
Contents

List of Contributors	xi
1 Architectures of Internet Switches and Routers	1
Xin Li, Lotfi Mhamdi, Jing Liu, Konghong Pun, and Mounir Hamdi	
1.1 Introduction	2
1.2 Bufferless Crossbar Switches	3
1.2.1 Introduction to Switch Fabrics	3
1.2.2 Output-queued Switches	4
1.2.3 Input-queued Switches	4
1.2.4 Scheduling Algorithms for VOQ Switches	5
1.2.5 Combined Input–Ouput-queued Switches	9
1.3 Buffered Crossbar Switches	12
1.3.1 Buffered Crossbar Switches Overview	12
1.3.2 The VOQ/BCS Architecture	13
1.4 Multi-stage Switching	19
1.4.1 Architecture Choice	19
1.4.2 The MSM Clos-network Architecture	20
1.4.3 The Bufferless Clos-network Architecture	23
1.5 Optical Packet Switching	27
1.5.1 Multi-rack Hybrid Opto-electronic Switch Architecture	27
1.5.2 Optical Fabrics	28
1.5.3 Reduced Rate Scheduling	30
1.5.4 Time Slot Assignment Approach	30
1.5.5 DOUBLE Algorithm	32
1.5.6 ADJUST Algorithm	32
1.6 Conclusion	34
2 Theoretical Performance of Input-queued Switches Using Lyapunov Methodology	39
Andrea Bianco, Paolo Giaccone, Emilio Leonardi, Marco Mellia, and Fabio Neri	
2.1 Introduction	39
2.2 Theoretical Framework	41

2.2.1	Description of the Queueing System	41
2.2.2	Stability Definitions for a Queueing System	43
2.2.3	Lyapunov Methodology	44
2.2.4	Lyapunov Methodology to Bound Queue Sizes and Delays	47
2.2.5	Application to a Single Queue	48
2.2.6	Final Remarks	49
2.3	Performance of a Single Switch	50
2.3.1	Stability Region of Pure Input-queued Switches	51
2.3.2	Delay Bounds for Maximal Weight Matching	54
2.3.3	Stability Region of CIOQ with Speedup 2	55
2.3.4	Scheduling Variable-size Packets.....	57
2.4	Networks of IQ Switches.....	58
2.4.1	Theoretical Performance.....	59
2.5	Conclusions	61
3	Adaptive Batched Scheduling for Packet Switching with Delays	65
	Kevin Ross and Nicholas Bambos	
3.1	Introduction	65
3.2	Switching Modes with Delays: A General Model	66
3.3	Batch Scheduling Algorithms.....	69
3.3.1	Fixed Batch Policies	70
3.3.2	Adaptive Batch Policies.....	72
3.3.3	The Simple-batch Static Schedule.....	73
3.4	An Interesting Application: Optical Networks	74
3.5	Throughput Maximization via Adaptive Batch Schedules	76
3.6	Summary	78
4	Geometry of Packet Switching: Maximal Throughput Cone Scheduling Algorithms	81
	Kevin Ross and Nicholas Bambos	
4.1	Introduction	81
4.2	Backlog Dynamics of Packet Switches.....	84
4.3	Switch Throughput and Rate Stability	86
4.4	Cone Algorithms for Packet Scheduling.....	88
4.4.1	Projective Cone Scheduling (PCS).....	89
4.4.2	Relaxation, Generalizations, and Delayed PCS (D-PCS).....	90
4.4.3	Argument Why PCS and D-PCS Maximize Throughput	92
4.4.4	Quality of Service and Load Balancing.....	93
4.5	Complexity in Cone Schedules – Scalable PCS Algorithms	95
4.5.1	Approximate PCS.....	95
4.5.2	Local PCS.....	95
4.6	Final Remarks	98
5	Fabric on a Chip: A Memory-management Perspective	101
	Itamar Elhanany, Vahid Tabatabaee, and Brad Matthews	
5.1	Introduction	101
5.1.1	Benefits of the Fabric-on-a-Chip Approach	102
5.2	Emulating an Output-queued Switch	103

5.3	Packet Placement Algorithm	105
5.3.1	Switch Architecture	105
5.3.2	Memory-management Algorithm and Related Resources	106
5.3.3	Sufficiency Condition on the Number of Memories.....	109
5.4	Implementation Considerations	114
5.4.1	Logic Dataflow.....	114
5.4.2	FPGA Implementation Results	119
5.5	Conclusions	120
6	Packet Switch with Internally Buffered Crossbars.....	121
	Zhen Guo, Roberto Rojas-Cessa, and Nirwan Ansari	
6.1	Introduction to Packet Switches.....	121
6.2	Crossbar-based Switches	122
6.3	Internally Buffered Crossbars	124
6.4	Combined Input–Crosspoint Buffered (CICB) Crossbars	126
6.4.1	FIFO–CICO Switches	126
6.4.2	VOQ–CICB Switches	128
6.4.3	Separating Matching into Input and Output Arbitrations	130
6.4.4	Weighted Arbitration Schemes.....	130
6.4.5	Arbitration Schemes based on Round-robin Selection	135
6.5	CICB Switches with Internal Variable-length Packets	141
6.6	Output Emulation by CICB Switches	141
6.7	Conclusions	144
7	Dual Scheduling Algorithm in a Generalized Switch: Asymptotic Optimality and Throughput Optimality.....	147
	Lijun Chen, Steven H. Low, and John C. Doyle	
7.1	Introduction	148
7.2	System Model	150
7.2.1	Queue Length Dynamics	151
7.2.2	Dual Scheduling Algorithm	152
7.3	Asymptotic Optimality and Fairness	153
7.3.1	An Ideal Reference System	153
7.3.2	Stochastic Stability	154
7.3.3	Asymptotic Optimality and Fairness	155
7.4	Throughput-optimal Scheduling	159
7.4.1	Throughput Optimality and Fairness	159
7.4.2	Optimality Proof	160
7.4.3	Flows with Exponentially Distributed Size	163
7.5	A New Scheduling Architecture	165
7.6	Conclusions	166
8	The Combined Input and Crosspoint Queued Switch.....	169
	Kenji Yoshigoe and Ken Christensen	
8.1	Introduction	169
8.2	History of the CICQ Switch	172
8.3	Performance of CICQ Cell Switching	175
8.3.1	Traffic Models	176

8.3.2	Simulation Experiments	177
8.4	Performance of CICQ Packet Switching	179
8.4.1	Traffic Models	179
8.4.2	Simulation Experiments	179
8.5	Design of Fast Round-robin Arbiters	181
8.5.1	Existing RR Arbiter Designs	182
8.5.2	A New Short-term Fair RR Arbiter – The Masked Priority Encoder (MPE)	183
8.5.3	A New Fast Long-term Fair RR Arbiter – The Overlapped RR (ORR) Arbiter	186
8.6	Future Directions – The CICQ with VCQ	188
8.6.1	Design of Virtual Crosspoint Queueing (VCQ)	189
8.6.2	Evaluation of CICQ Cell Switch with VCQ	190
8.7	Summary	192
9	Time–Space Label Switching Protocol (TSL-SP)	197
	Anpeng Huang, Biswanath Mukherjee, Linzhen Xie, and Zhengbin Li	
9.1	Introduction	197
9.2	Time Label	198
9.3	Space Label	200
9.4	Time–Space Label Switching Protocol (TSL-SP)	201
9.5	Illustrative Results	205
9.6	Summary	209
10	Hybrid Open Hash Tables for Network Processors	211
	Dale Parson, Qing Ye, and Liang Cheng	
10.1	Introduction	211
10.2	Conventional Hash Algorithms	213
10.2.1	Chained Hash Tables	214
10.2.2	Open Hash Tables	215
10.3	Performance Degradation Problem	216
10.3.1	Improvements	218
10.4	Hybrid Open Hash Tables	219
10.4.1	Basic Operations	219
10.4.2	Basic Ideas	219
10.4.3	Performance Evaluation	220
10.5	Hybrid Open Hash Table Enhancement	222
10.5.1	Flaws of Hybrid Open Hash Table	222
10.5.2	Dynamic Enhancement	223
10.5.3	Adaptative Enhancement	224
10.5.4	Timeout Enhancement	224
10.5.5	Performance Evaluation	224
10.6	Extended Discussions of Concurrency Issues	225
10.6.1	Insertion	225
10.6.2	Clean-to-copy Phase Change	226
10.6.2	Timestamps.....	227
10.7	Conclusion	227
Index	229