
Contents

Dedication	xvii
Foreword	xix
Preface	xxi
Acknowledgements	xxiii
PART I Introduction	1
1 Introduction to SDN	3
<i>Ruslan L. Smelyanskiy and Alexander Shalimov</i>	
1.1 Data centers	3
1.1.1 The new computing paradigm	3
1.1.2 DC network architecture	5
1.1.3 Traffic in DC	5
1.1.4 Addressing and routing in DC	7
1.1.5 Performance	8
1.1.6 TCP/IP stack issues	10
1.1.7 Network management system	11
1.1.8 Virtualization, scalability, flexibility	12
1.2 Software-defined networks	13
1.2.1 How can we split control plane and data plane?	13
1.2.2 OpenFlow protocol and programmable switching: basics	16
1.2.3 SDN controller, northbound API, controller applications	19
1.2.4 Open issues and challenges	22
1.3 Summary and conclusion	22
References	23
2 SDN implementations and protocols	27
<i>Cristian Hernandez Benet, Kyoomars Alizadeh Noghani, and Javid Taheri</i>	
2.1 How SDN is implemented	28
2.1.1 Implementation aspects	28
2.1.2 Existing SDN controllers	29
2.2 Current SDN implementation using OpenDaylight	30
2.2.1 OpenDaylight	30
2.3 Overview of OpenFlow devices	33
2.3.1 Software switches	34
2.3.2 Hardware switches	35

2.4	SDN protocols	36
2.4.1	ForCES	36
2.4.2	OpenFlow	37
2.4.3	Open vSwitch database management (OVSDB)	41
2.4.4	OpenFlow configuration and management protocol (OF-CONFIG)	42
2.4.5	Network configuration protocol (NETCONF)	43
2.5	Open issues and challenges	44
2.6	Summary and Conclusions	45
	References	46
3	SDN components and OpenFlow	49
	<i>Yanbiao Li, Dafang Zhang, Javid Taheri, and Keqin Li</i>	
3.1	Overview of SDN's architecture and main components	49
3.1.1	Comparison of IP and SDN in architectures	50
3.1.2	SDN's main components	51
3.2	OpenFlow	52
3.2.1	Fundamental abstraction and basic concepts	52
3.2.2	OpenFlow tables and the forwarding pipeline	54
3.2.3	OpenFlow channels and the communication mechanism	55
3.3	SDN controllers	57
3.3.1	System architectural overview	57
3.3.2	System implementation overview	59
3.3.3	Rule placement and optimization	60
3.4	OpenFlow switches	60
3.4.1	The detailed working flow	60
3.4.2	Design and optimization of table lookups	62
3.4.3	Switch designs and implementations	63
3.5	Open issues in SDN	65
3.5.1	Resilient communication	65
3.5.2	Scalability	65
	References	66
4	SDN for cloud data centres	69
	<i>Dimitrios Pezaros, Richard Cziva, and Simon Jouet</i>	
4.1	Overview	69
4.2	Cloud data centre topologies	70
4.2.1	Conventional architectures	70
4.2.2	Clos/Fat-Tree architectures	71
4.2.3	Server-centric architectures	73
4.2.4	Management network	75
4.3	Software-defined networks for cloud data centres	76
4.3.1	Challenges in cloud DC networks	76
4.3.2	Benefits of using SDN in cloud DCs	77

4.3.3	Current SDN deployments in cloud DC	79
4.3.4	SDN as the backbone for a converged resource control plane	80
4.4	Open issues and challenges	82
4.4.1	Network function virtualisation and SDN in DCs	82
4.4.2	The future of network programmability	83
4.5	Summary	85
	Acknowledgements	85
	References	86
5	Introduction to big data	91
	<i>Amir H. Payberah and Fatemeh Rahimian</i>	
5.1	Big data platforms: challenges and requirements	91
5.2	How to store big data?	93
5.2.1	Distributed file systems	94
5.2.2	Messaging systems	95
5.2.3	NoSQL databases	96
5.3	How to process big data?	99
5.3.1	Batch data processing platforms	99
5.3.2	Streaming data processing platforms	102
5.3.3	Graph data processing platforms	107
5.3.4	Structured data processing platforms	110
5.4	Concluding remarks	111
	References	112
6	Big Data processing using Apache Spark and Hadoop	115
	<i>Koichi Shirahata and Satoshi Matsuoka</i>	
6.1	Introduction	115
6.2	Big Data processing	117
6.2.1	Big Data processing models	118
6.2.2	Big Data processing implementations	119
6.2.3	MapReduce-based Big Data processing implementations	120
6.2.4	Computing platforms for Big Data processing	122
6.3	Apache Hadoop	123
6.3.1	Overview of Hadoop	123
6.3.2	Hadoop MapReduce	124
6.3.3	Hadoop distributed file system	125
6.3.4	YARN	126
6.3.5	Hadoop libraries	127
6.3.6	Research activities on Hadoop	128
6.4	Apache Spark	129
6.4.1	Overview of Spark	129
6.4.2	Resilient distributed dataset	129

6.4.3	Spark libraries	130
6.4.4	Using both Spark and Hadoop cooperatively	131
6.4.5	Research activities on Spark	132
6.5	Open issues and challenges	132
6.5.1	Storage	132
6.5.2	Computation	133
6.5.3	Network	134
6.5.4	Data analysis	135
6.6	Summary	136
	References	136
7	Big Data stream processing	139
	<i>Yidan Wang, M. Reza HoseinyFarahabady, Zahir Tari, and Albert Y. Zomaya</i>	
7.1	Introduction to stream processing	139
7.1.1	Background and motivation	139
7.1.2	Streamlined data processing framework	140
7.1.3	Stream processing systems	141
7.2	Apache storm [8, 9]	143
7.2.1	Reading path	143
7.2.2	Storm structure and composing components	143
7.2.3	Data stream and topology	144
7.2.4	Parallelism of topology	145
7.2.5	Grouping strategies	146
7.2.6	Reliable message processing	147
7.3	Scheduling and resource allocation in Apache Storm	148
7.3.1	Scheduling and resource allocation in cloud [4–7]	148
7.3.2	Scheduling of Apache Storm [8, 9]	149
7.3.3	Advanced scheduling schemes for Storm	150
7.4	Quality-of-service-aware scheduling	151
7.4.1	Performance metrics [16]	151
7.4.2	Model predictive control-based scheduling	152
7.4.3	Experimental performance analysis	153
7.5	Open issues in stream processing	155
7.6	Conclusion	156
	Acknowledgement	156
	References	157
8	Big Data in cloud data centers	159
	<i>Gunasekaran Manogaran and Daphne Lopez</i>	
8.1	Introduction	159
8.2	Needs for the architecture patterns and data sources for Big Data storage in cloud data centers	160

8.3	Applications of Big Data analytics with cloud data centers	162
8.3.1	Disease diagnosis	162
8.3.2	Government organizations	163
8.3.3	Social networking	163
8.3.4	Computing platforms	163
8.3.5	Environmental and natural resources	163
8.4	State-of-the-art Big Data architectures for cloud data centers	163
8.4.1	Lambda architecture	164
8.4.2	NIST Big Data Reference Architecture (NBDRA)	166
8.4.3	Big Data Architecture for Remote Sensing	167
8.4.4	The Service-On Line-Index-Data (SOLID) architecture	169
8.4.5	Semantic-based Architecture for Heterogeneous Multimedia Retrieval	170
8.4.6	LargeScale Security Monitoring Architecture	171
8.4.7	Modular software architecture	172
8.4.8	MongoDB-based Healthcare Data Management Architecture	173
8.4.9	Scalable and Distributed Architecture for Sensor Data Collection, Storage and Analysis	174
8.4.10	Distributed parallel architecture for “Big Data”	176
8.5	Challenges and potential solutions for Big Data analytics in cloud data centers	177
8.6	Conclusion	180
	References	181

PART II How SDN helps Big Data 183

9 SDN helps volume in Big Data 185

*Kyoomars Alizadeh Noghani, Cristian Hernandez Benet,
and Javid Taheri*

9.1	Big Data volume and SDN	186
9.2	Network monitoring and volume	187
9.2.1	Legacy traffic monitoring solutions	188
9.2.2	SDN-based traffic monitoring	189
9.3	Traffic engineering and volume	191
9.3.1	Flow scheduling	192
9.3.2	TCP incast	196
9.3.3	Dynamically change network configuration	197
9.4	Fault tolerant and volume	198
9.5	Open issues	201
9.5.1	Scalability	202
9.5.2	Resiliency and reliability	202
9.5.3	Conclusion	202
	References	203

10 SDN helps velocity in Big Data	207
<i>Van-Giang Nguyen, Anna Brunstrom, Karl-Johan Grinnemo, and Javid Taheri</i>	
10.1 Introduction	208
10.1.1 Big Data velocity	208
10.1.2 Type of processing	208
10.2 How SDN can help velocity?	211
10.3 Improving batch processing performance with SDN	212
10.3.1 FlowComb	212
10.3.2 Pythia	213
10.3.3 Bandwidth-aware scheduler	214
10.3.4 Phurti	215
10.3.5 Cormorant	216
10.3.6 SDN-based Hadoop for social TV analytics	217
10.4 Improving real-time and stream processing performance with SDN	218
10.4.1 Firebird	218
10.4.2 Storm-based NIDS	219
10.4.3 Crosslayer scheduler	220
10.5 Summary	221
10.5.1 Comparison table	221
10.5.2 Generic SDN-based Big Data processing framework	221
10.6 Open issues and research directions	223
10.7 Conclusion	225
References	225
11 SDN helps value in Big Data	229
<i>Harald Gjermundrød</i>	
11.1 Private centralized infrastructure	232
11.1.1 Adaptable network platform	232
11.1.2 Adaptable data flows and application deployment	233
11.1.3 Value of dark data	233
11.1.4 New market for the cloud provider	235
11.2 Private distributed infrastructure	236
11.2.1 Adaptable resource allocation	236
11.2.2 Value of dark data	238
11.3 Public centralized infrastructure	238
11.3.1 Adaptable data flows and programmable network	238
11.3.2 Usage of dark data	240
11.3.3 Data market	240
11.4 Public distributed infrastructure	242
11.4.1 Usage of dark data	242
11.4.2 Data market	243
11.4.3 Data as a service	247

11.5 Open issues and challenges	247
11.6 Chapter summary	249
References	249
12 SDN helps other Vs in Big Data	253
<i>Pradeeban Kathiravelu and Luís Veiga</i>	
12.1 Introduction to other Vs in Big Data	254
12.1.1 Variety in Big Data	254
12.1.2 Volatility in Big Data	255
12.1.3 Validity and veracity in Big Data	256
12.1.4 Visibility in Big Data	256
12.2 SDN for other Vs of Big Data	257
12.2.1 SDN for variety of data	258
12.2.2 SDN for volatility of data	259
12.2.3 SDN for validity and veracity of data	261
12.2.4 SDN for visibility of data	262
12.2.5 More Vs into Big Data	263
12.3 SDN for Big Data diversity	264
12.3.1 Use cases for SDN in heterogeneous Big Data	264
12.3.2 Architectures for variety and quality of data	265
12.3.3 QoS-aware Big Data applications	266
12.3.4 Multitenant SDN and data isolation	267
12.4 Open issues and challenges	268
12.4.1 Scaling Big Data with SDN	268
12.4.2 Scaling Big Data beyond data centers	270
12.5 Summary and conclusion	270
References	271
13 SDN helps Big Data to optimize storage	275
<i>Ali R. Butt, Ali Anwar, and Yue Cheng</i>	
13.1 Software defined key-value storage systems for datacenter applications	275
13.2 Related work, features, and shortcomings	276
13.2.1 Shortcomings	277
13.3 SDN-based efficient data management	280
13.4 Rules of thumb of storage deployment in software defined datacenters	281
13.4.1 Summary of rules-of-thumb	285
13.5 Experimental analysis	286
13.5.1 Evaluating data management framework in software defined datacenter environment	286
13.5.2 Evaluating micro-object-store architecture in software defined datacenter environment	289

13.6	Open issue and future directions in SDN-enabled Big Data management	292
13.6.1	Open issues in data management framework in software defined datacenter	292
13.6.2	Open issues in micro-object-store architecture in software defined datacenter environment	293
13.7	Summary	294
	References	294
14	SDN helps Big Data to optimize access to data	297
	<i>Yuankun Fu and Fengguang Song</i>	
14.1	Introduction	297
14.2	State of the art and related work	299
14.3	Performance analysis of message passing and parallel file system I/O	300
14.4	Analytical modeling-based end-to-end time optimization	302
14.4.1	The problem	302
14.4.2	The traditional method	303
14.4.3	Improved version of the traditional method	303
14.4.4	The fully asynchronous pipeline method	304
14.4.5	Microbenchmark for the analytical model	305
14.5	Design and implementation of DataBroker for the fully asynchronous method	309
14.6	Experiments with synthetic and real applications	310
14.6.1	Synthetic and real-world applications	310
14.6.2	Accuracy of the analytical model	311
14.6.3	Performance speedup	312
14.7	Open issues and challenges	314
14.8	Conclusion	315
	Acknowledgments	315
	References	315
15	SDN helps Big Data to become fault tolerant	319
	<i>Abdelmounaam Rezgui, Kyoomars Alizadeh Noghani, Javid Taheri, Amir Mirzaeinia, Hamdy Soliman, and Nickolas Davis</i>	
15.1	Big Data workloads and cloud data centers	320
15.2	Network architectures for cloud data centers	321
15.2.1	Switch-centric data centers	321
15.2.2	Server-centric data centers	321
15.3	Fault-tolerant principles	324
15.4	Traditional approaches to fault tolerance in data centers	325
15.4.1	Reactive approaches	326
15.4.2	Proactive approaches	327
15.4.3	Problems with legacy fault-tolerant solutions	327

15.5	Fault tolerance in SDN-based data centers	328
15.5.1	Failure detection in SDN	329
15.5.2	Failure recovery in SDN	329
15.6	Reactive fault-tolerant approach in SDN	330
15.7	Proactive fault-tolerant approach in SDN	330
15.7.1	Failure prediction in cloud data centers	332
15.7.2	Traffic patterns of Big Data workloads	332
15.8	Open issues and challenges	333
15.8.1	Problems with SDN-based fault-tolerant methods	333
15.8.2	Fault tolerance in the control plane	334
15.9	Summary and conclusion	334
	References	334
PART III How Big Data helps SDN		337
16	How Big Data helps SDN with data protection and privacy	339
	<i>Lothar Fritsch</i>	
16.1	Collection and processing of data to improve performance	339
16.1.1	The promise of Big Data in SDN: data collection, analysis, configuration change	339
16.2	Data protection requirements and their implications for Big Data in SDN	340
16.2.1	Data protection requirements in Europe	340
16.2.2	Personal data in networking information	343
16.2.3	Issues with Big Data processing	344
16.3	Recommendations for privacy design in SDN Big Data projects	344
16.3.1	Storage concepts	345
16.3.2	Filtration, anonymization and data minimization	345
16.3.3	Privacy-friendly data mining	346
16.3.4	Purpose-binding and obligations management	346
16.3.5	Data subject consent management techniques	347
16.3.6	Algorithmic accountability concepts	347
16.3.7	Open issues for protecting privacy using Big Data and SDN	349
16.4	Conclusion	350
	Acknowledgment	350
	References	350
17	Big Data helps SDN to detect intrusions and secure data flows	353
	<i>Li-Chun Wang and Yu-Jia Chen</i>	
17.1	Introduction	353
17.2	Security issues of SDN	354
17.2.1	Security issues in control channel	354
17.2.2	Denial-of-service (DoS) attacks	354

17.2.3 Simulation of control channel attack on SDN	357
17.3 Big Data techniques for security threats in SDN	359
17.3.1 Big Data analytics	360
17.3.2 Data analytics for threat detection	361
17.4 QoS consideration in SDN with security services	361
17.4.1 Delay guarantee for security traversal	361
17.4.2 Traffic load balancing	365
17.5 Big Data applications for securing SDN	368
17.5.1 Packet inspection	368
17.6 Open issues and challenge	371
17.7 Summary and conclusion	371
References	372

18 Big Data helps SDN to manage traffic 375

Jianwu Wang and Qiang Duan

Abstract	375
18.1 Introduction	375
18.2 State of art of traffic management in IP and SDN networks	377
18.2.1 General concept and procedure of network traffic management	377
18.2.2 Traffic management in IP networks	378
18.2.3 Traffic management in SDN networks	379
18.3 Potential benefits for traffic management in SDN using Big Data techniques	381
18.3.1 Big Data in SDN networks	381
18.3.2 How Big Data analytics could help SDN networks	382
18.4 A framework for Big Data-based SDN traffic management	382
18.5 Possible Big Data applications for SDN traffic analysis and control	384
18.5.1 Big graph data analysis for SDN traffic analysis and long-term network topology improvement	384
18.5.2 Streaming-based Big Data analysis for real-time SDN traffic analysis and adaptation	384
18.5.3 Big Data mining for SDN network control and adaptation	385
18.6 Open issues and challenges	385
18.6.1 Data acquisition measurement and overhead	385
18.6.2 SDN controller management	386
18.6.3 New system architecture for Big Data-based traffic management in SDN	386
18.7 Conclusion	386
References	387

19 Big Data helps SDN to optimize its controllers	389
<i>Daewoong Cho, Saeed Bastani, Javid Taheri, and Albert Y. Zomaya</i>	
19.1 Introduction	389
19.2 What is a SDN controller?	390
19.3 SDN controller-related issues	391
19.3.1 Scalability	391
19.3.2 Resiliency	392
19.3.3 Solutions	393
19.4 Big Data for SDN controller optimization	394
19.4.1 System architecture	395
19.4.2 Big Data analytics techniques	395
19.4.3 Problem formulation	396
19.4.4 Optimization algorithm	398
19.4.5 Applicable scenarios	399
19.5 Open issues and challenges	404
19.6 Conclusion	405
References	405
20 Big Data helps SDN to verify integrity of control/data planes	409
<i>Qingsong Wen, Ren Chen, Yinglong Xia, Li Zhou, Juan Deng, Jian Xu, and Mingzhen Xia</i>	
20.1 Introduction	409
20.2 Related work	410
20.3 Finding top- K shortest simple paths	410
20.3.1 MPS algorithm for top K shortest simple paths	411
20.3.2 Improved MPS algorithm with efficient implementation	413
20.4 Routing check and detection	416
20.4.1 Subnet partition	417
20.4.2 Loop detection	418
20.4.3 Black hole detection	418
20.4.4 Reachability detection	419
20.5 Efficient graph engine	419
20.5.1 Edge-set representation	420
20.5.2 Consolidation	421
20.5.3 Multimodal organization	423
20.5.4 Scheduling and prefetching	423
20.6 Experiments	423
20.6.1 Performance evaluation of finding top- K shortest simple paths	423
20.6.2 Performance evaluation of the efficient graph engines	426
20.7 Open issues and challenges	428
20.8 Conclusions	429
References	429

21 Big Data helps SDN to improve application specific quality of service	433
<i>Susanna Schwarzmann, Andreas Blenk, Ognjen Dobrijevic, Michael Jarschel, Andreas Hotho, Thomas Zinner, and Florian Wamser</i>	
21.1 Introduction	433
21.2 Classification of SDN-based context-aware networking approaches	434
21.2.1 Monitoring of QoE influence factors (QoE-IFs)	435
21.2.2 Control actions of management approaches	436
21.2.3 Potential of Big Data for SDN QoE management	437
21.3 Big Data analytics to support QoS/QoE management	438
21.3.1 Big Data analytics	438
21.3.2 Current and ongoing work	440
21.4 Combining Big Data analytics and SDN: three use cases to improve QoS/QoE	442
21.4.1 Use case 1: improving the operation of networks	442
21.4.2 Use case 2: improving the quality of video-on-demand streaming based on business agreements	444
21.4.3 Use case 3: improving the quality of applications without business agreements	445
21.5 Vision: intelligent network-wide auto-optimization	446
21.6 Challenges and discussions	449
21.6.1 Challenges of SDN-based QoE management	449
21.6.2 Challenges of a Big Data-supported SDN architecture for enhancing application quality	450
21.7 Conclusion	452
Acknowledgments	452
References	453
Index	457