

# High-Volume Web Servers: Traffic Patterns, Performance Implications, and Resource Management

Dr. Mark S. Squillante

*IBM Thomas J. Watson Research Center*

## Abstract

The control, management and optimization of performance-based measures in high-volume Web sites require a fundamental understanding of the user request patterns and the performance implications of such traffic patterns. In this talk we will first present an analysis of the request patterns found at various Web server complexes, including high-volume commercial and sporting event Web sites. Our analysis demonstrates complex traffic patterns that include both short-range and long-rangedependence structures and both light-tailed and heavy-tailed behaviors. We then investigate the impact of these complex user request patterns on the response-time distribution of different Web server environments.

Our analysis illustrates some of the key difficulties in the management and control of system resources to satisfy certain types of service-level agreements. We then describe aspects of the system architecture and techniques employed at various commercial and sporting event Web sites that reduce the overhead of serving dynamic content and that improve the Web server response-time distribution. We will also present research on controlling the performance of different classes of user requests, including some new forms of service-level agreements. Various applications of our analysis and results will be discussed.

*Mark S. Squillante received the Ph.D. degree in computer science from the University of Washington, Seattle, WA, in 1990. He has been a Research Staff Member at the IBM Thomas J. Watson Research Center, Yorktown Heights, NY, since 1990, and an adjunct faculty member of the Department of Computer Science at Columbia University, New York, NY, from 1991 through 1996. He currently manages the Systems Analysis and Optimization group. From 1982 to 1985 he was a Member of the Technical Staff at Bell Telephone Laboratories, Murray Hill, NJ. His research interests concern fundamental problems in the mathematical analysis, modeling and optimization of computer systems, applications and services. Dr. Squillante is a Senior Member of IEEE, and a member of ACM, IFIP W.G. 7.3, INFORMS and SIAM. He currently is an Associate Editor of Operations Research and an Editor of Performance Evaluation.*