Next-Generation Internet Architectures and Protocols

With ever-increasing demands on capacity, quality of service, speed, and reliability, current Internet systems are under strain and under review. Combining contributions from experts in the field, this book captures the most recent and innovative designs, architectures, protocols, and mechanisms that will enable researchers to successfully build the next-generation Internet. A broad perspective is provided, with topics including innovations at the physical/transmission layer in wired and wireless media, as well as the support for new switching and routing paradigms at the device and sub-system layer. The proposed alternatives to TCP and UDP at the data transport layer for emerging environments are also covered, as are the novel models and theoretical foundations proposed for understanding network complexity. Finally, new approaches for pricing and network economics are discussed, making this ideal for students, researchers, and practitioners who need to know about designing, constructing, and operating the next-generation Internet.

Byrav Ramamurthy is an Associate Professor in the Department of Computer Science and Engineering at the University of Nebraska, Lincoln (UNL). He is a recipient of the UNL CSE Department Student Choice Outstanding Teaching Award, the author of Design of Optical WDM Networks (2000), and a co-author of Secure Group Communications over Data Networks (2004). His research areas include optical networks, wireless/sensor networks, network security, distributed computing, and telecommunications.

George N. Rouskas is a Professor of Computer Science at North Carolina State University. He has received several research and teaching awards, including an NSF CAREER Award, the NCSU Alumni Outstanding Research Award, and he has been inducted into the NCSU Academy of Outstanding Teachers. He is the author of the book Internet Tiered Services (2009) and his research interests are in network architectures and protocols, optical networks and performance evaluation.

Krishna Moorthy Sivalingam is a Professor in the Department of Computer Science and Engineering at the Indian Institute of Technology (IIT), Madras. Prior to this, he was a Professor at the University of Maryland, Baltimore County, and also conducted research at Lucent Technologies and AT&T Bell Labs. He has previously edited three books and holds three patents in wireless networking. His research interests include wireless networks, wireless sensor networks, optical networks, and performance evaluation.
Next-Generation Internet: Architectures and Protocols

Edited by

BYRAV RAMAMURTHY
University of Nebraska-Lincoln

GEORGE ROUSKAS
North Carolina State University

KRISHNA M. SIVALINGAM
Indian Institute of Technology Madras
To my mother, Mrs. Lalitha Ramamurthy – BR
To Magdalini and Alexander – GNR
To my family – KMS
Contents

Contributors xvi
Preface xix

Part I Enabling technologies 1

1 Optical switching fabrics for terabit packet switches 3
Davide Cuda, Roberto Gaudino, Guido A. Gavilanes Castillo, and Fabio Neri

1.1 Optical switching fabrics 5
   1.1.1 Wavelength-selective (WS) architecture 7
   1.1.2 Wavelength-routing (WR) architecture 8
   1.1.3 Plane-switching (PS) architecture 9

1.2 Modeling optical devices 10
   1.2.1 Physical model 11
   1.2.2 Device characterization 12
   1.2.3 Multi-plane-specific issues 15

1.3 Scalability analysis 16

1.4 Cost analysis 18

1.5 Results 21
   1.5.1 Scalability of the aggregate switching bandwidth 21
   1.5.2 CAPEX estimation 23

1.6 Conclusions 24

References 25

2 Broadband access networks: current and future directions 27
Abu (Sayeem) Reaz, Lei Shi, and Biswanath Mukherjee

2.1 Introduction 27
   2.1.1 Current broadband access solutions 27
   2.1.2 Passive Optical Network (PON) 28
   2.1.3 Extending the reach: Long-Reach PON (LR-PON) 30

2.2 Technologies and demonstrations 32
   2.2.1 Enabling technologies 32
   2.2.2 Demonstrations of LR-PON 33
## 2.3 Research challenges in LR-PON

2.3.1 Low-cost devices: colorless ONU

2.3.2 Resource allocation: DBA with Multi-Thread Polling

2.3.3 Traffic management: behavior-aware user assignment

## 2.4 Reaching the end-users: Wireless-Optical Broadband Access Network (WOBAN)

2.4.1 WOBAN architecture

2.4.2 Motivation of WOBAN

2.4.3 Research challenges in WOBAN

## 2.5 Conclusion

---

## 3 The optical control plane and a novel unified control plane architecture for IP/WDM networks

Georgios Ellinas, Antonis Hadjiantonis, Ahmad Khalil, Neophytos Antoniades, and Mohamed A. Ali

3.1 Introduction

3.2 Overview of optical control plane design

3.2.1 Link Management Protocol

3.2.2 GMPLS routing protocol

3.2.3 GMPLS signaling protocol

3.3 IP-over-WDM networking architecture

3.3.1 The overlay model

3.3.2 The peer and augmented models

3.4 A new approach to optical control plane design: an optical layer-based unified control plane architecture

3.4.1 Node architecture for the unified control plane

3.4.2 Optical layer-based provisioning

3.5 Conclusions

---

## 4 Cognitive routing protocols and architecture

Suyang Ju and Joseph B. Evans

4.1 Introduction

4.2 Mobility-aware routing protocol

4.2.1 Background

4.2.2 Approach

4.2.3 Benefits

4.2.4 Protocol architecture

4.3 Spectrum-aware routing protocol

4.3.1 Background

4.3.2 Approach
5 Grid networking

Anusha Ravula and Byrav Ramamurthy

5.1 Introduction
5.2 The Grid
  5.2.1 Grid Computing
  5.2.2 Lambda Grid networks
5.3 Cloud Computing
5.4 Resources
  5.4.1 Grid network resources
  5.4.2 Optical network testbeds and projects
  5.4.3 Computational resources
  5.4.4 Other resources
5.5 Scheduling
5.6 Optical Circuit Switching and Optical Burst Switching
  5.6.1 Studies on OCS-based Grids
  5.6.2 Studies on OBS based Grids
5.7 Conclusion

References

Part II Network architectures

6 Host identity protocol (HIP): an overview
Pekka Nikander, Andrei Gurtov, and Thomas R. Henderson

6.1 Introduction
6.2 Fundamental problems in the Internet today
  6.2.1 Loss of universal connectivity
  6.2.2 Poor support for mobility and multi-homing
  6.2.3 Unwanted traffic
  6.2.4 Lack of authentication, privacy, and accountability
6.3 The HIP architecture and base exchange
  6.3.1 Basics
  6.3.2 HITs and LSIs
  6.3.3 Protocols and packet formats
  6.3.4 Detailed layering
  6.3.5 Functional model
  6.3.6 Potential drawbacks
6.4 Mobility, multi-homing, and connectivity  
   6.4.1 HIP-based basic mobility and multi-homing 121  
   6.4.2 Facilitating rendezvous 122  
   6.4.3 Mobility between addressing realms and through NATs 123  
   6.4.4 Subnetwork mobility 124  
   6.4.5 Application-level mobility 126  
6.5 Privacy, accountability, and unwanted traffic 126  
   6.5.1 Privacy and accountability 126  
   6.5.2 Reducing unwanted traffic 127  
6.6 Current status of HIP 129  
6.7 Summary 131

References 131

7 Contract-Switching for Managing Inter-Domain Dynamics 136  
Murat Yuksel, Aparna Gupta, Koushik Kar, and Shiv Kalyanaraman

7.1 Contract-switching paradigm 137  
7.2 Architectural issues 138  
   7.2.1 Dynamic contracting over peering points 139  
   7.2.2 Contract routing 139  
7.3 A contract link: bailouts and forwards 143  
   7.3.1 Bailout forward contract (BFC) 144  
   7.3.2 Formalization for pricing a bailout forward contract (BFC) 144  
   7.3.3 Bailout forward contract (BFC) performance evaluation 147  
7.4 Summary 152

References 153

8 PHAROS: an architecture for optical networks 154  

8.1 Introduction 154  
8.2 Background 157  
8.3 PHAROS architecture: an overview 157  
8.4 Resource allocation 161  
   8.4.1 Resource management strategies 161  
   8.4.2 Protection 164  
   8.4.3 Playbooks 166  
   8.4.4 Sub-lambda grooming 168  
8.5 Signaling system 169  
   8.5.1 Control plane operation 171  
   8.5.2 Failure notification 172
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.6</td>
<td>Core node implementation</td>
<td>173</td>
</tr>
<tr>
<td>8.7</td>
<td>Performance analysis</td>
<td>175</td>
</tr>
<tr>
<td>8.8</td>
<td>Concluding remarks</td>
<td>176</td>
</tr>
<tr>
<td></td>
<td>References</td>
<td>177</td>
</tr>
<tr>
<td>9</td>
<td>Customizable in-network services</td>
<td>179</td>
</tr>
<tr>
<td>9.1</td>
<td>Background</td>
<td>179</td>
</tr>
<tr>
<td>9.1.1</td>
<td>Internet architecture</td>
<td>179</td>
</tr>
<tr>
<td>9.1.2</td>
<td>Next-generation Internet</td>
<td>180</td>
</tr>
<tr>
<td>9.1.3</td>
<td>Data path programmability</td>
<td>180</td>
</tr>
<tr>
<td>9.1.4</td>
<td>Technical challenges</td>
<td>181</td>
</tr>
<tr>
<td>9.1.5</td>
<td>In-network processing solutions</td>
<td>181</td>
</tr>
<tr>
<td>9.2</td>
<td>Network services</td>
<td>182</td>
</tr>
<tr>
<td>9.2.1</td>
<td>Concepts</td>
<td>182</td>
</tr>
<tr>
<td>9.2.2</td>
<td>System architecture</td>
<td>184</td>
</tr>
<tr>
<td>9.3</td>
<td>End-system interface and service specification</td>
<td>186</td>
</tr>
<tr>
<td>9.3.1</td>
<td>Service pipeline</td>
<td>186</td>
</tr>
<tr>
<td>9.3.2</td>
<td>Service composition</td>
<td>187</td>
</tr>
<tr>
<td>9.4</td>
<td>Routing and service placement</td>
<td>188</td>
</tr>
<tr>
<td>9.4.1</td>
<td>Problem statement</td>
<td>188</td>
</tr>
<tr>
<td>9.4.2</td>
<td>Centralized routing and placement</td>
<td>189</td>
</tr>
<tr>
<td>9.4.3</td>
<td>Distributed routing and placement</td>
<td>190</td>
</tr>
<tr>
<td>9.5</td>
<td>Runtime resource management</td>
<td>191</td>
</tr>
<tr>
<td>9.5.1</td>
<td>Workload and system model</td>
<td>191</td>
</tr>
<tr>
<td>9.5.2</td>
<td>Resource management problem</td>
<td>192</td>
</tr>
<tr>
<td>9.5.3</td>
<td>Task duplication</td>
<td>192</td>
</tr>
<tr>
<td>9.5.4</td>
<td>Task mapping</td>
<td>193</td>
</tr>
<tr>
<td>9.6</td>
<td>Summary</td>
<td>194</td>
</tr>
<tr>
<td></td>
<td>References</td>
<td>194</td>
</tr>
<tr>
<td>10</td>
<td>Architectural support for continuing Internet evolution and innovation</td>
<td>197</td>
</tr>
<tr>
<td>10.1</td>
<td>Toward a new Internet architecture</td>
<td>197</td>
</tr>
<tr>
<td>10.2</td>
<td>The problems with the current architecture</td>
<td>199</td>
</tr>
<tr>
<td>10.3</td>
<td>SILO architecture: design for change</td>
<td>201</td>
</tr>
<tr>
<td>10.4</td>
<td>Prior related work</td>
<td>206</td>
</tr>
<tr>
<td>10.5</td>
<td>Prototype and case studies</td>
<td>207</td>
</tr>
</tbody>
</table>
### Part III Protocols and practice

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>Separating routing policy from mechanism in the network layer</td>
<td>219</td>
</tr>
<tr>
<td></td>
<td>James Griffioen, Kenneth L. Calvert, Onur Ascigil, and Song Yuan</td>
<td></td>
</tr>
<tr>
<td>11.1</td>
<td>Introduction</td>
<td>219</td>
</tr>
<tr>
<td>11.2</td>
<td>PoMo design goals</td>
<td>220</td>
</tr>
<tr>
<td>11.3</td>
<td>Architecture overview</td>
<td>222</td>
</tr>
<tr>
<td>11.3.1</td>
<td>PFRI network structure and addressing</td>
<td>222</td>
</tr>
<tr>
<td>11.3.2</td>
<td>PFRI forwarding</td>
<td>223</td>
</tr>
<tr>
<td>11.3.3</td>
<td>PFRI routing policies</td>
<td>225</td>
</tr>
<tr>
<td>11.3.4</td>
<td>PFRI packet header mechanisms</td>
<td>226</td>
</tr>
<tr>
<td>11.4</td>
<td>Scaling the PFRI Architecture</td>
<td>227</td>
</tr>
<tr>
<td>11.5</td>
<td>Discussion</td>
<td>230</td>
</tr>
<tr>
<td>11.6</td>
<td>Experimental evaluation</td>
<td>232</td>
</tr>
<tr>
<td>11.7</td>
<td>Other clean-slate approaches</td>
<td>234</td>
</tr>
<tr>
<td></td>
<td>Acknowledgements</td>
<td>235</td>
</tr>
<tr>
<td>References</td>
<td></td>
<td>235</td>
</tr>
</tbody>
</table>

| 12      | Multi-path BGP: motivations and solutions | 238 |
|         | Francisca Valera, Ilijtsch van Beijnum, Alberto García-Martínez, Marcelo Bagnulo | |
| 12.1    | Introduction | 238 |
| 12.2    | Trilogy project | 239 |
| 12.2.1  | Objectives | 239 |
| 12.2.2  | Trilogy technologies | 240 |
| 12.3    | Multi-path routing | 241 |
| 12.3.1  | Higher network capacity | 242 |
| 12.3.2  | Scalable traffic engineering capabilities | 242 |
| 12.3.3  | Improved response to path changes | 242 |
| 12.3.4  | Enhanced security | 243 |
| 12.3.5  | Improved market transparency | 243 |
| 12.4    | Multi-path BGP | 244 |
| 12.4.1  | Intra-domain multi-path routing | 244 |
| 12.4.2  | Inter-domain multi-path routing | 245 |
12.4.3 Motivations for other solutions 247
12.4.4 mBGP and MpASS 248
12.5 Conclusions and future work 253

References 254

13 Explicit congestion control 257
Frank Kelly and Gaurav Raina

13.1 Fairness 258
13.1.1 Why proportional fairness? 260
13.2 Proportionally fair rate control protocol 260
13.2.1 Sufficient conditions for local stability 263
13.2.2 Illustrative simulation 264
13.2.3 Two forms of feedback? 264
13.2.4 Tatonnement processes 265
13.3 Admission management 265
13.3.1 Step-change algorithm 266
13.3.2 Robustness of the step-change algorithm 267
13.3.3 Guidelines for network management 268
13.3.4 Illustrating the utilization–robustness tradeoff 269
13.3.5 Buffer sizing and the step-change algorithm 270
13.4 Concluding remarks 272

References 273

14 KanseiGenie: software infrastructure for resource management and programmability of wireless sensor network fabrics 275
Mukundan Sridharan, Wenjie Zeng, William Leal, Xi Ju, Rajiv Ramnath, Hongwei Zhang, and Anish Arora

14.1 Introduction 275
14.2 Features of sensing fabrics 278
14.2.1 Generic Services 278
14.2.2 Domain-specific services 283
14.3 KanseiGenie architecture 284
14.3.1 The fabric model 284
14.3.2 KanseiGenie Architecture 285
14.3.3 GENI extension to KanseiGenie 287
14.3.4 Implementation of KanseiGenie 288
14.3.5 KanseiGenie federation 290
14.4 KanseiGenie customization and usage 292
14.4.1 How to customize KanseiGenie 292
14.4.2 Vertical APIs and their role in customization 293
14.4.3 KanseiGenie usage step-by-step run-through 294
14.5 Evolving research issues in next-generation networks 295
  14.5.1 Resource specifications for sensor fabrics 295
  14.5.2 Resource discovery 296
  14.5.3 Resource allocation 296
  14.5.4 Data as resource 297
  14.5.5 Network virtualization 297
14.6 Conclusion 298

References 298

Part IV Theory and models 301

15 Theories for buffering and scheduling in Internet switches 303
  Damon Wischik
  15.1 Introduction 303
  15.2 Buffer sizing and end-to-end congestion control 304
    15.2.1 Four heuristic arguments about buffer sizing 305
    15.2.2 Fluid traffic model and queue model 307
    15.2.3 Queuing delay, utilization, and synchronization 309
    15.2.4 Traffic burstiness 312
  15.3 Queueing theory for switches with scheduling 313
    15.3.1 Model for a switched network 313
    15.3.2 The capacity region, and virtual queues 314
    15.3.3 Performance analysis 315
  15.4 A proposed packet-level architecture 320

References 323

16 Stochastic network utility maximization and wireless scheduling 324
  Yung Yi and Mung Chiang
  16.1 Introduction 324
  16.2 LAD (Layering As optimization Decomposition) 326
    16.2.1 Background 326
    16.2.2 Key ideas and procedures 327
  16.3 Stochastic NUM (Network Utility Maximization) 328
    16.3.1 Session-level dynamics 328
    16.3.2 Packet-level dynamics 332
    16.3.3 Constraint-level dynamics 334
    16.3.4 Combinations of multiple dynamics 336
  16.4 Wireless scheduling 337
    16.4.1 Collision-free algorithms 339
    16.4.2 Collision-based algorithms 342
Contributors

Mohamed A. Ali, City University of New York, USA
Neophytos Antoniades, City University of New York, USA
Anish Arora, The Ohio State University, USA
Onur Asgicil, University of Kentucky, USA
Marcelo Bagmulo, Universidad Carlos III de Madrid, Spain
Ilia Baldine, Renaissance Computing Institute, USA
Kenneth L. Calvert, University of Kentucky, USA
Mung Chiang, Princeton University, USA
Davide Cuda, Politecnico di Torino, Turin, Italy
Rudra Dutta, North Carolina State University, USA
Georgios Ellinas, University of Cyprus, Cyprus
Joseph B. Evans, University of Kansas, USA
Alberto Garcia-Martinez, Universidad Carlos III de Madrid, Spain
Roberto Gaudino, Politecnico di Torino, Turin, Italy
Guido A. Gavilanes Castillo, Politecnico di Torino, Turin, Italy
James Griffioen, University of Kentucky, USA
Aparna Gupta, Rensselaer Polytechnic Institute, USA
Andrei Gurtov, University of Oulu, Finland
Antonis Hadjiantonis, University of Nicosia, Cyprus
Thomas R. Henderson, Boeing, USA
Alden W. Jackson, BBN Technologies, USA
John Jacob, BAE Systems, USA
Suyang Ju, University of Kansas, USA
Xi Ju, The Ohio State University, USA
Shiv Kalyanaraman, IBM Research, India
Koushik Kar, Rensselaer Polytechnic Institute, USA
Frank Kelly, University of Cambridge, UK
Ahmad Khalil, City University of New York, USA
William Leal, The Ohio State University, USA
Will E. Leland, BBN Technologies, USA
Baochun Li, University of Toronto, Canada
Zongpeng Li, University of Calgary, Canada
John H. Lowry, BBN Technologies, USA
Walter C. Milliken, BBN Technologies, USA
Biswanath Mukherjee, University of California Davis, USA
John Musacchio, University of California Santa Cruz, USA
Fabio Neri, Politecnico di Torino, Turin, Italy
Pekka Nikander, Ericsson Research, Finland
P. Pal, BBN Technologies, USA
Gaurav Raina, Indian Institute of Technology Madras, India
Byrav Ramamurthy, University of Nebraska-Lincoln, USA
Subramanian Ramanathan, BBN Technologies, USA
Rajiv Ramnath, The Ohio State University, USA
K. Rauschenbach, BBN Technologies, USA
Anusha Ravula, University of Nebraska-Lincoln, USA
Abu (Sayeem) Reaz, University of California Davis, USA
Ceaser A. Santivanez, BBN Technologies, USA
Galina Schwartz, University of California Berkeley, USA
Lei Shi, University of California Davis, USA
Mukundan Sridharan, The Ohio State University, USA
Francisco Valera, Universidad Carlos III de Madrid, Spain
Iljitsch van Beijnum, IMDEA Networks, Spain
Jean Walrand, University of California Berkeley, USA
Damon Wischik, University College London, UK
Daniel Wood, Verizon Network Systems, USA
Tilman Wolf, University of Massachusetts Amherst, USA
Hong Xu, University of Toronto, Canada
Yung Yi, Princeton University, USA
Song Yuan, University of Kentucky, USA
Murat Yuksel, University of Nevada - Reno, USA
Wenjie Zeng, The Ohio State University, USA
Hongwei Zhang, The Ohio State University, USA