To my wife Emilia and my son George

Foreword

The fundamental aspect of the Internet architecture that distinguishes it from other network technologies (such as X.25 and ATM) is that it is connectionless (vs. connection-oriented) and stateless (vs. stateful). The heated debate of whether connection-oriented or connectionless architecture is better has lasted for several decades. Proponents of the connectionless architecture point out the great robustness and scalability properties of the architecture, as demonstrated by the Internet. One well-known articulation of this philosophy is the "End-to-End Arguments". Opponents argue, rightfully, that there is no known solution that can provide quantitative performance assurances or guaranteed QoS in a connectionless network. It has been widely recognized that QoS is a must-have feature as the Internet technology evolves to the next stage. However, all existing solutions that provide guaranteed QoS require routers to maintain per flow (another name for connection used by the Internet community) state, which is the fundamental element of a connection-oriented architecture. The apparent conflicting goals of having a stateless network and supporting QoS have presented a great dilemma for Internet architects. As an example, Dave Clark, one of the most respected Internet architects and the author of the famous "End-to-End Arguments" paper, was also a key designer of the Internet Integrated Services Architecture that requires routers to maintain per flow state.

Dr. Ion Stoica's dissertation addresses this most pressing and difficult problem facing the Internet community today: how to enhance the Internet to support rich functionalities (such as QoS and traffic management) while still maintaining the scalability and robustness properties embodied in the original Internet architecture.

In his dissertation, Dr. Stoica proposes a novel architecture called SCORE (Stateless Core) that does not require core routers to maintain per flow state yet can provide services similar to those provided by stateful networks. This is achieved by a family of SCORE distributed algorithms that approximate the services provided by idealized stateful networks. The key technique used to implement a SCORE network is Dynamic Packet State (DPS), which uses extra state carried in packet headers to coordinate distributed algorithms implemented by routers. Such an architecture has both important theoretical and practical significances. From a conceptual point of view, this architecture is the first that combines the advantages of stateful and stateless networks, and can therefore achieve QoS, scalability, and robustness simultaneously. From a practical point of view, the industry and the IETF have been struggling to make a choice between two QoS architectures: the stateful Intserv, which can provide hard QoS guarantees but is less scalable and robust, and the stateless Diffserv, which is more scalable and robust but cannot provide services with high assurance. The SCORE architecture provides a third approach that is superior.

I believe that this research represents one of the most important and innovative contributions in networking research in the past decade. I hope that you will enjoy reading it and agree with me afterward.

Hui Zhang

Preface

This book contains the dissertation the author wrote at the Department of Electrical and Computer Engineering (ECE) at Carnegie Mellon University. This thesis was submitted to the ECE department in conformity with the requirements for the degree of Doctor of Philosophy in 2000. It was honored with the 2001 ACM Doctoral Dissertation Award.

Abstract

Today's Internet provides one simple service: best-effort datagram delivery. This minimalist service allows the Internet to be *stateless*, that is, routers do not need to maintain any fine-grained information about traffic. As a result of this stateless architecture, the Internet is both highly *scalable* and *robust*. However, as the Internet evolves into a global commercial infrastructure that is expected to support a plethora of new applications, such as IP telephony, interactive TV, and e-commerce, the existing best-effort service will no longer be sufficient. As a consequence, there is an urgent need to provide more powerful services such as guaranteed services, differentiated services, and flow protection.

Over the past decade, there has been intense research toward achieving this goal. Two classes of solutions have been proposed: those maintaining the *stateless* property of the original Internet (e.g., differentiated services), and those requiring a new *stateful* architecture (e.g., integrated services). While stateful solutions can provide more powerful and flexible services, such as per flow bandwidth and delay guarantees, they are less scalable than stateless solutions. In particular, stateful solutions require each router to maintain and manage per flow state on the control path, and to perform per flow classification, scheduling, and buffer management on the data path. Since today's routers can handle millions of active flows, it is difficult, if not impossible, to implement such solutions in a scalable fashion. On the other hand, while stateless solutions are much more scalable, they offer weaker services.

The key contribution of this dissertation is to bridge this long-standing gap between stateless and stateful solutions in packet-switched networks such as the Internet. Our thesis is that "*it is actually possible to provide services as powerful and as flexible as the ones implemented by a stateful network using a*

stateless network". To prove this thesis, we propose a novel technique called Dynamic Packet State (DPS). The key idea behind DPS is that, instead of having routers maintain per flow state, packets carry the state. In this way, routers are still able to process packets on a per flow basis, despite the fact that they do not maintain any per flow state. Based on DPS, we develop a network architecture called Stateless Core (SCORE) in which core routers do not maintain any per flow state. Yet, using DPS to coordinate actions of edge and core routers along the path traversed by a flow allows us to design distributed algorithms that emulate the behavior of a broad class of stateful networks in SCORE networks.

In this dissertation we describe complete solutions including architectures, algorithms, and implementations which address three of the most important problems in today's Internet: providing guaranteed services, differentiated services, and flow protection. Compared to existing solutions, our solutions *eliminate* the most complex operations on both the data and control paths in the network core, i.e., packet classification on the data path, and maintaining per flow state consistency on the control path. In addition, the complexities of buffer management and packet scheduling are greatly reduced. For example, in our flow protection solution these operations take constant time, while in previous solutions these operations may take time logarithmic in the number of flows traversing the router.

Acknowledgements

I am extremely grateful to Hui Zhang, my thesis advisor, for giving me the right amount of freedom and guidance during my graduate studies. From the very beginning, he treated me as a peer and as a friend. He was instrumental in maintaining my focus, and constantly reminding me that identifying the research problem is as important, if not more important, than finding the right solution. Hui not only taught me how to become a better researcher, but also helped me to become a better person. His engaging arguments and strong feedback contributed greatly to this dissertation. I hope for and look forward to continued collaboration with him in the future.

The genesis of this thesis can be traced back to my internship at Xerox PARC in the summer of 1997. It all started with Scott Shenker asking the intriguing question: "Can we approximate Fair Queueing without maintaining per flow state in a network cloud?" I am indebted to Scott for teaching me how to rigorously define a problem and then pursue its solution. During these years he was an invaluable source of feedback and support. His suggestions and insights had a great impact on this dissertation.

I am grateful to the other members of my committee, Garth Gibson, Thomas Gross, and Peter Steenkiste, for their feedback and for their advice that helped to shape my research skills. Garth taught me the art of asking the right questions in a technical discourse. His inquisitorial and sharp questions helped me to better understand the limitations of my research and motivated me to find better ways to explain my results. Thomas provided the right balance to my research by constantly encouraging me to not get buried in the algorithmic details, but to always try to put my work into perspective. Peter always found time to discuss research issues, and gave excellent feedback. He was one of the first to suggest using Dynamic Packet State to provide guaranteed services.

Thanks to Mihai Budiu, Yang-hua Chu, Urs Hengartner, Eugene Ng, Mahadev Satyanarayanan, and Jeannette Wing for their feedback and comments that helped to improve the overall quality of this dissertation. Thanks to Joan Digney for accommodating me in her busy schedule and proofreading this thesis, which helped to significantly improve the presentation.

I am grateful to all the friends and colleagues with whom I spent my time as a graduate student at Carnegie Mellon University. Thanks to my Romanian friends, Mihai and Raluca Budiu, Cristian Dima, Marius Minea, and George Necula, with whom I spent many hours discussing the most varied and exciting topics. Mihai sparkled many of these conversations with his wit, and by being a never empty reservoir of information. Thanks to my networking group colleagues and officemates Yang-hua Chu, Urs Hengartner, Nick Hopper, Tom Kang, Marco Mellia, Andrew Myers, Eugene Ng, Sanjay Rao, Chuck Rosenberg, Kay Sripanidkulchai, Donpaul Stephens, and Yinglian Xie. I treasure our animated Friday lunch discussions that always managed to end the week on a high note. Special thanks to Eugene for his help and patience with my countless questions. He was the default good-to-answer-all-questions person whom I asked about everything, from how to modify the `if_de` driver in FreeBSD, to which are the best restaurants around the campus.

The completion of this thesis marked the end of my many years as a student. Among many outstanding teachers who shaped my education and scientific career are: Hussein Abdel-Wahab, Irina Athanasiu, Kevin Jeffay, David Keyes, Trandafir Moisa, Stephan Olariu, Alex Pothen, and Nicolae Tapus. I am grateful to Kevin, who was instrumental in helping and then convincing me to come to Carnegie Mellon University. Many thanks to Irina who, during my studies at the "Politehnica" University of Bucharest, gave me the mentorship a student can only dream about.

I would like to express my earnest gratitude to my parents and my sister for their love and support, without which any of my achievements would not have been possible. Thanks to my father who sparkled my early interest in science and engineering. His undaunting confidence gave me the strength to overcome any difficulties and to maintain high goals. Thanks to my mother for her love and countless sacrifices to raise me and to give me the best possible education.

I am deeply indebted to my dear wife Emilia for her love and understanding through my graduate years. She was always behind me and gave her unconditional support even if that meant sacrificing the time we spent together. I thank my mother and my mother-in-law who devotedly took care of our son for two and a half years. Without their help, it would not have been possible to reach this stage in my career. Finally, thanks to our son, George, for the joy and the happiness he brings to me during our many moments together.

Berkeley, December 2003 *Ion Stoica*

Springer

http://www.springer.com/978-3-540-21960-6

Stateless Core: A Scalable Approach for Quality of Service in the Internet
Winning Thesis of the 2001 ACM Doctoral Dissertation Competition
Stoica, I.
2004, XVI, 219 p., Softcover
ISBN: 978-3-540-21960-6