
Preface

In the past decade we have witnessed explosive growth of Internet services. With the popularity of Web surfing, instant messaging, on-line shopping, and other Internet services, people rely more and more on the Web to share information, communicate to each other, and even conduct business. There is little doubt that the ever-expanding Internet services will continue to change the way we live.

Due to the unprecedented scale of the Internet, popular Internet services must be scalable so as to respond to thousands or even hundreds of thousands of concurrent requests in a timely manner. The requests can be from clients in different geographical places and with different types of access networks and devices. They may also have different expectations for quality of services (QoSs) for various reasons. This scalability requirement goes beyond the capability of any of today's uniprocessor servers. Large-scale Internet services are becoming an increasingly important class of driving applications for scalable computer systems. In particular, server clusters, which are locally or globally distributed, become commonplace behind popular mission-critical Internet services; Web caches, proxies, and content delivery networks are widely deployed at network edges to reduce access latency and save network bandwidth. For example, Google search engine used more than 6,000 Linux/Intel PCs to serve an average of more than 200 million queries per day as of 2003; Akamai formed an Internet overlay network based on over 14,000 servers deployed in more than 1,100 networks to help reap the performance and reliability of media-rich content delivery services; peer-to-peer (P2P) computing model on the horizon pushes the envelope even further by allowing Internet users to serve each other directly. More research projects are ongoing for scalable and highly available Internet services.

Scalability aside, security is another primary concern in Internet services because client and server are often in different administrative domains and the underlying TCP/IP communication protocols are insecure by design. Many technologies have been developed over the years for secure Internet services. A notable is secure socket layer (SSL) protocol for secure electronic transactions. It lays a security foundation for more than \$5.5 billion e-commerce business in the 2003 U.S. retail section. Digital signature technology and its legitimation have paved a way to paperless business transactions on the Web. Java Virtual Machine (JVM) provides an agglomeration of security technologies toward general-purpose trusted computing on the Internet.

On the other hand, with the penetration of Internet services into our daily lives, people become more concerned than ever about security and privacy of cyber activities. People are suspicious of service scalability technologies such as mobile code and prefetching, and more reluctant to accept those without proof of security. Secu-

ity has a strong interplay with scalability in Internet services.

This book is intended to provide readers with an overview of scalable and secure Internet services and architecture and in-depth analysis of a number of key scaling technologies on the horizon. The topics include

- Server clusters and load balancing
- QoS-aware resource management
- Server capacity planning
- Web caching and prefetching
- P2P overlay network
- Mobile code and security
- Mobility support for adaptive grid computing

The coverage of each topic starts with a problem definition, a comprehensive review of current representative approaches for the problem. It is then followed by a detailed description of novel technologies that we recently developed at Wayne State University. The book stresses the underlying principles of the technologies and the role of these principles in practice with a balanced coverage of concepts and engineering trade-offs. It demonstrates the effectiveness of the technologies by rigorous mathematical modeling/analyses, simulation, and practical implementations. Most of the technologies were originally presented in peer-reviewed technical conferences and journals. This book is not a collection of these published works. It blends concepts, principles, design, analysis, and engineering implementations of a wide array of technologies in a unified framework for scalable and secure Internet services. It represents a systematic treatment of the subject, based on our own cutting-edge research experience over the years.

This book features a balanced coverage in breadth and depth of advanced scaling technologies in support of media streaming, e-commerce, grid computing, personalized content delivery services, distributed file sharing, network management, and other timely Internet applications. A number of software packages that we have recently developed as companions of our research publications are also covered.

Intended Audience

This book is meant for researchers, students, and practicing engineers in the fields of distributed computing systems and Internet applications. The relevance for these fields is obvious, given the increasing scalability and security concerns in Internet services. This book should provide researchers and students with a thorough understanding of major issues, current practices, and remaining problems in the area. For engineers who are designing and developing Internet services, this book provides insights into the design space and engineering trade-offs of scalability and security issues. They may also find the technologies described in this book applicable to real world problems. This book assumes that readers have general knowledge about computer systems and the Internet. It can be used as a text for senior and graduate students, as well as a reference for information technology (IT) professionals.

Acknowledgments

I wish to thank all those who contributed to this book. I thank Ramzi Basharahil, Aharon Brodie, Guihai Chen, Wanli Chen, Ngo-Tai Fong, Song Fu, Daniel Grosu, Tamer Ibrahim, Young-Sik Jeong, Manoj Kona, Shiyong Lu, Jayashree Ravi, Haiying Shen, Weisong Shi, Philip Sokolowski, Ravindra Sudhindra, Le Yi Wang, Jianbin Wei, Brain Wims, Minghua Xu, George Yin, Xiliang Zhong, Xiaobo Zhou, and other former and current CIC group members and colleagues of Wayne State University, without whom this book would not be possible. Part of this book draws on the papers that we have written together in the area of scalable and secure Internet services in the past several years.

A number of other colleagues have contributed to this book in various ways. I would especially thank Dharma P. Agrawal, Laxmi N. Bhuyan, Monica Brockmeyer, Jiannong Cao, George Cybenko, Vipin Chaudhary, Sajal K. Das, Minyi Guo, Kai Hwang, Weijia Jia, Hai Jin, Vijay Karamcheti, Francis C. M. Lau, Minglu Li, Philip McKinley, Burkard Monien, Lionel M. Ni, David Nicol, Yi Pan, Alexander Reinefeld, Nabil J. Sarhan, Loren Schwiebert, Mukesh Singhal, Pradip K. Srimani, Xian-He Sun, Xiaodong Zhang, Cho-Li Wang, Jie Wu, Li Xiao, Zhiwei Xu, Tao Yang, Yuanyuan Yang, and Wei Zhao for reading the manuscript or offering comments and suggestions. I am grateful to Professor Sartaj K. Sahni, the Editor-in-Chief of *Computer and Information Science Series* of Chapman & Hall/CRC, for his encouragement and suggestions during the course of this project. He reviewed the very first version of this book during IPDPS'04 in Santa Fe, New Mexico. I appreciated very much his insightful comments and suggestions.

I also thank acquiring editor Bob Stern, project coordinator Theresa Del Forn, and project editor Sarah E. Blackmon for essential support throughout the arduous process of getting this book into print. Sarah scrupulously read the final book draft for presentation flaws. A thanks also goes to Zhiping Liu for the cover photo of the Ambassador Bridge in Detroit.

This research was supported in part by U.S. NSF CDA-9619900, EIA-9729828, CCR-9988266, ACI-0203592, NASA 03-OBPR-01-0049, and Wayne State University's Research Enhancement Program and Career Development Chair Award. Any opinions, findings, and conclusions expressed in this material are those of the author and do not necessarily reflect the views of the funding agencies.

After all, I would like to thank my wife for all she has done and put up with over the years. Without her constant love and support, this book would be an impossible mission. My two adorable daughters also deserve a big thank-you for the joy they bring. I dedicate this book to them in gratitude and love.

Cheng-Zhong Xu
Detroit, Michigan U.S.