The All-New Switch Book

The Complete Guide to LAN Switching Technology
Second Edition

Rich Seifert
Jim Edwards

Wiley Publishing, Inc.
## Preface

## Introduction

### Part One: Foundations of LAN Switches

#### Chapter 1: Laying the Foundation

<table>
<thead>
<tr>
<th>Network Architecture</th>
<th>Physical Layer</th>
<th>Data Link Layer</th>
<th>Network Layer</th>
<th>Transport Layer</th>
<th>Session Layer</th>
<th>Presentation Layer</th>
<th>Application Layer</th>
<th>Layering Makes a Good Servant but a Bad Master</th>
<th>Inside the Data Link Layer</th>
<th>Modes of Operation</th>
<th>Data Link Sublayering</th>
<th>Logical Link Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Addressing

- Local and Global Uniqueness
  - LAN Data Link Addresses
    - Unicast and Multicast Addresses
    - Globally Unique and Locally Unique MAC Addresses
    - How LAN Addresses Are Assigned
    - Written Address Conventions

### LAN Technology Review

- Ethernet
<table>
<thead>
<tr>
<th>Chapter 2</th>
<th>Transparent Bridges</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Principles of Operation</strong></td>
<td>63</td>
</tr>
<tr>
<td><strong>Unicast Operation</strong></td>
<td>65</td>
</tr>
<tr>
<td><strong>Unknown and Multicast Destinations</strong></td>
<td>66</td>
</tr>
<tr>
<td><strong>Generating the Address Table</strong></td>
<td>68</td>
</tr>
<tr>
<td><strong>Address Table Aging</strong></td>
<td>69</td>
</tr>
<tr>
<td><strong>Process Model of Table Operation</strong></td>
<td>70</td>
</tr>
<tr>
<td><strong>Custom Filtering and Forwarding</strong></td>
<td>72</td>
</tr>
<tr>
<td><strong>Multiple Bridge Topologies</strong></td>
<td>73</td>
</tr>
<tr>
<td><strong>Transparent Bridge Architecture</strong></td>
<td>74</td>
</tr>
<tr>
<td><strong>Maintaining the Link Invariants</strong></td>
<td>76</td>
</tr>
<tr>
<td><strong>The Hard Invariants Are Hard Indeed</strong></td>
<td>78</td>
</tr>
<tr>
<td><strong>Soft Invariants</strong></td>
<td>80</td>
</tr>
</tbody>
</table>

| Implementing the Bridge Address Table | 84 |
| **Table Operations** | 85 |
| **Search Algorithms** | 85 |
| **Hash Tables** | 85 |
| **Binary Search** | 88 |
| **Content-Addressable Memories** | 90 |
| **How Deep Is Your Table?** | 92 |
| **Aging Entries from the Table** | 93 |
Chapter 6  Source Routing  255

Overview of Source Routing Operation   256
Eine Kleine Sourceroutinggeschichte   257

Source Routing Concepts
Nontransparency, or "Peek-a-Boo — I See You!"   260
Who's the Boss?   260
Connection Orientation   261
Be All That You Can Be (Without Joining the Army)   263
Even Token Rings Need to Get Out of the Loop Sometimes   263
Ring and Bridge Numbering   264
Route Discovery 266
Maximum Transmission Unit Discovery 266
Source-Routed Frames 267
  Differentiating Source-Routed and Non-Source-Routed Frames 267
  Non-Source-Routed Frames 269
  Source-Routed Frame Format 269
  Routing Control Fields 269
  Route Descriptors 273
Source Routing Operation 274
  Route Discovery 275
    Route Discovery Algorithms 275
    Route Discovery Frames 277
    Route Selection 279
    Issues in Route Discovery 280
Station Operation 282
  Architectural Model of Source Routing 282
  End Station Transmit Behavior 282
  End Station Receive Behavior 284
Bridge Operation 285
  Bridge Behavior for Specifically Routed Frames 286
  Bridge Behavior for Explorer Frames (Both ARE and STE) 286
Interconnecting the Source-Routed and Transparently Bridged Universes 289
  Don’t Bridge — Route! 294
  The Source Routing-to-Transparent Bridge 295
  The Source Routing/Transparent Bridge 298
IEEE Standards and Source Routing 301
The Future of Source Routing 301

Part Two  Advanced LAN Switch Concepts

Chapter 7  Full Duplex Operation 305

  Why a MAC? 305
  Full Duplex Enablers 307
    Dedicated Media 307
    Dedicated LAN 310
  Full Duplex Ethernet 311
    “Ethernet Is CSMA/CD” 312
    Full Duplex Ethernet Operating Environment 313
    Subset of Half Duplex Operation 314
    Transmitter Operation 315
    Receiver Operation 315
    Ethernet Minimum Frame Size Constraint 316
  Dedicated Token Ring 317
  Implications of Full Duplex Operation 319
    Eliminating the Link Length Restriction of Half Duplex Ethernet 319


<table>
<thead>
<tr>
<th>Contents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Increasing the Link Capacity 320</td>
</tr>
<tr>
<td>Increasing Switch Load 322</td>
</tr>
<tr>
<td>Full Duplex Application Environments 323</td>
</tr>
<tr>
<td>Switch-to-Switch Connections 323</td>
</tr>
<tr>
<td>Server and Router Connections 324</td>
</tr>
<tr>
<td>Long-Distance Connections 325</td>
</tr>
</tbody>
</table>

**Chapter 8**  
**LAN and Switch Flow Control** 327

- The Need for Flow Control 327
- Default Switch Behavior 330
- The Effect of Frame Loss 330
- End-to-End Flow Control 332
- Cost-Performance Tradeoffs 332

- Controlling Flow in Half Duplex Networks 333
- Backpressure 333
  - Aggressive Transmission Policies 337

- MAC Control 341
  - MAC Control Architecture 341
  - MAC Control Frame Format 343

- PAUSE Function 344
  - Overview of PAUSE Operation 346
  - PAUSE Frame Semantics 347
  - Configuration of Flow Control Capabilities 349

- IEEE 802.3x Flow Control Implementation Issues 350
  - Design Implications of PAUSE Function 351
    - Inserting PAUSE Frames in the Transmit Queue 351
    - Parsing Received PAUSE Frames 352
    - PAUSE Timing 353
    - Buffering Requirements 354

- Flow Control Policies and Use 356
  - Buffer Thresholds 356
  - Selection of PAUSE Times 357
  - Dealing with Unreliable Delivery 358

- Flow Control Symmetry 358
  - Symmetric Flow Control 359
  - Asymmetric Flow Control 359

**Chapter 9**  
**Link Aggregation** 361

- Link Aggregation Benefits 362
- Application of Link Aggregation 364
  - Switch-to-Switch Connections 365
  - Switch-to-Station (Server or Router) Connections 365
  - Station-to-Station Connections 367

- Aggregate or Upgrade? 367
- Issues in Link Aggregation 368
  - Addressing 368
  - Distributing Traffic Across an Aggregation 371
## Contents

Maintaining Link Invariants in an Aggregated Environment 372
Separating Traffic Flows 374
Conversation Determination Aids the Realization of Aggregation 375
Mapping the Distribution Function to the Physical Link 377
Conversations Above the Data Link Layer 377
Summary of Distribution Functions 380
Changing the Distribution 381
Performance 384
Technology Constraints (a.k.a. Link Aggravation) 384
Mixing LAN Technologies in a Single Aggregation 384
Mixing Data Rates in a Single Aggregation 385
Aggregation and Shared LANs 385
Configuration Control 385
IEEE 802.3ad Link Aggregation Standard 388
Scope of the Standard 388
Features and Benefits of the Standard 390
Link Aggregation Architectural Model 392
Binding Physical Ports to Aggregators 394
Binding, Distribution, and Collection 397
Addressing 397
Marker Protocol Operation 398
Link Aggregation Control Protocol 401
LACP Concepts 401
LACP Frame Format 406

### Chapter 10 Multicast Pruning 413

Multicast Usage 413
Who Assigns Multicast Addresses? 414
Application Use of Multicast 417
Implications of Default Behavior 419
Trimming the (Spanning) Tree 420
The Weekend Networker’s Guide to Tree Pruning 421
Receiver Declaration 421
Registration of the Declaration 422
Propagation of the Registration 423
Source Pruning 424
IEEE 802.1p 424
GARP Multicast Registration Protocol 424
Generic Attribute Registration Protocol 426
GMRP Use of GARP 430

### Chapter 11 Virtual LANs: Applications and Concepts 433

Applications of VLANs 434
The Software Patch Panel 434
LAN Security 437
User Mobility 439
Bandwidth Preservation 442
VLAN Concepts 443
Playing Tag on Your LAN 445
  Implicit Tags 445
  Explicit Tags 446
  VLAN Awareness and Tag Awareness 448
VLAN Awareness 448
  What It Means to Be VLAN-Aware 449
  VLAN-Aware Switches 449
  VLAN-Aware End Stations 454
  He Looks Around, Around, He Sees VLANs in the
  Architecture, Spinning in Infinity... 456
Shared Media and VLAN Awareness 458
  Non-VLAN-Aware Switches and End Stations 458
VLAN Association Rules (Mapping Frames to VLANs) 459
  Port-Based VLAN Mapping 460
  MAC Address-Based VLAN Mapping 461
  Protocol-Based VLAN Mapping 462
  IP Subnet-Based VLAN Mapping 465
  A VLAN Phenomenon: The One-Armed Router 466
  Application-Based VLAN Mapping 469
  The Rules Follow the Application 471
Frame Forwarding 472

Chapter 12 Virtual LANs: The IEEE Standard 475
Overview and Scope of the Standard 477
Elements of the Standard 478
Tag and Frame Formats 480
  VLAN Protocol Identifier 481
  Tag Control Information Field 482
  Embedded Routing Information Field 485
    Route Control Portion 486
    Route Descriptor Portion 487
  Tagged Ethernet Frames 488
  Flash! Ethernet MTU Increases by 4 Bytes! 492
  Tagged Token Ring Frames 495
  Tagged FDDI Frames 495
  VLAN Tags on Other LAN Technologies 496
  A Word on Bit and Byte Order 496
IEEE 802.1Q Switch Operation 497
  Ingress Process 499
    Acceptable Frame Filter 499
    Ingress Rules 499
    Ingress Filter 500
  Progress Process 500
    Forwarding in a VLAN-Aware Switch 500
Maintaining the Filtering Database 501
Egress Process 502
Egress Rules 502
Egress Filter 504
System-Level Switch Constraints 506
GARP VLAN Registration Protocol 506
GVRP Use of GARP 507
Multicast Registration and VLAN Context 508
VLANs and the Spanning Tree 508
The Multiple Spanning Tree Protocol 511
So Exactly What Are They Trying to Accomplish Here? 511
What the Heck Does This All Mean? 512
Tha-tha-tha-tha-tha...That’s Right Folks! 512
Multiple Spanning Tree Instance 513
MST Regions 514

Chapter 13 **Priority Operation**  517
Why Priority? 517
LAN Priority Mechanisms 519
Token Ring Priority Mechanisms 520
FDDI Priority Mechanisms 521
Ethernet Priority Mechanisms 522
VLAN and Priority Tagging 525
Getting into the Priority Business 526
Priority Operation in Switches 529
The Ordering Invariant — Redux 530
IEEE 802.1p 530
Switch Process Flow for Priority Operation 532
Determining Frame Priority on Input 533
Tag, You’re It! 533
LAN-Specific User Priority Indication 533
Implicit Priority Determination, or “Whose Clues Do You Use?” 534
Priority Regeneration 535
Mapping Input Priority to Class-of-Service 536
Class of Service Versus Quality of Service 536
How Many Queues Do You Chueues? 538
Default Priority Mappings 540
Output Scheduling 541
Scheduling Algorithms 541
Indicating the Priority in Transmitted Frames 544
Mapping User Priority to Access Priority at the Output Port 545

Chapter 14 **LAN Security**  547
Network Security Overview 548
Hackers, Crackers, Viruses, and Those Confounded Worms 549
Hac and Crac, the Ker Brothers. 549
Chapter 16  Network Troubleshooting Strategies

The Trouble with Troubleshooting
Housekeeping
  Running the Network Baseline
Proactive Troubleshooting
Troubleshooting Tools
  Troubleshooting Utilities
    ping
    trace route
    netstat
    route
    ARP
More Advanced Tools of the Trade
  Network Analyzers (or whatever they are calling them today)
  Other Testing Equipment
  ... and if all else fails
A Systematic Approach
  Defining the Problem
  Sharing the Known
  Determining the Issue
  Developing a Solution
  Resolving and Taking Action!
  Monitoring the Results
  The Final Step — Have a Beer!
Some Strategies for Layer 2 Troubleshooting
  Performing a Health Check
Software, Hardware, and Configuration
  Issues Relating to Software
  Issues Relating to Hardware
  Issues Relating to Configuration
Common Layer 2 Issues
  VLANS
  Duplex Mismatches
  Spanning Tree
Wrap Up

Chapter 17  Make the Switch!

Keeping House
  Housekeeping Functions
  Implementation and Performance
    (or, It’s Tough to Find a Good Housekeeper)
<table>
<thead>
<tr>
<th>Contents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Switch Data Receive Path Functions</td>
</tr>
<tr>
<td>Port Interfaces (Receive)</td>
</tr>
<tr>
<td>Receive Flow Control</td>
</tr>
<tr>
<td>Link Aggregation Collector</td>
</tr>
<tr>
<td>Classification Engine</td>
</tr>
<tr>
<td>Local Sinking of Reserved Multicast Addresses</td>
</tr>
<tr>
<td>VLAN Ingress Rules</td>
</tr>
<tr>
<td>Priority Ingress Rules</td>
</tr>
<tr>
<td>Do It Once and Save the Results</td>
</tr>
<tr>
<td>Implementation of the Classification Engine</td>
</tr>
<tr>
<td>VLAN Filters</td>
</tr>
<tr>
<td>Lookup Engine</td>
</tr>
<tr>
<td>Generating the Output Vector</td>
</tr>
<tr>
<td>Maintaining the Filtering Database</td>
</tr>
<tr>
<td>Lookup Implementation</td>
</tr>
<tr>
<td>Switch Fabrics</td>
</tr>
<tr>
<td>Shared Memory</td>
</tr>
<tr>
<td>Shared Memory Fabric Operation</td>
</tr>
<tr>
<td>Multicasting in a Shared Memory Architecture</td>
</tr>
<tr>
<td>Buffer Organization</td>
</tr>
<tr>
<td>Memory Bandwidth Limitations</td>
</tr>
<tr>
<td>Increasing the Memory Bandwidth</td>
</tr>
<tr>
<td>Shared Bus</td>
</tr>
<tr>
<td>Crosspoint Matrix</td>
</tr>
<tr>
<td>Multicasting in a Crosspoint Matrix Fabric</td>
</tr>
<tr>
<td>Crosspoint Matrix Implementation</td>
</tr>
<tr>
<td>The Head-of-Line Blocking Problem</td>
</tr>
<tr>
<td>Solving the Head-of-Line Blocking Problem</td>
</tr>
<tr>
<td>Priority Levels in the Switch Fabric</td>
</tr>
<tr>
<td>Input Versus Output Queues</td>
</tr>
<tr>
<td>Input Queues and Shared Memory Switch Fabrics</td>
</tr>
<tr>
<td>Input Queues, Output Queues, and Flow Control</td>
</tr>
<tr>
<td>Switch Data Transmit Path Functions</td>
</tr>
<tr>
<td>Output Filters</td>
</tr>
<tr>
<td>Output Queues and Priority Handling</td>
</tr>
<tr>
<td>Link Aggregation Distributor</td>
</tr>
<tr>
<td>Transmit Flow Control</td>
</tr>
<tr>
<td>Hey, Kids! What Time Is It?</td>
</tr>
<tr>
<td>Port Interfaces (Transmit)</td>
</tr>
</tbody>
</table>

Appendix: Protocol Parsing | 699 |
References | 703 |
Glossary | 711 |
Index | 753 |