting as an intermediary between a client and server. Although the notion of proxies predates the Web, the rapid growth of the Web in the mid 1990s spurred a strong interest in using proxies to reduce user-perceived latency, as well as server and network load. Many organizations deployed proxy caches to move popular Web content closer to their users. In addition, proxies acted as gateways to translate messages from one protocol to another.

Over time, proxies began to perform a wider range of functions. By sending requests to servers on behalf of a collection of clients, a proxy typically obscures the identity of individual users. Some proxies go one step further and remove information (such as e-mail addresses or cookies) that might compromise the users' privacy. On the other hand, some organizations use proxies to collect logs of the users' requests or to block access to certain Web pages based on the URL or the contents. A proxy may transform or adapt the server response message to compress the data, insert advertisements, or execute a Java applet. Proxies form the core part of content distribution networks that transfer Web resources from a variety of locations in the Internet.

The growing popularity of multimedia applications has led to proxy services for audio and video streams. In contrast to the transfer of traditional Web content, multimedia streaming imposes strict timing requirements on the delivery of data to the receiver. Caching all or part of an audio or video stream at a proxy can reduce the playback delay experienced by users. In some cases, the proxy may be able to handle client VCR functions, such as rewind and fast forward, without contacting the server. Proxies can improve the audio and video quality by retransmitting packets that have been lost in the network. Alternatively, a proxy can react to network congestion by selectively discarding less important parts of the stream, or by transcoding the audio and video data to a lower bandwidth version. A group of proxies can be used to form an application-level multicast tree for streaming data to multiple clients distributed throughout the Internet.

Extending the client/server model to incorporate proxies introduces a number of challenging research problems. Many studies have proposed new proxy services that can improve performance, reduce overhead, or enrich a particular application. Other work focuses on evaluating the performance of proxy services based on analytic models or simulations using workloads derived from traffic measurements. Researchers have also proposed and evaluated new algorithms for important tasks such as cache replacement and prefetching. Other algorithmic work has considered where to place proxies in the network or how to assign clients to particular proxies. Many of these problems are addressed in the papers in this special issue, which provides a view of the state-of-the-art in research on Internet proxy services.

II. Overview of the Issue

The 12 papers in this issue fall into four main areas: Web caching, multimedia streaming, server replication, and dynamically generated content.

A. Web Caching

Caching has been a popular proxy service since the early days of the Web. The issue's first paper, "Implications of Proxy Caching for Provisioning Networks and Servers," investigates the benefits of proxy caching during periods of heavy load. The study analyzes measurement data to show that, although proxies are generally effective in reducing the demands on servers and the network, these benefits are not as dramatic when the load is high. The next two papers investigate the use of distributed caching proxies for delivering Web content. The paper on "Limitations and Benefits of Cooperative Proxy Caching" presents an analytical model of user-perceived latency for two architectures for cooperating proxies. The study shows that having access to multiple proxies offers a marginal improvement in average user response time and a somewhat larger reduction in the number of transfers that experience high delay. The third paper is entitled "Hierarchical Web Caching Systems: Modeling, Design, and Experimental Results." This paper presents an analytical model of a hierarchical system, where each cache applies the Least Recently Used replacement algorithm. The analytic results motivate the creation of a cooperative proxy caching scheme that reduces the memory and processing requirements compared to a traditional hierarchical architecture.

B. Multimedia Streaming

The next four papers focus on proxy services for multimedia streaming. Caching is an important topic in this area as well. However, the large size of audio and video streams often precludes storing entire objects at the proxy. The first paper, "Scalable Proxy Caching of Video Under Storage Constraints," presents algorithms for selective caching of video frames at a proxy. The frames are chosen carefully to reduce the network resource requirements for transmitting a variable bit-rate stream, within the limitations of the storage space at

Publisher Item Identifier 10.1109/JSAC.2002.802747.

the proxy. The second paper is entitled "Silo, Rainbow, and Caching Token: Schemes for Scalable, Fault Tolerant Stream Caching." This work proposes several ways to cache different portions of a stream at a distributed collection of proxies. The paper describes data placement and replacement techniques that minimize the storage requirements and startup latency, as well as the load on the servers and the network.

The third paper, "Optimal Multicast Smoothing of Streaming Video Over the Internet," also considers how to exploit the buffer space available at multiple proxies inside the network. In this paper, the proxies are part of an application-level multicast tree that distributes a variable bit-rate video stream to a collection of clients. Each proxy performs workahead transmission of the data to the downstream buffers to reduce the network resource requirements. The fourth paper is entitled "A Multiplexing Scheme for H.323 Voice-Over-IP Applications." This work describes the design of a proxy that combines small voice packets from multiple sources into a single IP packet to improve the bandwidth efficiency of existing IP telephony applications. A second proxy performs demultiplexing to deliver the original voice packets to the receiving applications.

C. Server Replication

The next three papers focus on the placement of Web content at multiple locations in the network and the dynamic selection of a replica to satisfy each client request. The first paper, "Constrained Mirror Placement on the Internet," presents several algorithms for choosing locations for replicas from a fixed set of candidate sites. The work evaluates algorithms in terms of the reduction in client latency and server load and shows that a relatively simple greedy algorithm performs well compared to more complex approaches. The second paper is entitled "Placement Problems for Transparent Data Replication Proxy Services." This work considers where to locate server replicas that allow both read and write access to the data. The proposed algorithms consider the frequency of read and write operations to each of the replicated objects. The third paper, "Multicast Server Selection: Problems, Complexity, and Algorithmic Solutions," focuses on servers that transmit data to multiple clients simultaneously. The paper proposes algorithms that assign each client to an appropriate multicast server in environments where the set of clients may change over time and the servers may adapt their transmission rates to the prevailing network conditions.

D. Dynamically-Generated Content

The final two papers propose new services where the proxy participates in the creation of content for the user. The first paper is entitled "WebGraph: A Framework for Managing and Improving Performance of Dynamic Web Content." The work proposes dividing a Web page into static and dynamic components. Satisfying a request for the Web page involves retrieving the static portions, which may be cached at proxies and combining them with the dynamic components, which are generated on-demand. The second paper, "Spout: A Transparent Distributed Execution Proxy for Java Applets," addresses the security problems associated with Java virtual machines. The work proposes an distributed execution environment, where a proxy performs the applet's application logic and the end host simply executes the graphical user interface. This approach protects the user's machine from malicious Java applets without requiring modifications to the Web browser or the class libraries running on the end host.

ERNST BIERSACK, *Guest Editor*
Institute Eurecom
Corporate Communications Department
Sophia Antipolis, 06904 France

PETER DANZIG, *Guest Editor*
University of Southern California
Los Angeles, CA 90089 USA

JENNIFER REXFORD, *Guest Editor*
AT&T Laboratories—Research
Florham Park, NJ 07932 USA

ELLEN W. ZEGURA, *Guest Editor*
Georgia Institute of Technology
College of Computing
Atlanta, GA 30332–0280 USA

N. F. MAXEMCHUK, *J-SAC Board Representative*

REFERENCES

[1] J. Crowcroft, M. Handley, and I. Wakeman, *Internetworking Multimedia*. San Mateo, CA: Morgan Kaufmann, 1999.
[2] B. Krishnamurthy and J. Rexford, *Web Protocols and Practice: HTTP/1.1, Networking Protocols, Caching, and Traffic Measurement*. Reading, MA: Addison-Wesley, 2001.
[3] A. Luotonen, *Web Proxy Servers*. Englewood Cliffs, NJ: Prentice-Hall, 1997.
[4] M. Rabinovich and O. Spatscheck, *Web Caching and Replication*. Reading, MA: Addison-Wesley, 2001.
[5] J. Wang, "A survey of web caching schemes for the internet," *ACM Comput. Commun. Rev.*, vol. 29, no. 5, Oct. 1999.
[6] D. Wessels, *Web Caching*: O'Reilly Internet Ser., June 2001.

Ernst Biersack (M'88) received the M.S. and Ph.D. degrees in computer science from the Technische Universität München, Munich, Germany.

Since March 1992, he has been a Professor in Telecommunications at Institut Eurecom, Sophia Antipolis, France. His current research is on architectures for overlay networks, scalable video distribution, and network tomography. He is currently an Associate Editor of *ACM Multimedia Systems*.

Dr. Biersack is currently an Associate Editor of the IEEE *Network Magazine* and IEEE/ACM TRANSACTIONS ON NETWORKING. He received the Outstanding Paper Award of the IEEE Conference on Multimedia Computing and Systems, in 1996, for his work on synchronization in video servers (with W. Geyer). He received the 1999 W. R. Bennet Price of the IEEE for the best original paper published in the 1998 IEEE/ACM TRANSACTIONS ON NETWORKING for his work on reliable multicast (with J. Nonnenmacher and D. Towsley).

Peter Danzig received the B.S. degree in applied physics from University of California, Davis, and the Ph.D. degree in computer science from the University of California, Berkeley.

He is an Adjunct Associate Professor at the University Southern California, Los Angeles, CA. He is the former Vice President of Technology of Akamai Technology's west coast team, where he directed engineering and business development. He has published extensively. His research on Internet object caches became the reference implementation for Internet Web caches and the basis of the most successful commercial Content Delivery Networks, Squid, the Cisco cache engine, and Network Appliance's NetCache. He served as Chief Architect and CTO of Network Appliance's NetCache division. He founded Internet Middleware Corporation, to build the first industry-grade Web caches and sold the company to Network Appliance, in 1997.

Dr. Danzig has won the NSF National Young Investigator Award, two University of Southern California Innovative Teaching Awards, and the University of California Angelokous Award.

Jennifer Rexford (S'89–M'96–SM'01) received the B.S.E. degree in electrical engineering from Princeton University, NJ, in 1991 and the M.S.E. and Ph.D. degrees in computer science and electrical engineering from the University of Michigan, Ann Arbor, in 1993 and 1996, respectively.

She is now a member of the Internet and Networking Systems Center at AT&T Labs-Research, Florham Park, NJ. She is coauthor of *Web Protocols and Practice* (Reading, MA: Addison-Wesley, 2001), with B. Krishnamurthy. Her publications are available via the Web at <<http://www.research.att.com/~jrex>>. Her research interests include routing protocols and Internet traffic characterization.

Dr. Rexford is a Member of the Editorial Board of IEEE/ACM TRANSACTIONS ON NETWORKING.

Ellen W. Zegura (M'93) received the B.S. degree in computer science and electrical engineering, the M.S. degree in computer science, and the D.Sc. degree in computer science all from Washington University, St. Louis, MO, in 1987, 1990, and 1993, respectively.

She is currently an Associate Professor in the College of Computing, Georgia Institute of Technology, Atlanta, GA. Her research interests include active networking, server selection, anycast and multicast routing, and modeling large-scale internetworks.

Dr. Zegura is currently an Editor for IEEE/ACM TRANSACTIONS ON NETWORKING.