# Contents

## Preface

### 1 Internet Services

1.1 Introduction ............................................. 1

1.2 Requirements and Key Challenges ....................... 2
1.2.1 Transparency ........................................... 2
1.2.2 Scalability ............................................ 3
1.2.3 Heterogeneity ......................................... 4
1.2.4 Security .............................................. 5

1.3 Examples of Scalable Internet Services ................. 7
1.3.1 Search Engine ......................................... 7
1.3.2 On-line Shopping and E-Commerce ................... 8
1.3.3 Media Streaming Services ............................. 9
1.3.4 Peer-to-Peer File Sharing ............................ 10
1.3.5 Open Grid Service .................................... 11

1.4 Road Map to the Book .................................... 11

## 2 Network Load Balancing

2.1 The Load Balancing Problem ............................. 15

2.2 Server Load Balancing .................................... 17
2.2.1 Layer-4 Load Balancer ................................. 18
2.2.2 Layer-7 Load Balancer ................................. 20
2.2.3 Load Balancing Policies ............................... 22
2.2.4 Load Balancing with Session Persistence ............ 23
2.2.5 Distributed Approaches for Load Balancing ......... 25

2.3 Load Balancing in Service Overlay Networks .......... 26
2.3.1 HTTP Request Redirection and URL Rewriting ....... 26
2.3.2 DNS-Based Request Redirection ....................... 26
2.3.3 Content Delivery Networks ............................ 28

2.4 A Unified W5 Load Balancing Model .................... 30
2.4.1 Objectives of Load Balancing ........................ 30
2.4.2 Who Makes Load Balancing Decisions ................. 31
2.4.3 What Information Is the Decision Based on .......... 31
2.4.4 Which Task Is the Best Candidate for Migration .... 33
2.4.5 Where Should the Task Be Performed ................. 34
2.4.6 When or Why Is Migration Invoked ................... 35
### 6 Service Differentiation on E-Commerce Servers

6.1 Introduction .................................................. 93
6.2 2D Service Differentiation Model ............................ 95
   6.2.1 The Model ............................................. 95
   6.2.2 Objectives of Processing Rate Allocation .......... 97
6.3 An Optimal Processing Rate Allocation Scheme .......... 99
6.4 Effectiveness of 2D Service Differentiation ............ 101
   6.4.1 A Simulation Model ................................ 101
   6.4.2 Effect on Service Slowdown ......................... 103
   6.4.3 Controllability of Service Differentiation ...... 108
6.5 Concluding Remarks ........................................ 109

### 7 Feedback Control for Quality-of-Service Guarantees

7.1 Introduction .................................................... 111
7.2 Slowdown in an $M/G_P/1$ Queueing System ............. 112
   7.2.1 Slowdown Preliminaries ............................. 113
   7.2.2 Slowdown on Internet Servers ..................... 114
7.3 Processing Rate Allocation with Feedback Control ...... 115
   7.3.1 Queueing Theoretical Approach for Service Differentiation . 116
   7.3.2 Integrated Feedback Control Approach .......... 118
7.4 Robustness of the Integrated Approach ................. 121
7.5 QoS-Aware Apache Server with Feedback Control ...... 123
   7.5.1 Implementation of a QoS-Aware Apache Server 123
   7.5.2 QoS Guarantees in Real Environments .......... 124
7.6 Concluding Remarks ........................................ 126

### 8 Decay Function Model for Server Capacity Planning

8.1 Introduction .................................................... 129
8.2 The Decay Function Model ................................. 132
8.3 Resource Configuration and Allocation .................. 136
   8.3.1 Resource Configuration ............................. 136
   8.3.2 Optimal Fixed-Time Scheduling ................... 137
   8.3.3 Examples ............................................ 140
8.4 Performance Evaluation .................................... 142
   8.4.1 Capacity Configurations and Variances .......... 143
   8.4.2 Decay Function versus GPS Scheduling .......... 145
   8.4.3 Sensitivity to the Change of Traffic Intensity .... 146
   8.4.4 Quality-of-Service Prediction ................... 148
8.5 Concluding Remarks ........................................ 149

### 9 Scalable Constant-Degree Peer-to-Peer Overlay Networks

9.1 Introduction .................................................... 151
9.2 Topological Model of DHT-Based P2P Systems .......... 152
   9.2.1 A Generic Topological Model ..................... 153
   9.2.2 Characteristics of Representative DHT Networks . 155
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.3</td>
<td>Cycloid: A Constant-Degree DHT Network</td>
<td>157</td>
</tr>
<tr>
<td>9.3.1</td>
<td>CCC and Key Assignment</td>
<td>158</td>
</tr>
<tr>
<td>9.3.2</td>
<td>Cycloid Routing Algorithm</td>
<td>160</td>
</tr>
<tr>
<td>9.3.3</td>
<td>Self-Organization</td>
<td>161</td>
</tr>
<tr>
<td>9.4</td>
<td>Cycloid Performance Evaluation</td>
<td>163</td>
</tr>
<tr>
<td>9.4.1</td>
<td>Key Location Efficiency</td>
<td>163</td>
</tr>
<tr>
<td>9.4.2</td>
<td>Load Distribution</td>
<td>165</td>
</tr>
<tr>
<td>9.4.3</td>
<td>Network Resilience</td>
<td>168</td>
</tr>
<tr>
<td>9.5</td>
<td>Concluding Remarks</td>
<td>172</td>
</tr>
<tr>
<td>10</td>
<td>Semantic Prefetching of Web Contents</td>
<td>175</td>
</tr>
<tr>
<td>10.1</td>
<td>Introduction</td>
<td>175</td>
</tr>
<tr>
<td>10.2</td>
<td>Personalized Semantic Prefetching</td>
<td>177</td>
</tr>
<tr>
<td>10.2.1</td>
<td>Architecture</td>
<td>177</td>
</tr>
<tr>
<td>10.2.2</td>
<td>Neural Network-Based Semantics Model</td>
<td>178</td>
</tr>
<tr>
<td>10.3</td>
<td>NewsAgent: A News Prefetching System</td>
<td>181</td>
</tr>
<tr>
<td>10.3.1</td>
<td>Keywords</td>
<td>181</td>
</tr>
<tr>
<td>10.3.2</td>
<td>NewsAgent Architecture</td>
<td>182</td>
</tr>
<tr>
<td>10.3.3</td>
<td>Control of Prefetching Cache and Keyword List</td>
<td>183</td>
</tr>
<tr>
<td>10.4</td>
<td>Real-Time Simultaneous Evaluation Methodology</td>
<td>185</td>
</tr>
<tr>
<td>10.5</td>
<td>Experimental Results</td>
<td>186</td>
</tr>
<tr>
<td>10.5.1</td>
<td>NewsAgent Training</td>
<td>186</td>
</tr>
<tr>
<td>10.5.2</td>
<td>Effectiveness of Semantic Prefetching</td>
<td>188</td>
</tr>
<tr>
<td>10.5.3</td>
<td>Effects of Net Threshold and Learning Rate</td>
<td>190</td>
</tr>
<tr>
<td>10.5.4</td>
<td>Keyword List Management</td>
<td>192</td>
</tr>
<tr>
<td>10.6</td>
<td>Related Work</td>
<td>192</td>
</tr>
<tr>
<td>10.7</td>
<td>Concluding Remarks</td>
<td>194</td>
</tr>
<tr>
<td>11</td>
<td>Mobile Code and Security</td>
<td>197</td>
</tr>
<tr>
<td>11.1</td>
<td>Introduction</td>
<td>197</td>
</tr>
<tr>
<td>11.2</td>
<td>Design Issues in Mobile Agent Systems</td>
<td>200</td>
</tr>
<tr>
<td>11.2.1</td>
<td>Migration</td>
<td>200</td>
</tr>
<tr>
<td>11.2.2</td>
<td>Communication</td>
<td>201</td>
</tr>
<tr>
<td>11.2.3</td>
<td>Naming and Name Resolution</td>
<td>202</td>
</tr>
<tr>
<td>11.2.4</td>
<td>Security</td>
<td>203</td>
</tr>
<tr>
<td>11.3</td>
<td>Agent Host Protections</td>
<td>203</td>
</tr>
<tr>
<td>11.3.1</td>
<td>Security Requirements</td>
<td>203</td>
</tr>
<tr>
<td>11.3.2</td>
<td>Agent Authentication</td>
<td>204</td>
</tr>
<tr>
<td>11.3.3</td>
<td>Privilege Delegation and Agent Authorization</td>
<td>205</td>
</tr>
<tr>
<td>11.3.4</td>
<td>Agent-Oriented Access Control</td>
<td>206</td>
</tr>
<tr>
<td>11.3.5</td>
<td>Proof-Carrying Code</td>
<td>207</td>
</tr>
<tr>
<td>11.4</td>
<td>Mobile Agent Protections</td>
<td>209</td>
</tr>
<tr>
<td>11.4.1</td>
<td>Security Requirements</td>
<td>209</td>
</tr>
<tr>
<td>11.4.2</td>
<td>Integrity Detection</td>
<td>211</td>
</tr>
<tr>
<td>11.4.3</td>
<td>Cryptographic Protection of Agents</td>
<td>212</td>
</tr>
</tbody>
</table>
### 12 Naplet: A Mobile Agent Approach

12.1 Introduction ........................................... 219
12.2 Design Goals and Naplet Architecture ................ 220
  12.2.1 Naplet Class .................................. 221
  12.2.2 NapletServer Architecture ...................... 223
12.3 Structured Itinerary Mechanism ....................... 226
  12.3.1 Primitive Itinerary Constructs ................. 226
  12.3.2 Itinerary Programming Interfaces ............. 227
  12.3.3 Implementations of Itinerary Patterns .......... 230
12.4 Naplet Tracking and Location Finding ............... 232
  12.4.1 Mobile Agent Tracking Overview ............... 232
  12.4.2 Naplet Location Service ....................... 234
12.5 Reliable Agent Communication ........................ 235
  12.5.1 PostOffice Messaging Service ................. 235
  12.5.2 NapletSocket for Synchronous Communication ... 237
12.6 Security and Resource Management ................... 238
  12.6.1 Naplet Security Architecture ................. 239
  12.6.2 Resource Management .......................... 240
12.7 Programming for Network Management in Naplet ........ 242
  12.7.1 Privileged Service for Naplet Access to MIB ... 243
  12.7.2 Naplet for Network Management ............... 244

### 13 Itinerary Safety Reasoning and Assurance

13.1 Introduction ........................................... 247
13.2 MAIL: A Mobile Agent Itinerary Language ........... 249
  13.2.1 Syntax of MAIL ................................ 249
  13.2.2 Operational Semantics of MAIL ............... 252
13.3 Regular-Completeness of MAIL ....................... 255
13.4 Itinerary Safety Reasoning and Assurance .......... 257
  13.4.1 Itinerary Configuration and Safety ........... 257
  13.4.2 Itinerary Safety Reasoning and Assurance .... 259
13.5 Concluding Remarks ................................ 262

### 14 Security Measures for Server Protection

14.1 Introduction ........................................... 263
14.2 Agent-Oriented Access Control ....................... 265
  14.2.1 Access Control in Mobile Codes ............... 265
  14.2.2 Agent Naming and Authentication in Naplet .... 267
  14.2.3 Naplet Access Control Mechanism ............. 270
14.3 Coordinated Spatio-Temporal Access Control ....... 273
  14.3.1 Temporal Constraints .......................... 273
  14.3.2 Spatial Constraints ........................... 275
14.4 Concluding Remarks ................................ 277
15 Connection Migration in Mobile Agents  
15.1 Introduction ............................................. 279  
15.1.1 Related Work ........................................ 280  
15.2 NapletSocket: A Connection Migration Mechanism .... 281  
15.2.1 NapletSocket Architecture ......................... 281  
15.2.2 State Transitions ................................... 282  
15.3 Design Issues in NapletSocket ......................... 285  
15.3.1 Transparency and Reliability ....................... 285  
15.3.2 Multiple Connections ............................... 288  
15.3.3 Security ............................................. 289  
15.3.4 Socket Hand-Off .................................... 290  
15.3.5 Control Channel .................................... 291  
15.4 Experimental Results of NapletSocket ................. 292  
15.4.1 Effectiveness of Reliable Communication .......... 292  
15.4.2 Cost of Primitive NapletSocket Operations ....... 293  
15.4.3 NapletSocket Throughput ............................ 294  
15.5 Performance Model of Agent Mobility ................. 296  
15.5.1 Performance Model ................................. 296  
15.5.2 Simulation Results ................................. 298  
15.6 Concluding Remarks .................................... 300  

16 Mobility Support for Adaptive Grid Computing  
16.1 Introduction ............................................. 301  
16.2 An Agent-Oriented Programming Framework ............ 303  
16.2.1 The Architecture .................................... 303  
16.2.2 Strong Mobility of Multithreaded Agents .......... 305  
16.3 Distributed Shared Arrays for Virtual Machines .... 307  
16.3.1 DSA Architecture ................................... 308  
16.3.2 DSA APIs ........................................... 309  
16.3.3 DSA Run-Time Support ............................... 310  
16.4 Experimental Results ................................... 315  
16.4.1 Cost for Creating a Virtual Machine ............... 315  
16.4.2 Cost for DSA Read/Write Operations ............... 316  
16.4.3 Performance of LU Factorization and FFT .......... 318  
16.5 Concluding Remarks .................................... 321  

17 Service Migration in Reconfigurable Distributed Virtual Machines  
17.1 Introduction ............................................. 323  
17.2 M-DSA: DSA with Service Mobility Support ........... 325  
17.2.1 M-DSA Architecture ................................. 325  
17.2.2 An Example of M-DSA Programs ..................... 326  
17.3 Service Migration in M-DSA ............................. 328  
17.3.1 Performance Monitoring for Service Migration .... 328  
17.3.2 Service Packing and Restoration .................... 329  
17.3.3 Computational Agent State Capture .................. 331
17.3.4 Service Migration: A Summary .......................... 332
17.4 Interface to Globus Service ............................... 333
   17.4.1 Resource Management and Migration Decision ...... 333
   17.4.2 Security Protection ................................. 334
17.5 Experiment Results ........................................ 335
   17.5.1 Execution Time of LU and FFT on M-DSA .......... 336
   17.5.2 Service Migration Overhead Breakdown .............. 338
   17.5.3 Security Protection for Intercluster Communication 339
17.6 Related Work ............................................. 340
17.7 Concluding Remarks ...................................... 342

18 Migration Decision in Reconfigurable Distributed Virtual Machines 343
   18.1 Introduction .............................................. 343
   18.2 Reconfigurable Virtual Machine Model ................ 344
   18.3 Service Migration Decision ............................. 347
      18.3.1 Migration Candidate Determination and Service Migration Timing ........................................ 347
      18.3.2 Destination Server Selection ....................... 349
   18.4 Hybrid Migration Decision .............................. 353
      18.4.1 Agent Migration Decision .......................... 353
      18.4.2 Interplay between Service and Agent Migration .... 355
   18.5 Simulation Results ...................................... 356
   18.6 Concluding Remarks ..................................... 361

References ......................................................... 363
Index .................................................................. 391