NETWORK INFRASTRUCTURE AND ARCHITECTURE

Designing High-Availability Networks

KRZYSZTOF INIEWSKI
CARL McCROSKY
DANIEL MINOLI

A JOHN WILEY & SONS, INC., PUBLICATION

NETWORK INFRASTRUCTURE AND ARCHITECTURE

NETWORK INFRASTRUCTURE AND ARCHITECTURE

Designing High-Availability Networks

KRZYSZTOF INIEWSKI
CARL McCROSKY
DANIEL MINOLI

A JOHN WILEY & SONS, INC., PUBLICATION

Published by John Wiley & Sons, Inc., Hoboken, New Jersey.
Published simultaneously in Canada.

Library of Congress Cataloging-in-Publication Data:

Iniewski, Krzysztof.
Network infrastructure and architecture: designing high-availability networks / Krzysztof Iniewski, Carl McCrosky, Daniel Minoli.
p. cm.
Includes index.
ISBN 978-0-471-74906-6 (cloth)
1. Optical communications. 2. Integrated circuits—Very large scale integration. 3. Data transmission systems—Design and construction. I. McCrosky, Carl, 1948– II. Minoli, Daniel, 1952– III. Title.
TK5103.59.I49 2008
621.382′7—dc22

2007034273

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

For
Ferdynand Iniewski
Judy Berlyne McCrosky
Anna
with affection and thanks

CONTENTS

PREFACE xvii

PART I OPTICAL TRANSMISSION

1 Introduction to Networking 1

1.1 Introduction, 1
1.2 Transmission Media, 2
1.2.1 Copper Wire, 2
1.2.2 Coaxial Cable, 3
1.2.3 Optical Fiber, 4
1.2.4 Wireless Communication, 5
1.3 Basic Networking Concepts, 6
1.3.1 LAN, SAN, MAN, and WAN, 6
1.3.2 Network Topologies, 9
1.3.3 Circuit vs. Packet Switching, 11
1.3.4 Wavelength vs. Time vs. Statistical Multiplexing, 13
1.4 Open System Interconnection Model, 16
1.4.1 Basic Concept, 16
1.4.2 OSI Model and Data Encapsulation, 17
1.4.3 Network Overlay Hierarchy, 19
1.5 Networking Equipment, 20
1.5.1 Regenerators, Modems, Hubs, and Add–Drop Multiplexers, 21
1.5.2 Switches, 22
1.5.3 Routers, 22
1.5.4 Networking Service Models, 24
Key Points, 26
References, 28

2 Filer-Optic Transmission 31

2.1 Introduction, 31
2.2 Fiber Optic Communication, 32
2.2.1 Why Optical Fiber?, 32
2.2.2 Propagation: Single- and Multimode Fibers, 35
2.3 Light Emission and Detection, 39
2.3.1 Light Sources, 39
2.3.2 Photodetectors, 44
2.4 Optical Modulation, 46
2.4.1 Direct Modulation, 46
2.4.2 External Modulation, 47
2.5 Optical Amplification, 55
2.5.1 Erbium-Doped Fiber Amplifiers, 56
2.5.2 Raman Amplifiers, 59
2.5.3 EDFA vs. Raman Amplifier, 61
2.6 Fiber Transmission Impairments, 62
2.6.1 Chromatic Dispersion, 63
2.6.2 Dispersion Management Techniques, 66
2.6.3 Polarization Mode Dispersion, 72
2.6.4 Nonlinear Effects, 77
Key Points, 82
Acknowledgments, 83
References, 84

3 Wavelength-Division Multiplexing 87

3.1 Introduction, 87
3.2 WDM Technology, 88
3.2.1 WDM Basics, 88
3.2.2 WDM Bandwidth Capacity, 89
3.2.3 Coarse vs. Dense WDM Systems, 91
3.2.4 Future Extensions of DWDM Capacity, 92
3.3 Networking Equipment for WDM, 95
3.3.1 WDM Regenerators, 95
3.3.2 Optical Cross-Connects and Switches, 96
3.3.3 Optical Add–Drop Multiplexers, 100
3.4 WDM Networks, 102
3.4.1 WDM Network Provisioning, 102
3.4.2 Wavelength Blocking, 103
3.4.3 O-E-O Conversion in WDM Networks, 104
3.4.4 WDM Network Protection, 105
3.5 Case Study: WDM Link Design, 105
Key Points, 108
References, 109

PART II NETWORKING PROTOCOLS

4 SONET 111

4.1 Introduction, 111
4.2 SONET Networks, 112
4.2.1 SONET Transmission Rates, 112
4.2.2 SONET Network Architectures, 113
4.3 SONET Framing, 117
4.3.1 STS-1 Building Block, 117
4.3.2 Synchronous Payload Envelope, 120
4.3.3 SONET Virtual Tributaries, 125
4.3.4 SDH vs. SONET, 127
4.4 SONET Equipment, 128
4.4.1 SONET O-E-O Regenerator, 128
4.4.2 SONET ADM Multiplexer, 128
4.4.3 SONET Terminal Multiplexer, 129
4.5 SONET Implementation Features, 129
4.5.1 SONET Scrambling, 129
4.5.2 SONET Clock Distribution, 130
4.5.3 SONET Byte Stuffing, 133
Key Points, 134
References, 135

5 TCP/IP Protocol Suite 137

5.1 Introduction, 138
5.2 Structure of the Protocol Suite, 138
5.3 Internet Protocol, 145
5.3.1 IP Addresses, 145
5.3.2 IP Header Format and Function, 146
5.4 User Datagram Protocol, 148
5.5 Transmission Control Protocol, 149
5.5.1 TCP Header Format and Function, 151
5.5.2 Connection-Oriented Service, 153
5.5.3 Receiver Window, 153
5.6 TCP Flow Control, 155
5.6.1 Receiver-Based Flow Control, 156
5.6.2 Transmitter-Based Flow Control, 162
5.6.3 Fast Retransmit and Fast Recovery, 165
5.6.4 Delayed Acknowledgment, 166
5.6.5 Naigle's Algorithm, 167
5.7 IP Routing Mechanisms, 167
5.8 IP Route Calculations, 169

5.9 Difficulties with TCP and IP, 174
5.9.1 One Shortest Route, Regardless of Load Conditions, 174
5.9.2 Deliberate Congestion and Backoff, 175
5.9.3 Lack of Quality-of-Service Support, 175
5.9.4 Receiver Windows and Round-Trip Times, 175
5.9.5 Long, Fat TCP Pipes, 177
5.9.6 Big Packets on Thin Pipes, 178
5.10 IPv6: The Future?, 178
5.11 Conclusions, 180
Key Points, 180
References, 181

6 Protocol Stacks **183**

6.1 Introduction, 183
6.2 Difficulties with the TCP/IP Protocol Suite, 185
6.3 Supporting Protocols, 187
6.3.1 ATM, 188
6.3.2 Generic Framing Procedure, 190
6.3.3 Multiprotocol Label Switching, 192
6.3.4 Ethernet over the Internet, 195
6.3.5 Resilient Packet Rings, 196
6.3.6 G.709: Digital Wrapper Technology, 200
6.4 Legacy Solutions, 204
6.4.1 IP over SONET, 204
6.4.2 IP over ATM over SONET, 204
6.5 New Protocol Stack Solutions, 205
6.5.1 Using MPLS, 205
6.5.2 Future All- or Mostly Optical Networks, 207
6.5.3 Gigabit Ethernet over the Internet, 210
6.5.4 Storage Area Network Protocols over the Internet, 211
Key Points, 215
References, 217

PART III VLSI CHIPS

7 VLSI Integrated Circuits **219**

7.1 Introduction, 220
7.1.1 Integrated Circuits, VLSI, and CMOS, 220

7.1.2 Classification of Integrated Circuits, 221
7.1.3 Looking Ahead, 223
7.2 Integrated Circuits for Data Networking, 223
7.2.1 PMD and PHY Devices (Layer 1), 225
7.2.2 Framers and Mappers (Layer 2), 227
7.2.3 Packet Processing Devices (Layer 3), 230
7.3 Chip I/O Interfaces, 231
7.3.1 Serial vs. Parallel I/O, 232
7.3.2 Networking I/O Standards, 234
7.3.3 Design of Data Networking I/O Interfaces, 240
7.3.4 Memory I/O Interfaces, 243
7.3.5 Microprocessor I/O Interfaces, 245
7.4 Examples of Chip Architectures, 247
7.4.1 Time-Slice Architecture, 247
7.4.2 SONET Framer Architecture, 250
7.4.3 Network Processor Architecture, 252
7.5 VLSI Design Methodology, 253
7.5.1 Design Specification, 257
7.5.2 Functional Design and RTL Coding, 257
7.5.3 Functional Verification, 259
7.5.4 Design Synthesis, 260
7.5.5 Physical Design and Verification, 260
Key Points, 261
Acknowledgments, 263
References, 263

8 Circuits for Optical-to-Electrical Conversion 265

8.1 Introduction, 265
8.2 Optical to Electrical-to-Optical Conversion, 266
8.2.1 Principle of Operation, 266
8.2.2 Optical Transceiver Architectures, 267
8.2.3 Integrated Circuit Technology for Optical Transceivers, 268
8.3 Signal Amplification, 272
8.3.1 Trans-Impedance Amplifier, 272
8.3.2 Limited Amplifier, 273
8.3.3 Laser Driver, 274
8.4 Phase-Locked Loop, 275
8.4.1 Phase-Locked-Loop Architecture, 275
8.4.2 Voltage-Controlled Oscillator, 275
8.4.3 Phase and Frequency Detectors, 278

8.5 Clock Synthesis and Recovery, 279
8.5.1 Clock Synthesis, 279
8.5.2 Clock and Data Recovery, 283
8.5.3 Jitter Requirements, 285
8.6 Preemphasis and Equalization, 287
8.6.1 High-Speed Signal Impairments, 287
8.6.2 Preemphasis, 289
8.6.3 Equalization, 289
Key Points, 290
References, 290

PART IV DATA SWITCHING

9 Physical Circuit Switching 293

9.1 Introduction, 293
9.2 Switching and Why It Is Important, 294
9.3 Three Types of Switching, 298
9.3.1 Switching of Physical Circuits, 298
9.3.2 Switching of Time-Division-Multiplexed Signals, 299
9.3.3 Switching of Cell and/or Packets, 300
9.4 Quality of Service, 300
9.5 Special Services, 302
9.6 Switching in One or More Stages, 303
9.7 Cost Model for Switch Implementations, 304
9.8 Crossbar Switch Concept, 305
9.9 Optical Crossbar Switches, 308
9.10 Digital Electronic Crossbar Switches, 310
9.10.1 Control of Digital Crossbar Switches, 314
9.10.2 Cost Model for Digital Electronic Crossbar Switches, 315
9.10.3 Growth Limits of Digital Electronic Crossbar Switches, 317
9.10.4 Commercial Examples of Electronic Crossbar Switches, 317
9.11 Multistage Crossbar-Based Switches, 318
9.11.1 Routing and Blocking in Clos Networks, 321
9.11.2 Multicast in Clos Networks, 332
9.11.3 Implementation Costs of Clos Networks, 338
9.12 Desirability of Single-Stage Fabrics and Limits to Multistage Fabrics, 339
Key Points, 340
References, 341

10 Time-Division-Multipleured Switching 343

10.1 Introduction, 343
10.2 TDM Review, 344
10.3 TDM Switching Problem, 346
10.3.1 Temporal Alignment, 348
10.3.2 Dual Control Pages, 349
10.3.3 Strictly Nonblocking Design, 349
10.3.4 Varying Port Configurations, 350
10.4 Central Memory TDM Switches, 350
10.4.1 Cost Model for Central Memory TDM Switches, 352
10.4.2 Limits of Central Memory Design, 353
10.5 Ingress Buffered TDM Switches, 355
10.6 Egress Buffered Self-Select TDM Switches, 357
10.7 Sliced Single-Stage SNB TDM Fabrics, 358
10.8 Time–Space Multistage TDM Fabrics, 362
10.8.1 Architecture and Costs of Time–Space–Time Switch Fabrics, 365
10.8.2 Blocking and OPA, 366
10.8.3 Space–Time–Space Switching, 374
10.9 Multistage Memory Switches, 377
10.10 Summary, 379
Key Points, 380
References, 381

11 Packet and Cell Switching and Queuing 383

11.1 Introduction, 384
11.2 Packet–Cell Switching Problem, 384
11.3 Traffic Patterns, 386
11.3.1 Realistic Loads, 387
11.3.2 Responses to Congestion, 388
11.4 Logical Queue Structures and Their Behavior, 389
11.4.1 Queues, 389
11.4.2 Flows and Logical Queues, 390
11.4.3 Queuing Systems, 390
11.4.4 Speedup and Blocking, 392
11.4.5 Possible Queuing Systems, 395
11.5 Queue Locations and Buffer Sharing, 400
11.6 Filling and Draining Queues, 404
11.6.1 Filling, 404
11.6.2 Draining, 405
11.7 Central Memory Packet–Cell Switches, 406
11.8 Ingress Buffered Packet–Cell Switches, 410