Distributed Systems
Third edition
Version 3.02 (2018)
(Contains minor corrections in comparison to 3.01)

Maarten van Steen
Andrew S. Tanenbaum

Get a free copy of this book at:

https://www.distributed-systems.net/index.php/books/ds3/
# Contents

<table>
<thead>
<tr>
<th>Preface</th>
<th>xi</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1 Introduction</strong></td>
<td>1</td>
</tr>
<tr>
<td>1.1 What is a distributed system?</td>
<td>2</td>
</tr>
<tr>
<td>Characteristic 1: Collection of autonomous computing elements</td>
<td>2</td>
</tr>
<tr>
<td>Characteristic 2: Single coherent system</td>
<td>4</td>
</tr>
<tr>
<td>Middleware and distributed systems</td>
<td>5</td>
</tr>
<tr>
<td>1.2 Design goals</td>
<td>7</td>
</tr>
<tr>
<td>Supporting resource sharing</td>
<td>7</td>
</tr>
<tr>
<td>Making distribution transparent</td>
<td>8</td>
</tr>
<tr>
<td>Being open</td>
<td>12</td>
</tr>
<tr>
<td>Being scalable</td>
<td>15</td>
</tr>
<tr>
<td>Pitfalls</td>
<td>24</td>
</tr>
<tr>
<td>1.3 Types of distributed systems</td>
<td>24</td>
</tr>
<tr>
<td>High performance distributed computing</td>
<td>25</td>
</tr>
<tr>
<td>Distributed information systems</td>
<td>34</td>
</tr>
<tr>
<td>Pervasive systems</td>
<td>40</td>
</tr>
<tr>
<td>1.4 Summary</td>
<td>52</td>
</tr>
<tr>
<td><strong>2 Architectures</strong></td>
<td>55</td>
</tr>
<tr>
<td>2.1 Architectural styles</td>
<td>56</td>
</tr>
<tr>
<td>Layered architectures</td>
<td>57</td>
</tr>
<tr>
<td>Object-based and service-oriented architectures</td>
<td>62</td>
</tr>
<tr>
<td>Resource-based architectures</td>
<td>64</td>
</tr>
<tr>
<td>Publish-subscribe architectures</td>
<td>66</td>
</tr>
<tr>
<td>2.2 Middleware organization</td>
<td>71</td>
</tr>
<tr>
<td>Wrappers</td>
<td>72</td>
</tr>
<tr>
<td>Interceptors</td>
<td>73</td>
</tr>
<tr>
<td>Modifiable middleware</td>
<td>75</td>
</tr>
<tr>
<td>2.3 System architecture</td>
<td>76</td>
</tr>
</tbody>
</table>
CONTENTS

4.4 Multicast communication ........................................ 221
  Application-level tree-based multicasting ...................... 222
  Flooding-based multicasting ...................................... 226
  Gossip-based data dissemination ................................ 229
4.5 Summary ............................................................ 234

5 Naming ............................................................................. 237
  5.1 Names, identifiers, and addresses ............................... 238
  5.2 Flat naming ............................................................ 241
    Simple solutions .................................................... 241
    Home-based approaches .......................................... 245
    Distributed hash tables ........................................... 246
    Hierarchical approaches ......................................... 251
  5.3 Structured naming .................................................. 256
    Name spaces ........................................................ 256
    Name resolution ..................................................... 259
    The implementation of a name space ......................... 264
    Example: The Domain Name System ........................... 271
    Example: The Network File System ............................ 278
  5.4 Attribute-based naming ............................................ 283
    Directory services .................................................. 283
    Hierarchical implementations: LDAP ......................... 285
    Decentralized implementations ................................ 288
  5.5 Summary ............................................................ 294

6 Coordination ...................................................................... 297
  6.1 Clock synchronization ............................................. 298
    Physical clocks ...................................................... 299
    Clock synchronization algorithms .............................. 302
  6.2 Logical clocks ....................................................... 310
    Lamport’s logical clocks ......................................... 310
    Vector clocks ........................................................ 316
  6.3 Mutual exclusion ..................................................... 321
    Overview .............................................................. 322
    A centralized algorithm ......................................... 322
    A distributed algorithm ......................................... 323
    A token-ring algorithm ......................................... 325
    A decentralized algorithm ..................................... 326
  6.4 Election algorithms ................................................ 329
    The bully algorithm ............................................... 330
    A ring algorithm ................................................... 332
    Elections in wireless environments ......................... 333
    Elections in large-scale systems ............................... 335
  6.5 Location systems .................................................... 336
# CONTENTS

GPS: Global Positioning System ........................................ 337  
When GPS is not an option ............................................ 339  
Logical positioning of nodes ....................................... 339  
6.6 Distributed event matching ..................................... 343  
Centralized implementations ....................................... 343  
6.7 Gossip-based coordination ...................................... 349  
Aggregation ................................................................. 349  
A peer-sampling service ............................................. 350  
Gossip-based overlay construction ................................. 352  
6.8 Summary ................................................................. 353  

7 Consistency and replication ................................. 355  
7.1 Introduction ........................................................... 356  
Reasons for replication ............................................. 356  
Replication as scaling technique ................................. 357  
7.2 Data-centric consistency models ............................. 358  
Continuous consistency ............................................. 359  
Consistent ordering of operations ............................... 364  
Eventual consistency .................................................. 373  
7.3 Client-centric consistency models ......................... 375  
Monotonic reads ......................................................... 377  
Monotonic writes ....................................................... 379  
Read your writes ........................................................ 380  
Writes follow reads ...................................................... 382  
7.4 Replica management ................................................ 383  
Finding the best server location ................................ 383  
Content replication and placement ............................. 385  
Content distribution .................................................... 388  
Managing replicated objects ....................................... 393  
7.5 Consistency protocols ............................................... 396  
Continuous consistency ............................................. 396  
Primary-based protocols ........................................... 398  
Replicated-write protocols ......................................... 401  
Cache-coherence protocols ......................................... 403  
Implementing client-centric consistency .................... 407  
7.6 Example: Caching and replication in the Web ......... 409  
7.7 Summary ................................................................. 420  

8 Fault tolerance ......................................................... 423  
8.1 Introduction to fault tolerance ............................... 424  
Basic concepts ............................................................ 424  
Failure models ............................................................ 427  
Failure masking by redundancy .................................. 431  
8.2 Process resilience ..................................................... 432