

Architecture of **NETWORK SYSTEMS**

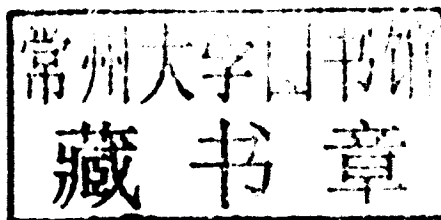
MK
MORGAN KAUFMANN

Dimitrios **SERPANOS** Tilman **WOLF**

Architecture of Network Systems

Dimitrios Serpanos

Tilman Wolf



ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SYDNEY • TOKYO

Morgan Kaufmann Publishers is an imprint of Elsevier



Acquiring Editor: Todd Green
Editorial Assistant: Robyn Day
Project Manager: Andre Cuello
Designer: Kristen Davis

Morgan Kaufmann is an imprint of Elsevier
30 Corporate Drive, Suite 400, Burlington, MA 01803, USA

© 2011 Elsevier, Inc. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices

Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods or professional practices may become necessary. Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information or methods described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

Library of Congress Cataloging-in-Publication Data

Serpanos, Dimitrios Nikolaou.

Architecture of network systems / Dimitrios Serpanos, Tilman Wolf.

p. cm. — (The Morgan Kaufmann series in computer architecture and design)

Includes bibliographical references and index.

ISBN 978-0-12-374494-4 (pbk.)

1. Computer network architectures. I. Wolf, Tilman. II. Title.

TK5105.5.S4234 2011

004.6—dc22

2010045063

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

ISBN: 978-0-12-374494-4

Printed in the United States of America

11 12 13 14 10 9 8 7 6 5 4 3 2 1

Working together to grow
libraries in developing countries

www.elsevier.com | www.bookaid.org | www.sabre.org

ELSEVIER

BOOK AID
International

Sabre Foundation

For information on all MK publications visit our website at www.mkp.com

Architecture of Network Systems

To the memory of my father Nikolaos. -DS
To Ana and our daughter Susana. -TW

Preface

WHY WRITE THIS BOOK?

Data communication networks are widely used today and are an integral part of our daily life. The Internet is a medium for business, personal, and government communication, and it is difficult to envision today's society without this essential infrastructure. The continued success of the Internet is dependent on our ability to maintain and improve the functionality, performance, and scalability of these networks. As a basis for obtaining the necessary knowledge about networks, there is a clear need for textbooks that provide an introduction to the foundations of this topic as well as a detailed understanding of more advanced issues.

While many available books cover the design of network protocols and their operation, there has been precious little focus on the systems that implement networks. Early data communication networks struggled with the scarcity of transmission bandwidth, which led to significant efforts to improve management and efficient use of this resource. Over the last three decades, advances in transmission technology have led to the availability of vast amounts of bandwidth, thus shifting the main bottleneck of networks from the transmission medium to the switching and processing of transmitted data. As a result, modern networks and the Internet are not only in need of appropriate protocols for the wide deployment of applications and services, but also in need of efficient systems that enable the timely processing and forwarding of network traffic.

TARGET AUDIENCE

This book aims to serve as a textbook and reference for designers and implementers of networking technology, networking students, and networking researchers. The goal of the book is to present the systems issues of network systems, approaching them from the architecture, design, and implementation point of view. Considering that network systems are embedded systems that implement network protocols, readers of the book would benefit from being familiar with the basic concepts of networking, embedded systems, and computer systems organization and architecture.

APPROACH

This book is about computing systems or, more specifically, about a class of special-purpose embedded systems used in networking devices. As such, the main thrust of the book is the promotion of systems architectures and designs. Therefore, this book can be classified as a computing systems book.

In contrast to typical architecture and computing systems books, this book follows a structure analogous to those of typical networking books. Following the OSI reference model for network protocols, a model that has been proven highly valuable from an education point of view, we classify and present network systems and their designs. Later in the book, we also discuss specific components of network systems, similarly to books on system architecture.

Since this book is the first systematic effort to present the architecture of the complete range of network systems as a whole, we focus on promoting key concepts for all types of network systems. We attempt to present the main architectures and designs of systems and components that cover this area. Clearly, the book does not and cannot cover all existing material in this area, which has experienced fast growth in recent years. Instead, the book focuses on major concepts. In addition to presenting state of the art, we have tried to demonstrate the progress in the field during the last couple of decades through specific examples that illustrate the improvements of technology and indicate the path of evolution for network systems in the future.

COURSE USE

This book has been specifically developed for use in college courses at the upper undergraduate level and graduate level. Much of the material presented in this book has been used previously by the authors in a graduate and advanced undergraduate course on network systems architecture of the Department of Electrical and Computer Engineering at the University of Patras and in a graduate level networking course of the Department of Electrical and Computer Engineering at the University of Massachusetts Amherst. It is expected that students taking a course based on this book have some prior exposure to computer networks and computer system organization. While the Appendix of the book provides a brief overview on the network protocols used in the Internet, a more detailed course on this topic may be of value.

It is not necessary to cover the entire book in a course. While there are some dependencies between chapters, different courses can emphasize different characteristics of network systems. For courses that address the networking aspects of network systems, we suggest a focus on Chapters 4 through 10, while for courses that address the embedded system aspects of network systems, we suggest a focus on Chapters 11 through 14.

WE WANT TO HEAR FROM YOU

We would appreciate receiving any feedback you may have about the book. Tell us if you find mistakes, if you have suggestions for improvements, what you like about the book, how you have used it in a course, etc. You can contact the authors via email at serpanos@upatras.gr (Dimitrios Serpanos) and wolf@ecs.umass.edu (Tilman Wolf).

ACKNOWLEDGMENTS

We thank the many individuals who have helped us in making this book a reality. First and foremost, we thank our editor Todd Green and Nancy Hoffmann for guiding us through the publication process and helping us meet our deadlines. We are also grateful for the technical input and editorial comments by current and former students, specifically C. Datsios, G. Keramidas, A. Papalambrou, and A. Voyiatzis. In addition, we thank the anonymous reviewers of book chapters, whose valuable comments enabled us to address and promote the appropriate technical concepts and to deliver a more readable book. Any failure to achieve these goals is, of course, the responsibility of the authors. Last, but definitely not least, we acknowledge the support and patience demonstrated by our families during the long process of developing this book. Without their understanding and extreme patience, this book would never have been completed.

About the Authors

Dimitrios Serpanos is a Professor of Electrical and Computer Engineering at the University of Patras, Greece, and the Director of the Industrial Systems Institute (ISI/RC Athena). Currently, he is also the chairman of the governing board of the newly founded University of Western Greece. His research interests include embedded systems, with focus on network systems, security systems and multimedia systems, computer architecture, and parallel and distributed systems.

Serpanos holds Ph.D. and M.A. degrees in Computer Science from Princeton University, Princeton, since 1988 and 1990, respectively. He received his diploma in Computer Engineering and Informatics from the University of Patras in 1985.

Between 1990 and 1996 he was a research staff member at IBM Research, T.J. Watson Research Center, New York. Since November 1996 he has been a faculty member in Greece. Between 1996 and 2000 he was with the Department of Computer Science, University of Crete. Currently, he is with the Department of Electrical and Computer Engineering, University of Patras, where he is a professor. During his term in Crete he conducted research at ICS-FORTH; while in Patras he has been conducting research at ISI. Professor Serpanos' research has received funding from the EU, the Greek government, and the private sector in Europe and the United States.

Serpanos is a senior member of the IEEE, a member of the ACM, a member of the New York Academy of Sciences, and an educational member of USENIX. He is serving or has served as associate editor for technical journals, including *ACM Transactions on Embedded Computing Systems*, *IEEE Transactions on Industrial Informatics*, the *Journal of Internet Engineering*, and *International Journal on Computers and Their Applications*. He has served as general chair and TPC chair of several conferences and workshops, as well as a TPC member of more than 120 conferences and workshops. In addition, he has served as guest editor to special issues of *IBM Journal of Research and Development*, *IEEE Network*, *ACM Transactions on Embedded Computing Systems*, *IEICE/IEEE joint issues*, and *Telecommunication Systems*.

Tilman Wolf is an Associate Professor in the Department of Electrical and Computer Engineering at the University of Massachusetts Amherst. He is engaged in research and teaching in the areas of computer networks, router design, embedded systems, and network and system security. Wolf received a diploma in Informatics from the University of Stuttgart, Germany, and holds a D.Sc. in Computer Science and two M.S. degrees from Washington University in St. Louis.

Wolf started working in the area of router design and network processor design in the late 1990s when commercial interest in routers with programmable data paths started taking off. His work focused on the design of high-performance routers that use programmable port processors to provide dynamically changing functionality, with an emphasis on performance modeling and benchmarking.

Since joining the University of Massachusetts in 2002, Wolf and his students in the Network Systems Laboratory have designed and prototyped run-time systems for workload management across multiple embedded processor cores in packet processors. Their work also addresses the question of how to redesign Internet architecture to balance the need for custom networking functions with the need for simplicity and manageability. More recently, he and his students have explored security vulnerabilities in processing components of network systems that require embedded protection mechanisms.

Wolf is a senior member of the IEEE and the ACM. He has served as associate editor for the *ACM/IEEE Transactions on Networking*, as program committee member and organizing committee member of numerous professional conferences, including IEEE INFOCOM and ACM SIGCOMM, and as TPC chair and general chair for ICCCN. He has served as treasurer for the ACM SIGCOMM society for several years. He has received several recognitions for his educational activities, including a college outstanding teacher award.

Contents

Preface	xiii
About the Authors	xvii
CHAPTER 1 Architecture of network systems overview	1
Computer networks	1
Embedded systems	3
Protocols and network systems	5
Organization of this book	6
CHAPTER 2 Network protocols and network systems.....	11
Introduction.....	11
The open systems interconnection reference model	12
Protocol stacks and protocol elements	13
Network systems: Definition and types.....	17
Structure of network systems.....	21
Summary	24
CHAPTER 3 Requirements of network systems	25
Introduction.....	25
Requirements of network applications and services	26
Qualitative requirements	27
Quantitative requirements	27
Example requirements	28
The throughput preservation challenge	29
Traffic models and benchmarks.....	32
Summary	33
CHAPTER 4 Interconnects and switching fabrics.....	35
Introduction.....	35
Crossbars and interconnection networks	38
Switch organization	40
Switch scheduling.....	43
Centralized scheduling	46
Distributed scheduling.....	49
Use of randomization	57
Real-time traffic	61
Summary	61

CHAPTER 5	Network adapters	63
	Introduction	63
	Basic network adapter	64
	Adapter operation analysis	66
	Memory organization	70
	Memory management unit with local memory	70
	Memory management unit and processor with local memory ...	73
	Intelligent DMA	75
	Multiprocessor processing element	77
	Single SPE configuration	79
	Multi-SPE configuration—protocol multiprocessing	80
	Multi-SPE configuration—spatial parallelism	81
	Memory management schemes	84
	Detached MMU	86
	Distributed MMUs	87
	Summary	87
CHAPTER 6	Bridges and layer 2 switches	89
	Introduction	89
	Types of bridges	91
	Transparent bridging and special-purpose subsystems	92
	High-performance transparent bridge architecture	97
	Transparent bridging support unit	101
	Network attachment unit design	104
	Internal switching component operation	107
	Partitioned data memory organization	109
	Layer 2 switches	110
	Summary	110
CHAPTER 7	Routers	111
	Network layer	111
	Functionality of the network layer	111
	Systems requirements	112
	Generic router design	115
	Data plane	117
	Route lookup	118
	Queuing and buffering management	126
	Control plane	129
	Routing algorithms	129
	Error handling	136
	Example network layer systems	139
	Summary	140

CHAPTER 8	Transport layer systems	141
	Transport layer.....	141
	Functionality of the transport layer	141
	Network flows	143
	Packet classification	147
	Transport layer systems.....	156
	Summary	160
CHAPTER 9	Application layer systems.....	161
	Application layer	161
	Functionality of the application layer.....	162
	Application layer protocols	163
	Network system support for application layer	172
	Payload inspection.....	172
	Load balancing	180
	Summary	182
CHAPTER 10	Quality of service and security	183
	Cross-layer issues	183
	Quality of service	183
	Quality of service concepts.....	184
	Network support for QoS.....	187
	Link scheduling algorithms.....	188
	Quality of service summary	199
	Security	199
	Security concepts.....	199
	Cryptography overview	200
	Security in network protocols	204
	Denial-of-service attacks.....	207
	Summary	210
CHAPTER 11	Specialized hardware components	211
	Hardware support for packet forwarding	211
	General-purpose packet processors.....	211
	Performance vs flexibility	212
	Systems-on-a-chip	213
	Network processors	214
	Design choices	220
	Example systems	223
	Special-purpose hardware accelerators	224
	Trade-offs in use of accelerators	224
	Example accelerators.....	225

Accelerator implementations.....	226
Summary	227
CHAPTER 12 Power issues in network systems	229
Introduction.....	229
Lookup engines	230
Network processors	235
Summary	237
CHAPTER 13 Networks on chips	239
Introduction.....	239
Network-on-chip architectures	240
Network-on-chip routing schemes	241
Technology and circuit design.....	242
Bus delay model.....	243
Crossbar delay model.....	244
Power issues	247
Summary	248
CHAPTER 14 Run-time support systems.....	249
Software support for network systems	249
Network dynamics.....	250
Run-time components.....	250
Software interfaces	252
Operating system support for networking	253
Networking software in operating systems	253
Software interactions	254
Performance considerations.....	256
Specialized router software	259
Router operating systems	259
Packet processing software	262
Summary	265
CHAPTER 15 Next-generation Internet architecture.....	267
Need for next-generation Internet.....	267
Networking trends	267
Limitations in the Internet	268
Extended reach of the Internet.....	269
New networking paradigms	270
Implications	273
Network virtualization.....	273
Diversity of protocols.....	273

Single infrastructure	274
Virtualized networks	275
Programmability	277
Programmability in next-generation networks	277
Implementing programmability.....	279
Experimental testbeds.....	280
Commercial incentives	281
Summary	281
APPENDIX The layered Internet architecture	
and network protocols.....	283
Layered Internet architecture	283
Hourglass architecture	284
Protocol processing in network systems.....	285
Example protocol stack	286
Link layer.....	287
Medium access control.....	287
Ethernet.....	290
Network layer	292
Internet protocol suite	292
Internet protocol	292
Transport layer.....	294
Reliable data transfer	295
Transmission control protocol.....	297
Summary	299
References	301
Index.....	313

Architecture of network systems overview

1

Computer networks have become critical infrastructure on which we rely for personal, business, and government use. Network systems are the hardware and software components from which these networks are built. Network systems determine what functionality a computer network can provide and what performance it can achieve. Due to this critical role, we believe it is important to study the architecture and operation of these network systems.

Network systems draw from concepts and technologies in computer networks, embedded systems, computer organization, and distributed computing. The convergence of these very diverse technical areas makes the study of network systems particularly exciting. This diversity also requires a thorough understanding of the relationship between these areas and how they influence network system design. We hope to provide these insights in this book.

COMPUTER NETWORKS

The advances of transmission technology for more than two decades have brought significant changes in networking as well as computing. In the 1970s and 1980s, standard networks provided limited connectivity, achieving bandwidth in the order of kilobits per second (kbps) up to a few megabits per second (Mbps) for local area networks, where the maximum speed reached 10 to 16 Mbps. From the middle of the 1980s, the development and commercialization of high-speed links that provided bandwidth of several Mbps for point-to-point connectivity enabled development of a new generation of networks and protocols that enable communication at very high speeds, reaching today hundreds of gigabits per second (Gbps).

In parallel with the dramatic progress in transmission technology, in the last decade of the 20th century the Internet was commercialized, moving it from research use to commercial use. The need to provide Internet connectivity to end users at home and at work not only exploited the high-speed transmission technology that had been developed, but also led to significant progress in access technologies. This trend led to development of a wide range of access protocols to connect end users to the Internet through telephone lines, cable TV infrastructure, satellites, and so forth.

The deployment of high-speed links and networks, as well as the Internet, provided the infrastructure for the development of new computing paradigms, mainly network-centric computing. In this paradigm, newly developed system infrastructures are used to support computing and storage-intensive applications and services. An early characteristic example is the development of networks of workstations, a multiprocessor architecture that relies on high-speed connectivity among workstations. This multiprocessor model is a natural advance of traditional distributed systems, which connected autonomous computing systems; the single view of the network of workstations as one system, necessary for a multiprocessor, was enabled by the high-speed networks that had become available. This abstraction enabled the efficient management of distributed resources through appropriate computing models and enabled a unified view of the networked workstations to the users. In a different direction, the ability to provide access to data and computational resources over the Internet enabled a vast number of new services for users and customers of commercial enterprises. These services are based on the well-known client/server distributed computing model and include examples ranging from banking to news feeds and from video conferencing to digital libraries.

The provision of all these services and applications over networks, including the Internet, requires technological advances at two fronts: protocols and network systems. Network protocols define the methods and mechanisms necessary to achieve reliable communication between two parties (or more than two in the case of multicasting or broadcasting). For example, network protocols define methods with which data units are encoded for transmission, mechanisms to detect transmission errors, methods for retransmission of data in case they are lost or transmitted with errors, and methods for regulating the flow of information between communicating peers to ensure that the receiver is not flooded with incoming data. Importantly, network protocols do not define any aspect of the systems that execute these protocols in order to implement data communication. For example, protocols do not define the type of processors, their speed, the size of memory, or any other systemic characteristic of the devices that implement these protocols.

Network systems are the systems and subsystems that realize the implementation of network protocols. Network systems need to be designed to meet the functional requirements specified by protocols. They also need to meet the performance requirements determined by the ever-increasing speed of transmission links. This relationship between network systems and related areas is illustrated in Figure 1-1. The demands for executing protocols at high speed led to the need for advanced, sophisticated system architectures, component designs, and implementations. These network systems constitute the focus of this book.

Network systems represent a distinct area of embedded systems architecture. Network systems are embedded systems because they are embedded in autonomous systems and devices that have specific purposes. For example, network systems are present in the infrastructure of networks, such as in switches, bridges, routers, and modems. Importantly, network systems also include network adapters,