

VSI OpenVMS

Cluster Systems

Document Number: DO-DCLUSY-01A

Publication Date: August 2021

This manual describes cluster concepts, procedures, and guidelines for configuring and managing OpenVMS Cluster systems. Except where noted, the procedures and guidelines apply equally to Integrity servers and Alpha systems. This manual also includes information for providing high availability, building-block growth, and unified system management across coupled systems.

Revision Update Information: This is a new manual.

Operating System and Version: VSI OpenVMS I64 Version 8.4-1H1
VSI OpenVMS Alpha 8.4-2L1

Cluster Systems



VMS Software

Copyright © 2021 VMS Software, Inc. (VSI), Burlington, Massachusetts, USA

Legal Notice

Confidential computer software. Valid license from VSI required for possession, use or copying. Consistent with FAR 12.211 and 12.212, Commercial Computer Software, Computer Software Documentation, and Technical Data for Commercial Items are licensed to the U.S. Government under vendor's standard commercial license.

The information contained herein is subject to change without notice. The only warranties for VSI products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. VSI shall not be liable for technical or editorial errors or omissions contained herein.

HPE, HPE Integrity, HPE Alpha, and HPE Proliant are trademarks or registered trademarks of Hewlett Packard Enterprise.

Preface	xv
1. About VSI	xv
2. Introduction	xv
3. Who Should Use This Manual	xv
4. How This Manual Is Organized	xv
5. Related Documents	xvi
6. VSI Encourages Your Comments	xvii
7. OpenVMS Documentation	xvii
8. Conventions	xvii
Chapter 1. Introduction to OpenVMS Cluster System Management	1
1.1. Overview	1
1.1.1. Uses	1
1.1.2. Benefits	1
1.2. Hardware Components	2
1.2.1. Computers	2
1.2.2. Physical Interconnects	2
1.2.3. OpenVMS Galaxy SMC1	3
1.2.4. Storage Devices	4
1.3. Software Components	4
1.3.1. OpenVMS Cluster Software Functions	5
1.4. Communications	5
1.4.1. System Communications	6
1.4.2. Application Communications	7
1.4.3. Cluster Alias	7
1.4.4. fail SAFE IP	7
1.5. System Management	8
1.5.1. Ease of Management	8
1.5.2. Tools and Utilities	8
1.5.3. System Management Tools from OpenVMS Partners	11
1.5.4. Other Configuration Aids	12
Chapter 2. OpenVMS Cluster Concepts	13
2.1. OpenVMS Cluster System Architecture	13
2.1.1. Port Layer	14
2.1.2. SCS Layer	15
2.1.3. System Applications (SYSAPs) Layer	16
2.1.4. Other Layered Components	16
2.2. OpenVMS Cluster Software Functions	16
2.2.1. Functions	16
2.3. Ensuring the Integrity of Cluster Membership	17
2.3.1. Connection Manager	17
2.3.2. Cluster Partitioning	17
2.3.3. Quorum Algorithm	17
2.3.4. System Parameters	18
2.3.5. Calculating Cluster Votes	18
2.3.6. Example	19
2.3.7. Sub-Cluster Selection	19
2.3.8. Quorum Disk	19
2.3.9. Quorum Disk Watcher	20
2.3.10. Rules for Specifying Quorum	21
2.4. State Transitions	21
2.4.1. Adding a Member	21

2.4.2. Losing a Member	22
2.5. OpenVMS Cluster Membership	24
2.5.1. Cluster Group Number	24
2.5.2. Cluster Password	25
2.5.3. Location	25
2.5.4. Example	25
2.6. Synchronizing Cluster Functions by the Distributed Lock Manager	26
2.6.1. Distributed Lock Manager Functions	26
2.6.2. System Management of the Lock Manager	26
2.6.3. Large-Scale Locking Applications	28
2.7. Resource Sharing	28
2.7.1. Distributed File System	28
2.7.2. RMS and the Distributed Lock Manager	28
2.8. Disk Availability	28
2.8.1. MSCP Server	28
2.8.2. Device Serving	29
2.8.3. Enabling the MSCP Server	29
2.9. Tape Availability	29
2.9.1. TMSCP Server	29
2.9.2. Enabling the TMSCP Server	30
2.10. Queue Availability	30
2.10.1. Controlling Queues	30
Chapter 3. OpenVMS Cluster Interconnect Configurations	31
3.1. Overview	31
3.2. OpenVMS Cluster Systems Interconnected by LANs	32
3.2.1. Design	32
3.2.1.1. PEDRIVER Fast Path Support	32
3.2.2. Cluster Group Numbers and Cluster Passwords	32
3.2.3. Servers	32
3.2.4. Satellites	33
3.2.5. Satellite Booting (Alpha)	33
3.2.6. Satellite Booting (Integrity servers)	34
3.2.7. Configuring Multiple LAN Adapters	34
3.2.7.1. System Characteristics	34
3.2.7.2. System Requirements	34
3.2.7.3. Guidelines	35
3.2.8. LAN Examples	35
3.2.9. Fast Path for LAN Devices	36
3.2.10. LAN Bridge Failover Process	36
3.2.11. Virtual LAN Support in OpenVMS	36
3.2.11.1. VLAN Design	37
3.2.11.2. VLAN Support Details	38
3.3. Cluster over IP	38
3.3.1. Design	39
3.3.1.1. PEDRIVER Support for UDP	39
3.3.1.2. TCP/IP Services Boot Time Loading and Initialization	40
3.3.2. Availability	40
3.3.3. System Characteristics	40
3.3.4. Software Requirements	41
3.3.5. Configuration Overview	41
3.3.5.1. Configuration Files	42
3.3.6. Satellite Node Support	42

3.3.7. High Availability Configuration using Logical LAN	42
3.3.8. Performance Guidelines	43
3.3.9. Example	43
3.4. OpenVMS Cluster Systems Interconnected by MEMORYCHANNEL (Alpha Only)	44
3.4.1. Design	44
3.4.2. Examples	44
3.5. Mixed-Interconnect OpenVMS Cluster Systems	45
3.5.1. Availability	45
3.5.2. Examples	45
3.6. Multihost SCSI OpenVMS Cluster Systems	46
3.6.1. Design for OpenVMS Alpha Configurations	46
3.6.2. Design for OpenVMS Integrity server Configurations	47
3.6.3. Examples	47
3.7. Serial Attached SCSI (SAS) (Integrity servers Only)	49
3.8. Multihost Fibre Channel OpenVMS Cluster Systems	49
3.8.1. Design	49
Chapter 4. The OpenVMS Cluster Operating Environment	51
4.1. Preparing the Operating Environment	51
4.2. Installing the OpenVMS Operating System	51
4.2.1. System Disks	51
4.2.2. Where to Install	52
4.2.3. Information Required	52
4.3. Installing Software Licenses	56
4.3.1. Guidelines	56
4.4. Installing Layered Products	56
4.4.1. Procedure	56
4.5. Configuring and Starting a Satellite Booting Service	57
4.5.1. Configuring and Starting the LANCP Utility	59
4.5.2. Booting Satellite Nodes with LANCP	59
4.5.3. Data Files Used by LANCP	59
4.5.4. Using LAN MOP Services in New Installations	60
4.5.5. Using LAN MOP Services in Existing Installations	60
4.5.6. Configuring DECnet	63
4.5.7. Starting DECnet	65
4.5.8. What is Cluster Alias?	65
4.5.9. Enabling Alias Operations	66
4.5.10. Configuring TCP/IP	67
Chapter 5. Preparing a Shared Environment	69
5.1. Shareable Resources	69
5.1.1. Local Resources	69
5.1.2. Sample Configuration	70
5.1.3. Storage in a Mixed-Architecture Cluster	70
5.2. Common-Environment and Multiple-Environment Clusters	71
5.3. Directory Structure on Common System Disks	72
5.3.1. Directory Roots	73
5.3.2. Directory Structure Example	73
5.3.3. Search Order	73
5.4. Clusterwide Logical Names	75
5.4.1. Default Clusterwide Logical Name Tables	75
5.4.2. Translation Order	76
5.4.3. Creating Clusterwide Logical Name Tables	76

5.4.4. Alias Collisions Involving Clusterwide Logical Name Tables	77
5.4.5. Creating Clusterwide Logical Names	77
5.4.6. Management Guidelines	78
5.4.7. Using Clusterwide Logical Names in Applications	79
5.4.7.1. Clusterwide Attributes for \$TRNLNM System Service	79
5.4.7.2. Clusterwide Attribute for \$GETSYI System Service	79
5.4.7.3. Creating Clusterwide Tables with the \$CRELNT System Service	80
5.4.8. Defining and Accessing Clusterwide Logical Names	80
5.4.8.1. Defining Clusterwide Logical Names in SYSTARTUP_VMS.COM	80
5.4.8.2. Defining Certain Logical Names in SYLOGICALS.COM	80
5.4.8.3. Using Conditional Definitions for Startup Command Procedures	81
5.4.9. Displaying Clusterwide Logical Names	81
5.5. Coordinating Startup Command Procedures	83
5.5.1. OpenVMS Startup Procedures	83
5.5.2. Building Startup Procedures	84
5.5.3. Combining Existing Procedures	85
5.5.4. Using Multiple Startup Procedures	85
5.6. Providing OpenVMS Cluster System Security	86
5.6.1. Security Checks	86
5.6.2. Files Relevant to OpenVMS Cluster Security	87
5.7. Network Security	92
5.7.1. Mechanisms	93
5.8. Coordinating System Files	93
5.8.1. Procedure	93
5.8.2. Network Database Files	94
5.9. System Time on the Cluster	94
5.9.1. Setting System Time	95
Chapter 6. Cluster Storage Devices	97
6.1. Data File Sharing	97
6.1.1. Access Methods	97
6.1.2. Examples	98
6.1.3. Specifying a Preferred Path	101
6.2. Naming OpenVMS Cluster Storage Devices	102
6.2.1. Allocation Classes	102
6.2.2. Specifying Node Allocation Classes	103
6.2.2.1. Assigning Node Allocation Class Values on Computers	104
6.2.2.2. Node Allocation Class Example With a DSA Disk and Tape	104
6.2.2.3. Node Allocation Class Example With Mixed Interconnects	105
6.2.2.4. Node Allocation Classes and RAID Array 210 and 230 Devices	106
6.2.3. Reasons for Using Port Allocation Classes	107
6.2.3.1. Constraint of the SCSI Controller Letter in Device Names	108
6.2.3.2. Constraints Removed by Port Allocation Classes	108
6.2.4. Specifying Port Allocation Classes	109
6.2.4.1. Port Allocation Classes for Devices Attached to a Multi-Host Interconnect	109
6.2.4.2. Port Allocation Class 0 for Devices Attached to a Single-Host Interconnect	110
6.2.4.3. Port Allocation Class -1	110
6.2.4.4. How to Implement Port Allocation Classes	111
6.2.4.5. Clusterwide Reboot Requirements for SCSI Interconnects	111
6.3. MSCP and TMSCP Served Disks and Tapes	112
6.3.1. Enabling Servers	112

6.3.1.1. Serving the System Disk	114
6.3.1.2. Setting the MSCP and TMSCP System Parameters	115
6.4. MSCP I/O Load Balancing	115
6.4.1. Load Capacity	115
6.4.2. Available Serving Capacity	116
6.4.3. Static Load Balancing	116
6.4.4. Overriding MSCP I/O Load Balancing for Special Purposes	116
6.5. Managing Cluster Disks With the Mount Utility	116
6.5.1. Mounting Cluster Disks	117
6.5.2. Examples of Mounting Shared Disks	117
6.5.3. Mounting Cluster Disks With Command Procedures	118
6.5.4. Disk Rebuild Operation	118
6.5.5. Rebuilding Cluster Disks	118
6.5.6. Rebuilding System Disks	119
6.6. Shadowing Disks Across an OpenVMS Cluster	120
6.6.1. Purpose	120
6.6.2. Shadow Sets	120
6.6.3. I/O Capabilities	120
6.6.4. Supported Devices	121
6.6.5. Shadow Set Limits	121
6.6.6. Distributing Shadowed Disks	121
Chapter 7. Setting Up and Managing Cluster Queues	123
7.1. Introduction	123
7.2. Controlling Queue Availability	123
7.3. Starting a Queue Manager and Creating the Queue Database	124
7.4. Starting Additional Queue Managers	125
7.4.1. Command Format	125
7.4.2. Database Files	125
7.5. Stopping the Queuing System	126
7.6. Moving Queue Database Files	126
7.6.1. Location Guidelines	126
7.7. Setting Up Print Queues	126
7.7.1. Creating a Queue	127
7.7.2. Command Format	127
7.7.3. Ensuring Queue Availability	128
7.7.4. Examples	128
7.8. Setting Up Clusterwide Generic Print Queues	129
7.8.1. Sample Configuration	129
7.8.2. Command Example	130
7.9. Setting Up Execution Batch Queues	130
7.9.1. Before You Begin	130
7.9.2. Batch Command Format	131
7.9.3. Autostart Command Format	131
7.9.4. Examples	132
7.10. Setting Up Clusterwide Generic Batch Queues	132
7.10.1. Sample Configuration	132
7.11. Starting Local Batch Queues	133
7.11.1. Startup Command Procedure	133
7.12. Using a Common Command Procedure	133
7.12.1. Command Procedure	134
7.12.2. Examples	134
7.12.3. Example	135

7.13. Disabling Autostart During Shutdown	137
7.13.1. Options	137
Chapter 8. Configuring an OpenVMS Cluster System	139
8.1. Overview of the Cluster Configuration Procedures	139
8.1.1. Before Configuring the System	141
8.1.2. Data Requested by the Cluster Configuration Procedures	143
8.1.3. Invoking the Procedure	148
8.2. Adding Computers	149
8.2.1. Controlling Conversational Bootstrap Operations	150
8.2.2. Common AUTOGEN Parameter Files	150
8.2.3. Examples	151
8.2.3.1. Creating and Configuring a Two-Node Cluster Using IP	161
8.2.3.2. Adding a new Node to a Cluster over IP	171
8.2.3.3. Adding a new Node to a Cluster over IP with a Shared System Disk	172
8.2.3.4. Adding an Integrity server Satellite node to a Cluster over IP	179
8.2.3.5. Adding an Integrity server Node to a Cluster over IP with Logical LAN Failover set	188
8.2.4. Adding a Quorum Disk	191
8.3. Removing Computers	192
8.3.1. Example	193
8.3.2. Removing a Quorum Disk	194
8.4. Changing Computer Characteristics	195
8.4.1. Preparation	197
8.4.2. Examples	200
8.5. Creating a Duplicate System Disk	209
8.5.1. Preparation	209
8.5.2. Example	209
8.6. Postconfiguration Tasks	210
8.6.1. Updating Parameter Files	212
8.6.2. Shutting Down the Cluster	213
8.6.3. Shutting Down a Single Node	214
8.6.4. Updating Network Data	214
8.6.5. Altering Satellite Local Disk Labels	215
8.6.6. Changing Allocation Class Values	215
8.6.7. Rebooting	215
8.6.8. Rebooting Satellites Configured with OpenVMS on a Local Disk (Alpha only)	216
8.7. Running AUTOGEN with Feedback	217
8.7.1. Advantages	217
8.7.2. Initial Values	217
8.7.3. Obtaining Reasonable Feedback	218
8.7.4. Creating a Command File to Run AUTOGEN	218
Chapter 9. Building Large OpenVMS Cluster Systems	221
9.1. Setting Up the Cluster	221
9.2. General Booting Considerations	222
9.2.1. Concurrent Booting	222
9.2.2. Minimizing Boot Time	222
9.2.3. General Booting Considerations for Cluster over IP	223
9.3. Booting Satellites	223
9.3.1. Differences between Alpha and Integrity server Satellites	224
9.4. Configuring and Booting Satellite Nodes (Alpha)	224

9.4.1. Booting from a Single LAN Adapter	225
9.4.2. Changing the Default Boot Adapter	225
9.4.3. Booting from Multiple LAN Adapters (Alpha Only)	226
9.4.4. Enabling Satellites to Use Alternate LAN Adapters for Booting	226
9.4.5. Configuring MOP Service	229
9.4.6. Controlling Satellite Booting	229
9.5. Configuring and Booting Satellite Nodes (Integrity servers)	232
9.5.1. Collecting Information from the Satellite System	232
9.5.2. Setting up the Satellite System for Booting and Crashing	233
9.5.3. Defining the Satellite System to the Boot Server	233
9.5.4. Booting the Satellite	234
9.5.5. Additional Tasks on the Satellite System	234
9.6. Booting Satellites with IP interconnect (Integrity servers, Alpha)	234
9.7. System-Disk Throughput	235
9.7.1. Avoiding Disk Rebuilds	235
9.7.2. Offloading Work	236
9.7.3. Configuring Multiple System Disks	237
9.8. Conserving System Disk Space	239
9.8.1. Techniques	239
9.9. Adjusting System Parameters	239
9.9.1. The SCSRESPCNT Parameter	239
9.9.2. The CLUSTER_CREDITS Parameter	240
9.10. Minimize Network Instability	241
9.11. DECnet Cluster Alias	241
Chapter 10. Maintaining an OpenVMS Cluster System	243
10.1. Backing Up Data and Files	243
10.2. Updating the OpenVMS Operating System	244
10.2.1. Rolling Upgrades	244
10.3. LAN Network Failure Analysis	245
10.4. Recording Configuration Data	245
10.4.1. Record Information	245
10.4.2. Satellite Network Data	246
10.5. Controlling OPCOM Messages	246
10.5.1. Overriding OPCOM Defaults	247
10.5.2. Example	247
10.6. Shutting Down a Cluster	248
10.6.1. The NONE Option	248
10.6.2. The REMOVE_NODE Option	248
10.6.3. The CLUSTER_SHUTDOWN Option	248
10.6.4. The REBOOT_CHECK Option	249
10.6.5. The SAVE_FEEDBACK Option	249
10.6.6. Shutting Down TCP/IP	249
10.7. Dump Files	250
10.7.1. Controlling Size and Creation	250
10.7.2. Sharing Dump Files	251
10.8. Maintaining the Integrity of OpenVMS Cluster Membership	252
10.8.1. Cluster Group Data	252
10.8.2. Example	253
10.9. Adjusting Packet Size for LAN or IP Configurations	254
10.9.1. System Parameter Settings for LANs and IPs	254
10.9.2. How to Use NISCS_MAX_PKTSZ	254
10.9.3. How to Use NISCS_UDP_PKTSZ	254

10.9.4. Editing Parameter Files	255
10.10. Determining Process Quotas	255
10.10.1. Quota Values	255
10.10.2. PQL Parameters	255
10.10.3. Examples	256
10.11. Restoring Cluster Quorum	257
10.11.1. Restoring Votes	257
10.11.2. Reducing Cluster Quorum Value	258
10.12. Cluster Performance	259
10.12.1. Using the SHOW Commands	259
10.12.2. Using the Monitor Utility	260
10.12.3. Using VSI Availability Manager	261
10.12.4. Monitoring LAN Activity	261
10.12.5. LAN or PEDRIVER Fast Path Settings	262
Appendix A. Cluster System Parameters	263
A.1. Values	263
Appendix B. Building Common Files	279
B.1. Building a Common SYSUAF.DAT File	279
B.2. Merging RIGHTSLIST.DAT Files	280
Appendix C. Cluster Troubleshooting	283
C.1. Diagnosing Computer Failures	283
C.1.1. Preliminary Checklist	283
C.1.2. Sequence of Booting Events	283
C.2. Satellite Fails to Boot	286
C.2.1. Displaying Connection Messages	287
C.2.2. General OpenVMS Cluster Satellite-Boot Troubleshooting	287
C.2.3. MOP Server Troubleshooting	289
C.2.4. Disk Server Troubleshooting	290
C.2.5. Satellite Booting Troubleshooting	290
C.2.6. Alpha Booting Messages (Alpha Only)	291
C.3. Computer Fails to Join the Cluster	293
C.3.1. Verifying OpenVMS Cluster Software Load	293
C.3.2. Verifying Boot Disk and Root	293
C.3.3. Verifying SCSNODE and SCSSYSTEMID Parameters	294
C.3.4. Verifying Cluster Security Information	294
C.4. Startup Procedures Fail to Complete	295
C.5. Diagnosing LAN Component Failures	295
C.6. Diagnosing Cluster Hangs	295
C.6.1. Cluster Quorum is Lost	296
C.6.2. Inaccessible Cluster Resource	296
C.7. Diagnosing CLUEXIT Bugchecks	296
C.7.1. Conditions Causing Bugchecks	297
C.8. Port Communications	297
C.8.1. LAN Communications	297
C.8.2. System Communications Services (SCS) Connections	298
C.9. Diagnosing Port Failures	298
C.9.1. Hierarchy of Communication Paths	298
C.9.2. Where Failures Occur	298
C.9.3. Verifying Virtual Circuits	299
C.9.4. Verifying LAN Connections	300
C.10. Analyzing Error-Log Entries for Port Devices	300

C.10.1. Examine the Error Log	300
C.10.2. Formats	301
C.10.3. LAN Device-Attention Entries	301
C.10.4. Logged Message Entries	303
C.10.5. Error-Log Entry Descriptions	303
C.11. OPA0 Error-Message Logging and Broadcasting	310
C.11.1. OPA0 Error Messages	310
C.12. Integrity server Satellite Booting Messages	312
Appendix D. Sample Programs for LAN Control	315
D.1. Purpose of Programs	315
D.2. Starting the NISCA Protocol	315
D.2.1. Start the Protocol	315
D.3. Stopping the NISCA Protocol	316
D.3.1. Stop the Protocol	316
D.3.2. Verify Successful Execution	316
D.4. Analyzing Network Failures	317
D.4.1. Failure Analysis	317
D.4.2. How the LAVC\$FAILURE_ANALYSIS Program Works	317
D.5. Using the Network Failure Analysis Program	317
D.5.1. Create a Network Diagram	318
D.5.2. Edit the Source File	320
D.5.3. Assemble and Link the Program	323
D.5.4. Modify Startup Files	323
D.5.5. Execute the Program	324
D.5.6. Modify MODPARAMS.DAT	324
D.5.7. Test the Program	324
D.5.8. Display Suspect Components	324
Appendix E. Subroutines for LAN Control	327
E.1. Introduction	327
E.1.1. Purpose of the Subroutines	327
E.2. Starting the NISCA Protocol	327
E.2.1. Status	328
E.2.2. Error Messages	328
E.3. Stopping the NISCA Protocol	329
E.3.1. Status	329
E.3.2. Error Messages	330
E.4. Creating a Representation of a Network Component	331
E.4.1. Status	331
E.4.2. Error Messages	331
E.5. Creating a Network Component List	332
E.5.1. Status	333
E.5.2. Error Messages	334
E.6. Starting Network Component Failure Analysis	335
E.6.1. Status	335
E.6.2. Error Messages	335
E.7. Stopping Network Component Failure Analysis	335
E.7.1. Status	335
E.7.2. Error Messages	336
Appendix F. Troubleshooting the NISCA Protocol	337
F.1. How NISCA Fits into the SCA	337
F.1.1. SCA Protocols	337

F.1.2. Paths Used for Communication	341
F.1.3. PEDRIVER	342
F.2. Addressing LAN Communication Problems	343
F.2.1. Symptoms	343
F.2.2. Traffic Control	343
F.2.3. Excessive Packet Losses on LAN Paths	344
F.2.4. Preliminary Network Diagnosis	344
F.2.5. Tracing Intermittent Errors	344
F.2.6. Checking System Parameters	345
F.2.7. Channel Timeouts	346
F.3. Using SDA to Monitor LAN or IP Communications	346
F.3.1. Isolating Problem Areas	347
F.3.2. SDA Command SHOW PORT	347
F.3.3. Monitoring Virtual Circuits	347
F.3.4. Monitoring PEDRIVER for LAN devices	351
F.3.5. Monitoring PEDRIVER Buses for LAN Devices	351
F.3.6. Monitoring LAN Adapters	352
F.3.7. Monitoring PEDRIVER Buses for IP interfaces	355
F.3.8. Monitoring PEDRIVER Channels for IP Interfaces	356
F.4. Using SCACP to Monitor Cluster Communications	357
F.5. Troubleshooting NISCA Communications	358
F.5.1. Areas of Trouble	358
F.6. Channel Formation	358
F.6.1. How Channels Are Formed	358
F.6.2. Techniques for Troubleshooting	359
F.7. Retransmission Problems	360
F.7.1. Why Retransmissions Occur	360
F.7.2. Techniques for Troubleshooting	362
F.8. Understanding NISCA Datagrams	362
F.8.1. Packet Format	362
F.8.2. LAN Headers	362
F.8.3. Ethernet Header	362
F.8.4. Datagram Exchange (DX) Header	363
F.8.5. Channel Control (CC) Header	364
F.8.6. Transport (TR) Header	365
F.9. Using a LAN Protocol Analysis Program	367
F.9.1. Single or Multiple LAN Segments	367
F.9.2. Multiple LAN Segments	367
F.10. Data Isolation Techniques	368
F.10.1. All OpenVMS Cluster Traffic	368
F.10.2. Specific OpenVMS Cluster Traffic	368
F.10.3. Virtual Circuit (Node-to-Node) Traffic	368
F.10.4. Channel (LAN Adapter-to-LAN Adapter) Traffic	369
F.10.5. Channel Control Traffic	369
F.10.6. Transport Data	369
F.11. Setting Up an HP 4972A LAN Protocol Analyzer	369
F.11.1. Analyzing Channel Formation Problems	370
F.11.2. Analyzing Retransmission Problems	370
F.12. Filters	372
F.12.1. Capturing All LAN Retransmissions for a Specific OpenVMS Cluster	372
F.12.2. Capturing All LAN Packets for a Specific OpenVMS Cluster	372
F.12.3. Setting Up the Distributed Enable Filter	373

F.12.4. Setting Up the Distributed Trigger Filter	373
F.13. Messages	373
F.13.1. Distributed Enable Message	373
F.13.2. Distributed Trigger Message	374
F.14. Programs That Capture Retransmission Errors	374
F.14.1. Starter Program	374
F.14.2. Partner Program	375
F.14.3. Scribe Program	376
Appendix G. NISCA Transport Protocol Congestion Control	377
G.1. NISCA Congestion Control	377
G.1.1. Congestion Caused by Retransmission	377
G.1.1.1. OpenVMS VAX Version 6.0 or OpenVMS AXP Version 1.5, or Later	377
G.1.2. HELLO Multicast Datagrams	378
G.1.3. HELLO IP Unicast and IP Multicast Datagrams	379

Preface

This manual describes cluster concepts, procedures, and guidelines for configuring and managing OpenVMS Cluster systems. Except where noted, the procedures and guidelines apply equally to Integrity servers and Alpha systems. This manual also includes information for providing high availability, building-block growth, and unified system management across coupled systems.

1. About VSI

VMS Software, Inc. (VSI) is an independent software company licensed by Hewlett Packard Enterprise to develop and support the OpenVMS operating system.

VSI seeks to continue the legendary development prowess and customer-first priorities that are so closely associated with the OpenVMS operating system and its original author, Digital Equipment Corporation.

2. Introduction

VSI OpenVMS Cluster Systems describe system management for OpenVMS Cluster systems. Although the OpenVMS Cluster software for Integrity servers and Alpha computers is separately purchased, licensed, and installed, the difference between the two architectures lies mainly in the hardware used. Essentially, system management for Integrity servers and Alpha computers in an OpenVMS Cluster is identical. Exceptions are pointed out.

Note

This manual is applicable only for a combination of Integrity server systems and Alpha systems.

3. Who Should Use This Manual

This document is intended for anyone responsible for setting up and managing OpenVMS Cluster systems. To use the document as a guide to cluster management, you must have a thorough understanding of system management concepts and procedures, as described in the *VSI OpenVMS System Manager's Manual*.

4. How This Manual Is Organized

VSI OpenVMS Cluster Systems contains of the following chapters and appendixes.

Chapter 1 introduces OpenVMS Cluster systems.

Chapter 2 presents the software concepts integral to maintaining OpenVMS Cluster membership and integrity.

Chapter 3 describes various OpenVMS Cluster configurations and the ways they are interconnected.

Chapter 4 explains how to set up an OpenVMS Cluster system and coordinate system files.

Chapter 5 explains how to set up an environment in which resources can be shared across nodes in the OpenVMS Cluster system.

Chapter 6 discusses disk and tape management concepts and procedures and how to use Volume Shadowing for OpenVMS to prevent data unavailability.

Chapter 7 discusses queue management concepts and procedures.

Chapter 8 explains how to build an OpenVMS Cluster system once the necessary preparations are made, and how to reconfigure and maintain the cluster.

Chapter 9 provides guidelines for configuring and building large OpenVMS Cluster systems, booting satellite nodes, and cross-architecture booting.

Chapter 10 describes ongoing OpenVMS Cluster system maintenance.

Appendix A lists and defines OpenVMS Cluster system parameters.

Appendix B provides guidelines for building a cluster common user authorization file.

Appendix C provides troubleshooting information.

Appendix D presents three sample programs for LAN control and explains how to use the Local Area OpenVMS Cluster Network Failure Analysis Program.

Appendix E describes the subroutine package used with local area OpenVMS Cluster sample programs.

Appendix F provides techniques for troubleshooting network problems related to the NISCA transport protocol.

Appendix G describes how the interactions of work load distribution and network topology affect OpenVMS Cluster system performance, and discusses transmit channel selection by PEDRIVER.

5. Related Documents

This document is not a one-volume reference manual. The utilities and commands are described in detail in the *VSI OpenVMS System Manager's Manual*, the *VSI OpenVMS System Management Utilities Reference Manual*, and the *VSI OpenVMS DCL Dictionary*.

For additional information on the topics covered in this manual, see the following documents:

- *Guidelines for OpenVMS Cluster Configurations*
- *VSI OpenVMS Alpha Partitioning and Galaxy Guide*
- *VSI OpenVMS Guide to OpenVMS File Applications*
- *VSI OpenVMS Guide to System Security*
- *OpenVMS Alpha System Dump Analyzer Utility Manual*
- *VMS System Dump Analyzer Utility Manual*
- *VSI OpenVMS I/O User's Reference Manual*
- *VSI OpenVMS License Management Utility Manual*
- *VSI OpenVMS System Management Utilities Reference Manual*

- *VSI OpenVMS System Manager's Manual*
- *A Comparison of System Management on OpenVMS AXP and OpenVMS VAX*
- *VSI OpenVMS System Services Reference Manual*
- *VSI OpenVMS Volume Shadowing Guide*
- *VSI OpenVMS Software Product Descriptions* (available on the VSI OpenVMS website: <https://vmssoftware.com/resources/documentation/>)
- *VSI OpenVMS DECnet Network Management Utilities*
- *VSI OpenVMS DECnet Networking Manual*
- The VSI DECnet-Plus (formerly known as DECnet/OSI) documentation set
- The TCP/IP Services for OpenVMS documentation set

6. VSI Encourages Your Comments

You may send comments or suggestions regarding this manual or any VSI document by sending electronic mail to the following Internet address: <docinfo@vmssoftware.com>. Users who have OpenVMS support contracts through VSI can contact <support@vmssoftware.com> for help with this product.

7. OpenVMS Documentation

The full VSI OpenVMS documentation set can be found on the VMS Software Documentation webpage at <https://vmssoftware.com/resources/documentation/>

8. Conventions

The following conventions are used in this manual:

Convention	Meaning
...	A horizontal ellipsis in examples indicates one of the following possibilities: <ul style="list-style-type: none"> • Additional optional arguments in a statement have been omitted. • The preceding item or items can be repeated one or more times. • Additional parameters, values, or other information can be entered.
:	A vertical ellipsis indicates the omission of items from a code example or command format; the items are omitted because they are not important to the topic being discussed.
()	In command format descriptions, parentheses indicate that you must enclose choices in parentheses if you specify more than one.
[]	In command format descriptions, brackets indicate optional choices. You can choose one or more items or no items. Do not type the brackets on the command line. However, you must include the brackets in the syntax for

Convention	Meaning
	OpenVMS directory specifications and for a substring specification in an assignment statement.
	In command format descriptions, vertical bars separate choices within brackets or braces. Within brackets, the choices are optional; within braces, at least one choice is required. Do not type the vertical bars on the command line.
{ }	In command format descriptions, braces indicate required choices; you must choose at least one of the items listed. Do not type the braces on the command line.
bold text	This typeface represents the name of an argument, an attribute, or a reason.
<i>italic text</i>	Italic text indicates important information, complete titles of manuals, or variables. Variables include information that varies in system output (Internal error <i>number</i>), in command lines (<i>/PRODUCER= name</i>), and in command parameters in text (where <i>dd</i> represents the predefined code for the device type).
UPPERCASE TEXT	Uppercase text indicates a command, the name of a routine, the name of a file, or the abbreviation for a system privilege.
Monospace text	Monospace type indicates code examples and interactive screen displays. In the C programming language, monospace type in text identifies the following elements: keywords, the names of independently compiled external functions and files, syntax summaries, and references to variables or identifiers introduced in an example.
-	A hyphen at the end of a command format description, command line, or code line indicates that the command or statement continues on the following line.
numbers	All numbers in text are assumed to be decimal unless otherwise noted. Nondecimal radices—binary, octal, or hexadecimal—are explicitly indicated.

Chapter 1. Introduction to OpenVMS Cluster System Management

“Cluster” technology was pioneered by Digital Equipment Corporation in 1983 with the VAXcluster system. The VAXcluster system was built using multiple standard VAX computing systems and the VMS operating system. The initial VAXcluster system offered the power and manageability of a centralized system and the flexibility of many physically distributed computing systems.

Through the years, the technology has evolved to support mixed-architecture cluster systems and the name changed to OpenVMS Cluster systems. Initially, OpenVMS Alpha and OpenVMS VAX systems were supported in a mixed-architecture OpenVMS Cluster system. In OpenVMS Version 8.2, cluster support was introduced for the OpenVMS Integrity server systems either in a single architecture cluster or in a mixed-architecture cluster with OpenVMS Alpha systems. VSI continues to enhance and expand OpenVMS Cluster capabilities.

1.1. Overview

An **OpenVMS Cluster system** is a highly integrated organization of OpenVMS software, Alpha, VAX, or Integrity servers or a combination of Alpha and VAX or Alpha and Integrity servers, and storage devices that operate as a single system. The OpenVMS Cluster acts as a single virtual system, even though it is made up of many distributed systems. As members of an OpenVMS Cluster system, Alpha and VAX or Alpha and Integrity server systems can share processing resources, data storage, and queues under a single security and management domain, yet they can boot or shut down independently.

The distance between the computers in an OpenVMS Cluster system depends on the interconnects that you use. The computers can be located in one computer lab, on two floors of a building, between buildings on a campus, or on two different sites hundreds of miles apart.

An OpenVMS Cluster system, with computers located on two or more sites, is known as a multiple-site OpenVMS Cluster system. For more information about multiple site clusters, see the *Guidelines for OpenVMS Cluster Configurations*.

1.1.1. Uses

OpenVMS Cluster systems are an ideal environment for developing high-availability applications, such as transaction processing systems, servers for network client or server applications, and data-sharing applications.

1.1.2. Benefits

Computers in an OpenVMS Cluster system interact to form a cooperative, distributed operating system and derive a number of benefits, as shown in the following table.

Benefit	Description
Resource sharing	OpenVMS Cluster software automatically synchronizes and load balances batch and print queues, storage devices, and other resources among all cluster members.
Flexibility	Application programmers do not have to change their application code, and users do not have to know anything about the OpenVMS Cluster environment to take advantage of common resources.

Benefit	Description
High availability	System designers can configure redundant hardware components to create highly available systems that eliminate or withstand single points of failure.
Nonstop processing	The OpenVMS operating system, which runs on each node in an OpenVMS Cluster, facilitates dynamic adjustments to changes in the configuration.
Scalability	Organizations can dynamically expand computing and storage resources as business needs grow or change without shutting down the system or applications running on the system.
Performance	An OpenVMS Cluster system can provide high performance.
Management	Rather than repeating the same system management operation on multiple OpenVMS systems, management tasks can be performed concurrently for one or more nodes.
Security	Computers in an OpenVMS Cluster share a single security database that can be accessed by all nodes in a cluster.
Load balancing	OpenVMS Cluster systems distribute work across cluster members based on the current load of each member.

1.2. Hardware Components

OpenVMS Cluster system configurations consist of hardware components from the following general groups:

- Computers
- Interconnects
- Storage devices

References: Detailed OpenVMS Cluster configuration guidelines can be found in the VSI OpenVMS Cluster *Software Product Description (SPD)* and in *Guidelines for OpenVMS Cluster Configurations*.

1.2.1. Computers

Up to 96 computers, ranging from desktop to mainframe systems, can be **members** of an OpenVMS Cluster system. Active members that run the OpenVMS Alpha or OpenVMS Integrity server operating system and participate fully in OpenVMS Cluster negotiations can include:

- Integrity server computers or workstations
- Alpha computers or workstations

1.2.2. Physical Interconnects

An **interconnect** is a physical path that connects computers to other computers and to storage subsystems. OpenVMS Cluster systems support a variety of interconnects (also referred to as buses) so that members can communicate using the most appropriate and effective method possible:

- LANs
 - Ethernet

- Fast Ethernet
- Gigabit Ethernet
- 10 Gigabit Ethernet (Integrity servers only)
- Internet Protocol (IP)
 - Fast Ethernet
 - Gigabit Ethernet
 - 10 Gigabit Ethernet (Integrity servers only)
- MEMORY CHANNEL (node to node communications, Alpha only)
- Serial Attached SCSI (SAS) (node-to-storage only, Integrity servers only)
- Small Computer Systems Interface (SCSI) (node-to-storage only)
- Fibre Channel (FC) (node-to-storage only)

Note

The CI, DSSI, and FDDI interconnects are supported on Alpha and VAX systems. Memory Channel and ATM interconnects are supported only on Alpha systems.

Table 1.1. Interconnect Support by OpenVMS Platform

Interconnect	Platform Support	Comments
IP: UDP	Integrity servers and Alpha	Supports Fast Ethernet/Gigabit Ethernet/10 Gb Ethernet. 10 Gb Ethernet is supported on Integrity servers only.
Fibre Channel	Integrity servers and Alpha	Shared storage only
SAS	Integrity servers	
SCSI	Integrity servers and Alpha	Limited shared storage configurations only
LAN: Ethernet, Fast Ethernet, Gigabit Ethernet, and 10 Gb Ethernet	Integrity servers and Alpha	10 Gb Ethernet is supported on Integrity servers only
MEMORY CHANNEL	Alpha	Node-to-node communications only

For the most recent list of supported interconnects and speeds, see the VSI OpenVMS Cluster Software *Software Product Description*.

1.2.3. OpenVMS Galaxy SMCI

In addition to the physical interconnects listed in Section 1.2.2, another type of interconnect, a shared memory CI (SMCI) for OpenVMS Galaxy instances, is available. SMCI supports cluster communications between Galaxy instances.

For more information about SMCI and Galaxy configurations, see the *VSI OpenVMS Alpha Partitioning and Galaxy Guide*.

1.2.4. Storage Devices

A **shared** storage device is a disk or tape that is accessed by multiple computers in the cluster. Nodes access remote disks and tapes by means of the MSCP and TMSCP server software (described in Section 1.3.1).

Systems within an OpenVMS Cluster support a wide range of storage devices:

- Disks and disk drives, including:
 - Fibre Channel (FC) disks
 - SAS devices
 - SCSI devices
 - Embedded devices, such as IDE and USB devices
 - Digital Storage Architecture (DSA) disks
 - RF series integrated storage elements (ISEs)
 - Solid state disks
- Tapes and tape drives
- Controllers and I/O servers, including the following:

Controller	Interconnect
HSG and HSV	FC
LSI 1068 and LSI Logic 1068e	SAS
HSZ	SCSI

In addition, the K.scsi HSC controller allows the connection of the StorageWorks arrays with SCSI devices on the HSC storage subsystems.

Note: HSZ controllers support many combinations of SDIs (standard disk interfaces) and STIs (standard tape interfaces) that connect disks and tapes.

For the most recent list of supported storage devices, see the *VSI OpenVMS Version 8.4 Software Product Description*.

In case AlphaServer Supported Options Lists or Integrity servers Supported Options Lists are needed, please contact to the HPE Support: <https://support.hpe.com/>.

1.3. Software Components

The OpenVMS operating system, which runs on each node in an OpenVMS Cluster, includes several software components that facilitate resource sharing and dynamic adjustments to changes in the underlying hardware configuration.

If one computer becomes unavailable, the OpenVMS Cluster system continues operating because OpenVMS is still running on the remaining computers.

1.3.1. OpenVMS Cluster Software Functions

The following table describes the software components and their main function.

Component	Facilitates	Function
Connection manager	Member integrity	Coordinates participation of computers in the cluster and maintains cluster integrity when computers join or leave the cluster.
Distributed lock manager	Resource synchronization	Synchronizes operations of the distributed file system, job controller, device allocation, and other cluster facilities. If an OpenVMS Cluster computer shuts down, all locks that it holds are released so that processing can continue on the remaining computers.
Distributed file system	Resource sharing	Allows all computers to share access to mass storage and file records, regardless of the type of storage device (DSA, RF, SCSI, and solid state subsystem) or its location.
Distributed job controller	Queuing	Makes generic and execution queues available across the cluster.
MSCP server	Disk serving	Implements the proprietary mass storage control protocol in order to make disks available to all nodes that do not have direct access to those disks.
TMSCP server	Tape serving	Implements the proprietary tape mass storage control protocol in order to make tape drives available to all nodes that do not have direct access to those tape drives.

1.4. Communications

The System Communications Architecture (SCA) defines the communications mechanisms that allow nodes in an OpenVMS Cluster system to co-operate. SCA governs the sharing of data between resources at the nodes and binds together System Applications (SYSAPs) that run on different Integrity server systems and Alpha computers.

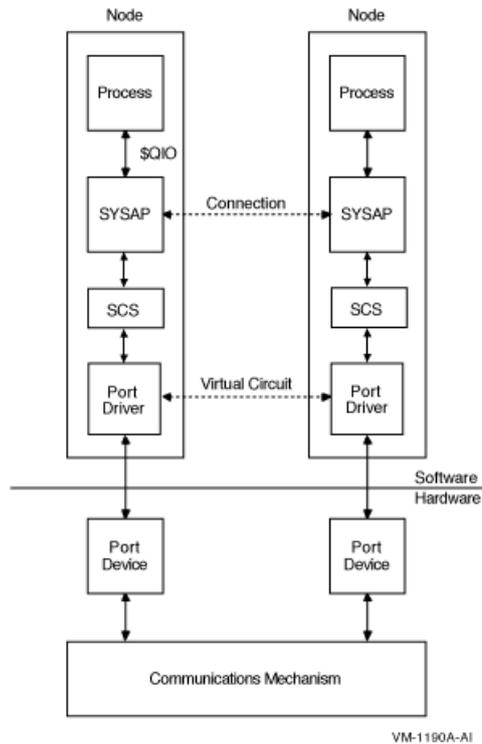
SCA consists of the following hierarchy of components:

Communications Software	Function
System applications (SYSAPs)	Consists of clusterwide applications (for example, disk and tape class drivers, connection manager, and MSCP server) that use SCS software for interprocessor communication.
System Communications Services (SCS)	Provides basic connection management and communication services, implemented as a logical path between system applications (SYSAPs) on nodes in an OpenVMS Cluster system.
Port drivers	Control the communication paths between local and remote ports.
Physical interconnects	Consists of ports or adapters for CI, DSSI, Ethernet, ATM, FDDI, and MEMORY CHANNEL interconnects. PEDRIVER is the port driver for LAN (Ethernet) interconnect and starting with OpenVMS Version 8.4 PEDRIVER is also enabled to use TCP/IP for cluster communication.

1.4.1. System Communications

Figure 1.1 shows the relationship between OpenVMS Cluster components.

Figure 1.1. OpenVMS Cluster System Communications



In Figure 1.1, processes in different nodes exchange information with each other.

- Processes can call the \$QIO system service and other system services directly from a program or indirectly using other mechanisms such as OpenVMS Record Management Services (RMS). The \$QIO system service initiates all I/O requests.
- A SYSAP on one OpenVMS Cluster node communicates with a SYSAP on another node using a logical path called **connection**. For example, a connection manager on one node communicates with the connection manager on another node, or a disk class driver on one node communicates with the MSCP server on another node. The following SYSAPs use SCS for cluster communication:
 - Disk and tape class drivers
 - MSCP server
 - TMSCP server
 - DECnet class driver
 - Connection manager
 - SCA\$TRANSPORT, which provides node-to-node communications to the intracluster communication (ICC) programming interface, available through ICC system services, and to the distributed queue manager

- SCS routines provide connection setup and services to format and transfer SYSAP messages to a port driver for delivery over a specific interconnect.
- Communications go through the port drivers to port drivers on other OpenVMS Cluster computers and storage controllers. A port driver manages a logical path, called a **virtual circuit**, between each pair of ports in an OpenVMS Cluster system. A virtual circuit provides reliable message delivery for the connections multiplexed upon it.

Starting with OpenVMS Version 8.4, cluster systems can use Transmission Control Protocol and Internet Protocol (TCP/IP) stack for cluster communication. PEDRIVER is enhanced with the capability to use TCP/IP in addition to LAN for cluster communication. For more information, see Chapter 3.

1.4.2. Application Communications

Applications running on OpenVMS Cluster systems use TCP/IP, DECnet, or ICC for application communication.

ICC allows processes to efficiently exchange messages with processes running on other cluster members using system communications services and the underlying cluster interconnect. The DECnet and TCP/IP communication services allow processes to locate or start remote servers and then exchange messages.

Note

The generic references to DECnet in this document mean either DECnet for OpenVMS or DECnet-Plus (formerly known as DECnet/OSI) software.

1.4.3. Cluster Alias

DECnet provides a feature known as a **cluster alias**. A cluster alias is a collective name for the nodes in an OpenVMS Cluster system.

Application software can use the cluster alias as the name to connect to a node in the OpenVMS Cluster. DECnet chooses the node to which the application makes a connection. The use of a cluster alias frees the application from keeping track of individual nodes in the OpenVMS Cluster system and results in design simplification, configuration flexibility, and application availability. It also provides a mechanism for load balancing by distributing incoming connections across the nodes comprising the cluster.

1.4.4. fail SAFE IP

TCP/IP provides a feature known as a failSAFE IP that allows IP addresses to failover when interfaces cease functioning on a system, where multiple interfaces have been configured with the same IP address.

You can configure a standby failover target IP address that failSAFE IP assigns to multiple interfaces on a node or across the OpenVMS Cluster system. When, for example, a Network Interface Controller fails or a cable breaks or disconnects, failSAFE IP activates the standby IP address so that an alternate interface can take over to maintain the network connection. If an address is not preconfigured with a standby, then failSAFE IP removes the address from the failed interface until it recovers. When the failed interface recovers, failSAFE IP detects this and can return its IP address.

1.5. System Management

The OpenVMS Cluster system manager must manage multiple users and resources for maximum productivity and efficiency while maintaining the necessary security.

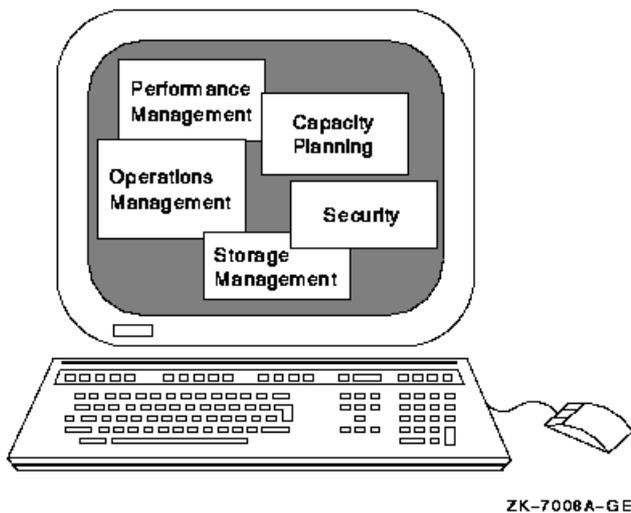
1.5.1. Ease of Management

An OpenVMS Cluster system is easily managed because the multiple members, hardware, and software are designed to co-operate as a single system:

- Smaller configurations usually include only one system disk (or two for an OpenVMS Cluster configuration with both OpenVMS Alpha and OpenVMS Integrity server operating systems), regardless of the number or location of computers in the configuration.
- Software must be installed only once for each operating system (Alpha or Integrity servers), and is accessible by every user and node of the OpenVMS Cluster.
- Users must be added once to access the resources of the entire OpenVMS Cluster.
- Several system management utilities and commands facilitate cluster management.

Figure 1.2 illustrates centralized system management.

Figure 1.2. Single-Point OpenVMS Cluster System Management



1.5.2. Tools and Utilities

The OpenVMS operating system supports a number of utilities and tools to assist you with the management of the distributed resources in OpenVMS Cluster configurations. Proper management is essential to ensure the availability and performance of OpenVMS Cluster configurations.

OpenVMS and its partners offer a wide selection of tools to meet diverse system management needs. Table 1.2 describes the products available for cluster management and indicates whether each is supplied with the operating system or is an optional product, which is purchased separately.

Table 1.2. System Management Tools

Tool	Supplied or Optional	Function
Accounting		
VMS Accounting	Supplied	Tracks how resources are being used.
Configuration and capacity planning		
LMF (License Management Facility)	Supplied	Helps the system manager determine which software products are licensed and installed on a standalone system and on each computer in an OpenVMS Cluster system.
SYSGEN (System Generation) utility	Supplied	Allows you to tailor your system for a specific hardware and software configuration. Use SYSGEN to modify system parameters, load device drivers, and create additional page and swap files.
CLUSTER_CONFIG.COM	Supplied	Automates the configuration or reconfiguration of an OpenVMS Cluster system and assumes the use of DECnet.
CLUSTER_CONFIG_LAN.COM	Supplied	Automates configuration or reconfiguration of an OpenVMS Cluster system without the use of DECnet.
HPE Management Agents for OpenVMS	Supplied	Consists of a web server for system management with management agents that allow you to look at devices on your OpenVMS systems.
HPE Insight Manager XE	Supplied with every HPNT server	Centralizes system management in one system to reduce cost, improve operational efficiency and effectiveness, and minimize system down time. You can use HPE Insight Manager XE on an NT server to monitor every system in an OpenVMS Cluster system. In a configuration of heterogeneous VSI systems, you can use HPE Insight Manager XE on an NT server to monitor all systems.
Event and fault tolerance		
OPCOM message routing	Supplied	Provides event notification.
Operations management		
Clusterwide process services	Supplied	Allows OpenVMS system management commands, such as SHOW USERS, SHOW SYSTEM, and STOP/ID=, to operate clusterwide.
Availability Manager	Supplied	From either an OpenVMS system or a Windows node, enables you to monitor one or more OpenVMS nodes on an extended LAN or wide area network (WAN). That is, the nodes for which you are collecting the information must be in the same extended LAN and there should be an interface that communicates with the collector nodes as well as the WAN analyzer.

Tool	Supplied or Optional	Function
		The Availability Manager collects system and process data from multiple OpenVMS nodes simultaneously, and then analyzes the data and displays the output using a native Java GUI.
HPE WBEM Services for OpenVMS	Supplied	WBEM (Web-Based Enterprise Management) enables management applications to retrieve system information and request system operations wherever and whenever required. It allows customers to manage their systems consistently across multiple platforms and operating systems, providing integrated solutions that optimize your infrastructure for greater operational efficiency.
SCACP (Systems Communications Architecture Control Program)	Supplied	Enables you to monitor, manage, and diagnose cluster communications and cluster interconnects.
DNS (Distributed Name Service)	Optional	Configures certain network nodes as name servers that associate objects with network names.
LATCP (Local Area Transport Control Program)	Supplied	Provides the function to control and obtain information from LAT port driver.
LANCP (LAN Control Program)	Supplied	Allows the system manager to configure and control the LAN software on OpenVMS systems.
NCP (Network Control Protocol) utility	Optional	Allows the system manager to supply and access information about the DECnet for OpenVMS (Phase IV) network from a configuration database.
NCL (Network Control Language) utility	Optional	Allows the system manager to supply and access information about the DECnet-Plus network from a configuration database.
POLYCENTER Software Installation Utility (PCSI)	Supplied	Provides rapid installations of software products.
Queue Manager	Supplied	Uses OpenVMS Cluster generic and execution queues to feed node-specific queues across the cluster.
Show Cluster utility	Supplied	Monitors activity and performance in an OpenVMS Cluster configuration, then collects and sends information about that activity to a terminal or other output device.
SDA (System Dump Analyzer)	Supplied	Allows you to inspect the contents of memory saved in the dump taken at crash time or as it exists in a running system. You can use SDA interactively or in batch mode.
SYSMAN (System Management utility)	Supplied	Enables device and processor control commands to take effect across an OpenVMS Cluster.
VMSINSTAL	Supplied	Provides software installations.
Performance		

Tool	Supplied or Optional	Function
AUTOGEN utility	Supplied	Optimizes system parameter settings based on usage.
Monitor utility	Supplied	Provides basic performance data.
Security		
Authorize utility	Supplied	Modifies user account profiles.
SET ACL command	Supplied	Sets complex protection on many system objects.
SET AUDIT command	Supplied	Facilitates tracking of sensitive system objects.
Storage management		
Backup utility	Supplied	Allows OpenVMS Cluster system managers to create backup copies of files and directories from storage media and then restore them. This utility can be used on one node to back up data stored on disks throughout the OpenVMS Cluster system.
Mount utility	Supplied	Enables a disk or tape volume for processing by one computer, a subset of OpenVMS Cluster computers, or all OpenVMS Cluster computers.
Volume Shadowing for OpenVMS	Optional	Replicates disk data across multiple disks to help OpenVMS Cluster systems survive disk failures.

1.5.3. System Management Tools from OpenVMS Partners

OpenVMS Partners offer a wide selection of tools to meet diverse system management needs. The types of tools are described in the following list:

- Schedule managers

Enable specific actions to be triggered at determined times, including repetitive and periodic activities, such as nightly backups.

- Event managers

Monitor a system and report occurrences and events that may require an action or that may indicate a critical or alarming situation, such as low memory or an attempted security breaking.

- Console managers

Enable a remote connection to and emulation of a system console so that system messages can be displayed and commands can be issued.

- Performance managers

Monitor system performance by collecting and analyzing data to allow proper tailoring and configuration of system resources. Performance managers might also collect historical data for capacity planning.

For information about OpenVMS Partners and the tools they provide, contact to the HPE Support: <https://support.hpe.com/>

1.5.4. Other Configuration Aids

In addition to these utilities and partner products, several commands are available that allow the system manager to set parameters on Fibre Channel, SCSI and SAS storage subsystems to help configure and manage the system. See the appropriate hardware documentation for more information.

Chapter 2. OpenVMS Cluster Concepts

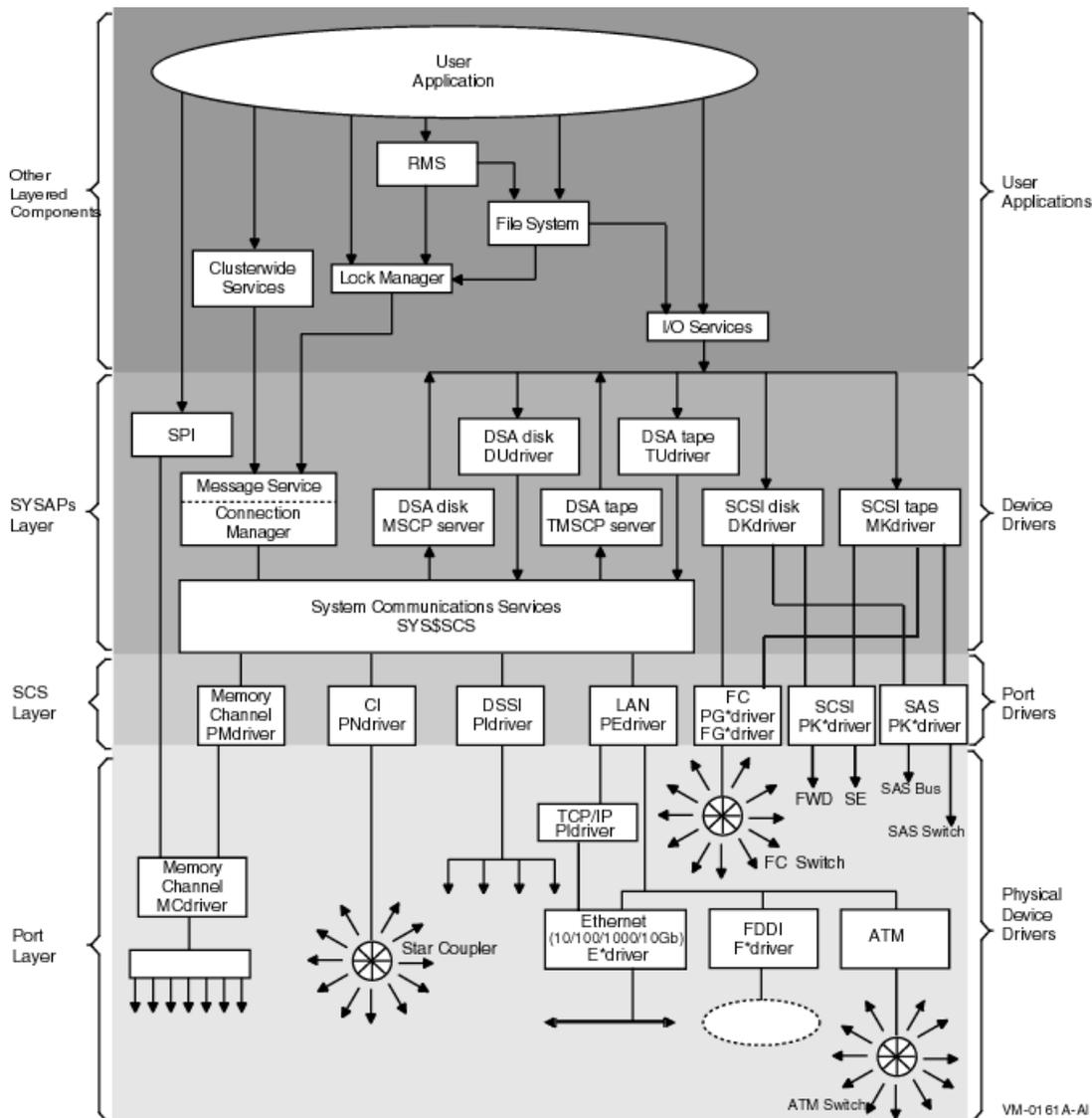
To help you understand the design and implementation of an OpenVMS Cluster system, this chapter describes its basic architecture.

2.1. OpenVMS Cluster System Architecture

Figure 2.1 illustrates the protocol layers within the OpenVMS Cluster system architecture, ranging from the communications mechanisms at the base of the figure to the users of the system at the top of the figure. These protocol layers include:

- Ports
- System Communications Services (SCS)
- System Applications (SYSAPs)
- Other layered components

Figure 2.1. OpenVMS Cluster System Architecture



Note

Not all interconnects are supported on all three architectures of OpenVMS. The CI, DSSI, and FDDI interconnects are supported on Alpha and VAX systems. Memory Channel and ATM interconnects are supported only on Alpha systems.

2.1.1. Port Layer

This lowest level of the architecture provides connections, in the form of communication ports and physical paths, between devices. The port layer can contain any of the following interconnects:

- LANs
 - Ethernet (Fast Ethernet, Gigabit Ethernet and 10 Gb Ethernet)
- Internet Protocol
 - Ethernet (Fast Ethernet, Gigabit Ethernet and 10 Gb Ethernet)

- MEMORY CHANNEL
- SAS
- SCSI
- Fibre Channel

Each interconnect is accessed by a port (also referred to as an adapter) that connects to the processor node. For example, the Fibre Channel interconnect is accessed by way of a Fibre Channel port.

2.1.2. SCS Layer

The SCS layer provides basic connection management and communications services in the form of datagrams, messages, and block transfers over each logical path. Table 2.1 describes these services.

Table 2.1. Communications Services

Service	Delivery Guarantees	Usage
Datagrams		
Information units that fit in 1 packet or less.	Delivery of datagrams is not guaranteed. Datagrams can be lost, duplicated, or delivered out of order.	Status and information messages whose loss is not critical. Applications that have their own reliability protocols such as DECnet or TCP/IP.
Messages		
Information units that fit in 1 packet or less.	Messages are guaranteed to be delivered and to arrive in order. Virtual circuit sequence numbers are used on the individual packets.	Disk read and write requests.
Block data transfers		
Copying (that is, reading or writing) any contiguous data between a local process or system virtual address space and an address on another node. Individual transfers are limited to the lesser of 2^{32-1} bytes, or the physical memory constraints of the host. Block data is a form of remote DMA transfer.	Delivery of block data is guaranteed. The sending and receiving ports and the port emulators cooperate in breaking the transfer into data packets and ensuring that all packets are correctly transmitted, received, and placed in the appropriate destination buffer. Block data transfers differ from messages in the size of the transfer.	Disk subsystems and disk servers to move data associated with disk read and write requests. Fast remastering of large lock trees. Transferring large ICC messages.

The SCS layer is implemented as a combination of hardware and software, or software only, depending upon the type of port. SCS manages connections in an OpenVMS Cluster and multiplexes messages between system applications over a common transport called a **virtual circuit**. A virtual circuit exists between each pair of SCS ports and a set of SCS connections that are multiplexed on that virtual circuit.

2.1.3. System Applications (SYSAPs) Layer

The next higher layer in the OpenVMS Cluster architecture consists of the SYSAPs layer. This layer consists of multiple system applications that provide, for example, access to disks and tapes and cluster membership control. SYSAPs can include:

- Connection manager
- MSCP server
- TMSCP server
- Disk and tape class drivers

These components are described in detail later in this chapter.

2.1.4. Other Layered Components

A wide range of OpenVMS components layer on top of the OpenVMS Cluster system architecture, including:

- Volume Shadowing for OpenVMS
- Distributed lock manager
- Process control services
- Distributed file system
- Record Management Services (RMS)
- Distributed job controller

These components, except for volume shadowing, are described in detail later in this chapter. Volume Shadowing for OpenVMS is described in Section 6.6.

2.2. OpenVMS Cluster Software Functions

The OpenVMS Cluster software components that implement OpenVMS Cluster communication and resource-sharing functions always run on every computer in the OpenVMS Cluster. If one computer fails, the OpenVMS Cluster system continues operating, because the components still run on the remaining computers.

2.2.1. Functions

The following table summarizes the OpenVMS Cluster communication and resource-sharing functions and the components that perform them.

Function	Performed By
Ensure that OpenVMS Cluster computers communicate with one another to enforce the rules of cluster membership	Connection manager
Synchronize functions performed by other OpenVMS Cluster components, OpenVMS products, and other software components	Distributed lock manager

Function	Performed By
Share disks and files	Distributed file system
Make disks available to nodes that do not have direct access	MSCP server
Make tapes available to nodes that do not have direct access	TMSCP server
Make queues available	Distributed job controller

2.3. Ensuring the Integrity of Cluster Membership

The connection manager ensures that computers in an OpenVMS Cluster system communicate with one another to enforce the rules of cluster membership.

Computers in an OpenVMS Cluster system share various data and system resources, such as access to disks and files. To achieve the coordination that is necessary to maintain resource integrity, the computers must maintain a clear record of cluster membership.

2.3.1. Connection Manager

The connection manager creates an OpenVMS Cluster when the first computer is booted and reconfigures the cluster when computers join or leave it during **cluster state transitions**. The overall responsibilities of the connection manager are to:

- Prevent partitioning (see Section 2.3.2).
- Track which nodes in the OpenVMS Cluster system are active and which are not.
- Deliver messages to remote nodes.
- Remove nodes.
- Provide a highly available message service in which other software components, such as the distributed lock manager, can synchronize access to shared resources.

2.3.2. Cluster Partitioning

A primary purpose of the connection manager is to prevent **cluster partitioning**, a condition in which nodes in an existing OpenVMS Cluster configuration divide into two or more independent clusters.

Cluster partitioning can result in data file corruption because the distributed lock manager cannot coordinate access to shared resources for multiple OpenVMS Cluster systems. The connection manager prevents cluster partitioning using a quorum algorithm.

2.3.3. Quorum Algorithm

The quorum algorithm is a mathematical method for determining if a majority of OpenVMS Cluster members exist so that resources can be shared across an OpenVMS Cluster system. **Quorum** is the number of votes that must be present for the cluster to function. Quorum is a dynamic value calculated by the connection manager to prevent cluster partitioning. The connection manager allows processing to occur only if a majority of the OpenVMS Cluster members are functioning.

2.3.4. System Parameters

Two system parameters, VOTES and EXPECTED_VOTES, are key to the computations performed by the quorum algorithm. The following table describes these parameters.

Parameter	Description
VOTES	<p>Specifies a fixed number of votes that a computer contributes toward quorum. The system manager can set the VOTES parameters on each computer or allow the operating system to set it to the following default values:</p> <ul style="list-style-type: none"> • For satellite nodes, the default value is 0. • For all other computers, the default value is 1. <p>Each Integrity server or an Alpha computer with a nonzero value for the VOTES system parameter is considered a voting member.</p>
EXPECTED_VOTES	<p>Specifies the sum of all VOTES held by OpenVMS Cluster members. The <i>initial</i> value is used to derive an estimate of the <i>correct</i> quorum value for the cluster. The system manager must set this parameter on each active Integrity server system or an Alpha system, including satellites in the cluster.</p>

2.3.5. Calculating Cluster Votes

The quorum algorithm operates as follows:

Step	Action				
1	<p>When nodes in the OpenVMS Cluster boot, the connection manager uses the largest value for EXPECTED_VOTES of all systems present to derive an estimated quorum value according to the following formula:</p> $\text{Estimated quorum} = (\text{EXPECTED_VOTES} + 2) / 2 \mid \text{Rounded down}$				
2	<p>During a state transition (whenever a node enters or leaves the cluster or when a quorum disk is recognized), the connection manager dynamically computes the cluster quorum value to be the <i>maximum</i> of the following:</p> <ul style="list-style-type: none"> • The current cluster quorum value (calculated during the last cluster transition). • Estimated quorum, as described in step 1. • The value calculated from the following formula, where the VOTES system parameter is the total votes held by all cluster members: $\text{QUORUM} = (\text{VOTES} + 2) / 2 \mid \text{Rounded down}$ <p>Note: Quorum disks are discussed in Section 2.3.8.</p>				
3	<p>The connection manager compares the cluster votes value to the cluster quorum value and determines what action to take based on the following conditions:</p> <table border="1"> <thead> <tr> <th>WHEN...</th> <th>THEN...</th> </tr> </thead> <tbody> <tr> <td>The total number of cluster votes is equal to at least the quorum value</td> <td>The OpenVMS Cluster system continues running.</td> </tr> </tbody> </table>	WHEN...	THEN...	The total number of cluster votes is equal to at least the quorum value	The OpenVMS Cluster system continues running.
WHEN...	THEN...				
The total number of cluster votes is equal to at least the quorum value	The OpenVMS Cluster system continues running.				

Step	Action	
	The current number of cluster votes drops below the quorum value (because of computers leaving the cluster)	The remaining OpenVMS Cluster members suspend all process activity and all I/O operations to cluster-accessible disks and tapes until sufficient votes are added (that is, enough computers have joined the OpenVMS Cluster) to bring the total number of votes to a value greater than or equal to quorum.

Note: When a node leaves the OpenVMS Cluster system, the connection manager does not decrease the cluster quorum value. In fact, the connection manager never decreases the cluster quorum value; the connection manager only increases the value, unless the REMOVE NODE option was selected during shutdown. However, system managers can decrease the value according to the instructions in Section 10.11.2.

2.3.6. Example

Consider a cluster consisting of three computers, each computer having its VOTES parameter set to 1 and its EXPECTED_VOTES parameter set to 3. The connection manager dynamically computes the cluster quorum value to be 2 (that is, $(3 + 2)/2$). In this example, any two of the three computers constitute a quorum and can run in the absence of the third computer. No single computer can constitute a quorum by itself. Therefore, there is no way the three OpenVMS Cluster computers can be partitioned and run as two independent clusters.

2.3.7. Sub-Cluster Selection

To select the optimal sub-cluster and to continue after the communication failure occurs, two possible sub-clusters are compared as follows:

1. The subset with the highest number of votes wins, if one of the subset has more votes.
2. If in case there is a tie in the number of votes:
 - The subset with the higher number of nodes wins.
 - If the number of nodes is also tied, then: OpenVMS arbitrarily, but deterministically selects one of the two subsets to "win" based on a comparison of SCS System ID values.

2.3.8. Quorum Disk

A cluster system manager can designate a disk a **quorum disk**. The quorum disk acts as a *virtual* cluster member whose purpose is to add one vote to the total cluster votes. By establishing a quorum disk, you can increase the availability of a two-node cluster; such configurations can maintain quorum in the event of failure of either the quorum disk or one node, and continue operating.

Note: Setting up a quorum disk is recommended only for OpenVMS Cluster configurations with two nodes. A quorum disk is neither necessary nor recommended for configurations with more than two nodes.

For example, assume an OpenVMS Cluster configuration with many satellites (that have no votes) and two nonsatellite systems (each having one vote) that downline load the satellites. Quorum is calculated as follows:

$$(\text{EXPECTED VOTES} + 2) / 2 = (2 + 2) / 2 = 2$$

Because there is no quorum disk, if either nonsatellite system departs from the cluster, only one vote remains and cluster quorum is lost. Activity will be blocked throughout the cluster until quorum is restored.

However, if the configuration includes a quorum disk (adding one vote to the total cluster votes), and the EXPECTED_VOTES parameter is set to 3 on each node, then quorum will still be 2 even if one of the nodes leaves the cluster. Quorum is calculated as follows:

$$(\text{EXPECTED_VOTES} + 2) / 2 = (3 + 2) / 2 = 2$$

Rules: Each OpenVMS Cluster system can include only one quorum disk. At least one computer must have a direct (not served) connection to the quorum disk:

- Any computers that have a direct, active connection to the quorum disk or that have the potential for a direct connection should be enabled as **quorum disk watchers**.
- Computers that cannot access the disk directly must rely on the quorum disk watchers for information about the status of votes contributed by the quorum disk.

Reference: For more information about enabling a quorum disk, see Section 8.2.4. Section 8.3.2 describes removing a quorum disk.

2.3.9. Quorum Disk Watcher

To enable a computer as a quorum disk watcher, use one of the following methods:

Method	Perform These Steps
Run the CLUSTER_CONFIG.COM procedure (described in Chapter 8)	<p>Invoke the procedure and:</p> <ol style="list-style-type: none"> 1. Select the CHANGE option. 2. From the CHANGE menu, select the item labeled “Enable a quorum disk on the local computer”. 3. At the prompt, supply the quorum disk device name. <p>The procedure uses the information you provide to update the values of the DISK_QUORUM and QDSKVOTES system parameters.</p>
Respond YES when the OpenVMS installation procedure asks whether the cluster will contain a quorum disk (described in Chapter 4)	<p>During the installation procedure:</p> <ol style="list-style-type: none"> 1. Answer Y when the procedure asks whether the cluster will contain a quorum disk. 2. At the prompt, supply the quorum disk device name. <p>The procedure uses the information you provide to update the values of the DISK_QUORUM and QDSKVOTES system parameters.</p>
Edit the MODPARAMS or AGEN\$ files (described in Chapter 8)	<p>Edit the following parameters:</p> <ul style="list-style-type: none"> • DISK_QUORUM: Specify the quorum disk name, in ASCII, as a value for the DISK_QUORUM system parameter. • QDSKVOTES: Set an appropriate value for the QDSKVOTES parameter. This parameter specifies the number of votes

Method	Perform These Steps
	contributed to the cluster votes total by a quorum disk. The number of votes contributed by the quorum disk is equal to the smallest value of the QDSKVOTES parameter on any quorum disk watcher.

Hint: If only one quorum disk watcher has direct access to the quorum disk, then remove the disk and give its votes to the node.

2.3.10. Rules for Specifying Quorum

For the quorum disk's votes to be counted in the total cluster votes, the following conditions must be met:

- On all computers capable of becoming watchers, you must specify the same *physical* device name as a value for the DISK_QUORUM system parameter. The remaining computers (which must have a blank value for DISK_QUORUM) recognize the name specified by the first quorum disk watcher with which they communicate.
- At least one quorum disk watcher must have a direct, active connection to the quorum disk.
- The disk must contain a valid format file named QUORUM.DAT in the master file directory. The QUORUM.DAT file is created automatically after a system specifying a quorum disk has booted into the cluster for the first time. This file is used on subsequent reboots.

Note: The file is not created if the system parameter STARTUP_P1 is set to MIN.

- To permit recovery from failure conditions, the quorum disk must be mounted by all disk watchers.
- The OpenVMS Cluster can include only one quorum disk.
- The quorum disk cannot be a member of a shadow set.

Hint: By increasing the quorum disk's votes to one less than the total votes from both systems (and by increasing the value of the EXPECTED_VOTES system parameter by the same amount), you can boot and run the cluster with only one node.

2.4. State Transitions

OpenVMS Cluster state transitions occur when a computer joins or leaves an OpenVMS Cluster system and when the cluster recognizes a quorum disk state change. The connection manager controls these events to ensure the preservation of data integrity throughout the cluster.

A state transition's duration and effect on users (applications) are determined by the reason for the transition, the configuration, and the applications in use.

2.4.1. Adding a Member

Every transition goes through one or more phases, depending on whether its cause is the addition of a new OpenVMS Cluster member or the failure of a current member.

Table 2.2 describes the phases of a transition caused by the addition of a new member.

Table 2.2. Transitions Caused by Adding a Cluster Member

Phase	Description
New member detection	<p>Early in its boot sequence, a computer seeking membership in an OpenVMS Cluster system sends messages to current members asking to join the cluster. The first cluster member that receives the membership request acts as the new computer's advocate and proposes reconfiguring the cluster to include the computer in the cluster. While the new computer is booting, no applications are affected.</p> <p>Note: The connection manager will not allow a computer to join the OpenVMS Cluster system if the node's value for EXPECTED_VOTES would readjust quorum higher than calculated votes to cause the OpenVMS Cluster to suspend activity.</p>
Reconfiguration	<p>During a configuration change due to a computer being added to an OpenVMS Cluster, all current OpenVMS Cluster members must establish communications with the new computer. Once communications are established, the new computer is admitted to the cluster. In some cases, the lock database is rebuilt.</p>

2.4.2. Losing a Member

Table 2.3 describes the phases of a transition caused by the failure of a current OpenVMS Cluster member.

Table 2.3. Transitions Caused by Loss of a Cluster Member

Cause	Description						
Failure detection	<p>The duration of this phase depends on the cause of the failure and on how the failure is detected.</p> <p>During normal cluster operation, messages sent from one computer to another are acknowledged when received.</p>						
	<table border="1"> <thead> <tr> <th>IF...</th> <th>THEN...</th> </tr> </thead> <tbody> <tr> <td>A message is not acknowledged within a period determined by OpenVMS Cluster communications software</td> <td>The repair attempt phase begins.</td> </tr> <tr> <td>A cluster member is shut down or fails</td> <td>The operating system causes datagrams to be sent from the computer shutting down to the other members. These datagrams state the computer's intention to sever communications and to stop sharing resources. The failure detection and repair attempt phases are bypassed, and the reconfiguration phase begins immediately.</td> </tr> </tbody> </table>	IF...	THEN...	A message is not acknowledged within a period determined by OpenVMS Cluster communications software	The repair attempt phase begins.	A cluster member is shut down or fails	The operating system causes datagrams to be sent from the computer shutting down to the other members. These datagrams state the computer's intention to sever communications and to stop sharing resources. The failure detection and repair attempt phases are bypassed, and the reconfiguration phase begins immediately.
	IF...	THEN...					
A message is not acknowledged within a period determined by OpenVMS Cluster communications software	The repair attempt phase begins.						
A cluster member is shut down or fails	The operating system causes datagrams to be sent from the computer shutting down to the other members. These datagrams state the computer's intention to sever communications and to stop sharing resources. The failure detection and repair attempt phases are bypassed, and the reconfiguration phase begins immediately.						
Repair attempt	<p>If the virtual circuit to an OpenVMS Cluster member is broken, attempts are made to repair the path. Repair attempts continue for an interval specified by the PAPOLLINTERVAL system parameter. (System managers can adjust the value of this parameter to suit local conditions.) Thereafter,</p>						

Cause	Description		
	the path is considered irrevocably broken, and steps must be taken to reconfigure the OpenVMS Cluster system so that all computers can once again communicate with each other and so that computers that cannot communicate are removed from the OpenVMS Cluster.		
Reconfiguration	If a cluster member is shut down or fails, the cluster must be reconfigured. One of the remaining computers acts as coordinator and exchanges messages with all other cluster members to determine an optimal cluster configuration with the most members and the most votes. This phase, during which all user (application) activity is blocked, usually lasts less than 3 seconds, although the actual time depends on the configuration.		
OpenVMS Cluster system recovery	Recovery includes the following stages, some of which can take place in parallel:		
	Stage	Action	
	I/O completion	When a computer is removed from the cluster, OpenVMS Cluster software ensures that all I/O operations that are started prior to the transition complete before I/O operations that are generated after the transition. This stage usually has little or no effect on applications.	
	Recovery includes the following stages, some of which can take place in parallel:	When a computer is removed from the cluster, OpenVMS Cluster software ensures that all I/O operations that are started prior to the transition complete before I/O operations that are generated after the transition. This stage usually has little or no effect on applications.	
		WHEN...	THEN...
		A computer leaves the OpenVMS Cluster	A rebuild is always performed.
		A computer is added to the OpenVMS Cluster	A rebuild is performed when the LOCKDIRWT system parameter is greater than 1.
	Caution: Setting the LOCKDIRWT system parameter to different values on the same model or type of computer can cause the distributed lock manager to use the computer with the higher value. This could cause undue resource usage on that computer.		
Disk mount verification	This stage occurs only when the failure of a voting member causes quorum to be lost. To protect data integrity, all I/O activity is blocked until quorum is regained. Mount verification is the mechanism used to block I/O during this phase.		
Quorum disk votes validation	If, when a computer is removed, the remaining members can determine that it has shut down or failed, the votes contributed by the quorum		

Cause	Description	
		disk are included without delay in quorum calculations that are performed by the remaining members. However, if the quorum watcher cannot determine that the computer has shut down or failed (for example, if a console halt, power failure, or communications failure has occurred), the votes are not included for a period (in seconds) equal to four times the value of the QDSKINTERVAL system parameter. This period is sufficient to determine that the failed computer is no longer using the quorum disk.
	Disk rebuild	If the transition is the result of a computer rebooting after a failure, the disks are marked as improperly dismantled. Reference: See Sections 6.5.5 and 6.5.6 for information about rebuilding disks.
	XFC cache change	If the XFC cache is active on this node, a check is made to determine if there are any nodes in the cluster that do not support the XFC cache. If so, any XFC cache data must be flushed before continuing with the cluster transition.
	Clusterwide logical name recovery	This stage ensures that all nodes in the cluster have matching clusterwide logical name information.
Application recovery	When you assess the effect of a state transition on application users, consider that the application recovery phase includes activities such as replaying a journal file, cleaning up recovery units, and users logging in again.	

2.5. OpenVMS Cluster Membership

OpenVMS Cluster systems based on LAN or IP network use a cluster group number and a cluster password to allow multiple independent OpenVMS Cluster systems to coexist on the same extended LAN or IP network and to prevent accidental access to a cluster by unauthorized computers.

Note

When using IP network for cluster communication, the remote node's IP address must be present in the SYS\$SYSTEM:PE\$IP_CONFIG.DAT local file.

2.5.1. Cluster Group Number

The **cluster group number** uniquely identifies each OpenVMS Cluster system on a LAN or IP or communicates by a common memory region (that is, communicating using SMCI). This group number must be either from 1 to 4095 or from 61440 to 65535.

Rule: If you plan to have more than one OpenVMS Cluster system on a LAN or an IP network, you must coordinate the assignment of cluster group numbers among system managers.

2.5.2. Cluster Password

The **cluster password** prevents an unauthorized computer using the cluster group number, from joining the cluster. The password must be from 1 to 31 characters; valid characters are letters, numbers, the dollar sign (\$), and the underscore (_).

2.5.3. Location

The cluster group number and cluster password are maintained in the cluster authorization file, SYS\$COMMON:[SYSEXEC]CLUSTER_AUTHORIZE.DAT. This file is created during the installation of the operating system, if you indicate that you want to set up a cluster that utilizes the shared memory or the LAN. The installation procedure then prompts you for the cluster group number and password.

Note

If you convert an OpenVMS Cluster that uses only the CI or DSSI interconnect to one that includes a LAN or shared memory interconnect, the SYS\$COMMON:[SYSEXEC]CLUSTER_AUTHORIZE.DAT file is created when you execute the CLUSTER_CONFIG.COM command procedure, as described in Chapter 8.

Reference: For information about OpenVMS Cluster group data in the CLUSTER_AUTHORIZE.DAT file, see Sections 8.4 and 10.8.

2.5.4. Example

If all nodes in the OpenVMS Cluster do not have the same cluster password, an error report similar to the following is logged in the error log file.

```

**** V3.4 ***** ENTRY 343
*****

Logging OS                1. OpenVMS
System Architecture       2. Alpha
OS version                XC56-BL2
Event sequence number    102.
Timestamp of occurrence   16-SEP-2009 16:47:48
Time since reboot        0 Day(s) 1:04:52
Host name                 PERK

System Model              AlphaServer ES45 Model 2

Entry Type                98. Asynchronous Device Attention

---- Device Profile ----
Unit                      PERK$PEA0
Product Name              NI-SCA Port

---- NISCA Port Data ----
Error Type and SubType    x0600 Channel Error, Invalid Cluster
Password
                           Received
Status                    x0000000000000000
Datalink Device Name     EIA8:

```

```
Remote Node Name          CHBOSE
Remote Address            x000064A9000400AA
Local Address             x000063B4000400AA
Error Count               1. Error Occurrences This Entry

----- Software Info -----
UCB$x_ERRCNT              6. Errors This Unit
```

2.6. Synchronizing Cluster Functions by the Distributed Lock Manager

The **distributed lock manager** is an OpenVMS feature for synchronizing functions required by the distributed file system, the distributed job controller, device allocation, user-written OpenVMS Cluster applications, and other OpenVMS products and software components.

The distributed lock manager uses the connection manager and SCS to communicate information between OpenVMS Cluster computers.

2.6.1. Distributed Lock Manager Functions

The functions of the distributed lock manager include the following:

- Synchronizes access to shared clusterwide resources, including:
 - Devices
 - Files
 - Records in files
 - Any user-defined resources, such as databases and memory

Each resource is managed clusterwide by an OpenVMS Cluster computer.

- Implements the \$ENQ and \$DEQ system services to provide clusterwide synchronization of access to resources by allowing the locking and unlocking of resource names.

Reference: For detailed information about system services, refer to the *VSI OpenVMS System Services Reference Manual*.

- Queues process requests for access to a locked resource. This queuing mechanism allows processes to be put into a wait state until a particular resource is available. As a result, cooperating processes can synchronize their access to shared objects, such as files and records.
- Releases all locks that an OpenVMS Cluster computer holds if the computer fails. This mechanism allows processing to continue on the remaining computers.
- Supports clusterwide deadlock detection.

2.6.2. System Management of the Lock Manager

The lock manager is fully automated and usually requires no explicit system management. However, the LOCKDIRWT and LOCKRMWT system parameters can be used to adjust the distribution of activity and control of lock resource trees across the cluster.

A lock resource tree is an abstract entity on which locks can be placed. Multiple lock resource trees can exist within a cluster. For every resource tree, there is one node known as the directory node and another node known as the lock resource master node.

A lock resource master node controls a lock resource tree and is aware of all the locks on the lock resource tree. All locking operations on the lock tree must be sent to the resource master. These locks can come from any node in the cluster. All other nodes in the cluster only know about their specific locks on the tree.

Furthermore, all nodes in the cluster have many locks on many different lock resource trees, which can be mastered on different nodes. When creating a new lock resource tree, the directory node must first be queried if a resource master already exists.

The LOCKDIRWT parameter allocates a node as the directory node for a lock resource tree. The higher a node's LOCKDIRWT setting, the higher the probability that it will be the directory node for a given lock resource tree.

For most configurations, large computers and boot nodes perform optimally when LOCKDIRWT is set to 1 and satellite nodes have LOCKDIRWT set to 0. These values are set automatically by the CLUSTER_CONFIG.COM procedure. Nodes with a LOCKDIRWT of 0 will not be the directory node for any resources unless all nodes in the cluster have a LOCKDIRWT of 0.

In some circumstances, you may want to change the values of the LOCKDIRWT parameter across the cluster to control the extent to which nodes participate as directory nodes.

LOCKRMWT influences which node is chosen to remaster a lock resource tree. Because there is a performance advantage for nodes mastering a lock resource tree (as no communication is required when performing a locking operation), the lock resource manager supports remastering lock trees to other nodes in the cluster. Remastering a lock resource tree means to designate another node in the cluster as the lock resource master for that lock resource tree and to move the lock resource tree to it.

A node is eligible to be a lock resource master node if it has locks on that lock resource tree. The selection of the new lock resource master node from the eligible nodes is based on each node's LOCKRMWT system parameter setting and each node's locking activity.

LOCKRMWT can contain a value between 0 and 10; the default is 5. The following list describes how the value of the LOCKRMWT system parameter affects resource tree mastery and how lock activity can affect the decision:

- Any node that has a LOCKRMWT value of 0 will attempt to remaster a lock tree to another node which has locks on that tree, as long as the other node has a LOCKRMWT greater than 0.
- Nodes with a LOCKRMWT value of 10 will be given resource trees from other nodes that have a LOCKRMWT less than 10.
- Otherwise, the difference in LOCKRMWT is computed between the master and the eligible node. The higher the difference, the more activity is required by the eligible node for the lock tree to move.

In most cases, maintaining the default value of 5 for LOCKRMWT is appropriate, but there may be cases where assigning some nodes a higher or lower LOCKRMWT is useful for determining which nodes master a lock tree. The LOCKRMWT parameter is dynamic, hence it can be adjusted, if necessary.

2.6.3. Large-Scale Locking Applications

The Enqueue process limit (ENQLM), which is set in the SYSUAF.DAT file and which controls the number of locks that a process can own, can be adjusted to meet the demands of large scale databases and other server applications.

Prior to OpenVMS Version 7.1, the limit was 32767. This limit was removed to enable the efficient operation of large scale databases and other server applications. A process can now own up to 16,776,959 locks, the architectural maximum. By setting ENQLM in SYSUAF.DAT to 32767 (using the Authorize utility), the lock limit is automatically extended to the maximum of 16,776,959 locks. \$CREPRC can pass large quotas to the target process if it is initialized from a process with the SYSUAF Enqlm quota of 32767.

Reference: See the *VSI OpenVMS Programming Concepts Manual* for additional information about the distributed lock manager and resource trees. See the *VSI OpenVMS System Manager's Manual* for more information about Enqueue Quota.

2.7. Resource Sharing

Resource sharing in an OpenVMS Cluster system is enabled by the distributed file system, RMS, and the distributed lock manager.

2.7.1. Distributed File System

The OpenVMS Cluster **distributed file system** allows all computers to share mass storage and files. The distributed file system provides the same access to disks, tapes, and files across the OpenVMS Cluster that is provided on a standalone computer.

2.7.2. RMS and the Distributed Lock Manager

The distributed file system and OpenVMS Record Management Services (RMS) use the distributed lock manager to coordinate clusterwide file access. RMS files can be shared to the record level.

Almost any disk or tape device can be made available to the entire OpenVMS Cluster system. The devices can be:

- Connected to a supported storage subsystem
- A local device that is served to the OpenVMS Cluster

All cluster-accessible devices appear as if they are connected to every computer.

2.8. Disk Availability

Locally connected disks can be served across an OpenVMS Cluster by the MSCP server.

2.8.1. MSCP Server

The **MSCP server** makes locally connected disks, including the following, available across the cluster:

- DSA disks local to OpenVMS Cluster members using SDI

- HSG and HSV disks in an OpenVMS Cluster using mixed interconnects
- SCSI and HSZ disks
- SAS, LSI 1068 SAS and LSI Logic 1068e SAS disks
- FC and HSG disks
- Disks on boot servers and disk servers located anywhere in the OpenVMS Cluster

In conjunction with the disk class driver (DUDRIVER), the MSCP server implements the storage server portion of the MSCP protocol on a computer, allowing the computer to function as a storage controller. The MSCP protocol defines conventions for the format and timing of messages sent and received for certain families of mass storage controllers and devices designed by VSI. The MSCP server decodes and services MSCP I/O requests sent by remote cluster nodes.

Note: The MSCP server is not used by a computer to access files on locally connected disks.

2.8.2. Device Servicing

Once a device is set up to be served:

- Any cluster member can submit I/O requests to it.
- The local computer can decode and service MSCP I/O requests sent by remote OpenVMS Cluster computers.

2.8.3. Enabling the MSCP Server

The MSCP server is controlled by the MSCP_LOAD and MSCP_SERVE_ALL system parameters. The values of these parameters are set initially by answers to questions asked during the OpenVMS installation procedure (described in Section 8.4), or during the CLUSTER_CONFIG.COM procedure (described in Chapter 8).

The default values for these parameters are as follows:

- MSCP is not loaded on satellites.
- MSCP is loaded on boot server and disk server nodes.

Reference: See Section 6.3 for more information about setting system parameters for MSCP servicing.

2.9. Tape Availability

Locally connected tapes can be served across an OpenVMS Cluster by the TMSCP server.

2.9.1. TMSCP Server

The **TMSCP server** makes locally connected tapes, available across the cluster including the following:

- HSG and HSV tapes
- SCSI tapes

- SAS tapes

The TMSCP server implements the TMSCP protocol, which is used to communicate with a controller for TMSCP tapes. In conjunction with the tape class driver (TUDRIVER), the TMSCP protocol is implemented on a processor, allowing the processor to function as a storage controller.

The processor submits I/O requests to locally accessed tapes, and accepts the I/O requests from any node in the cluster. In this way, the TMSCP server makes locally connected tapes available to all nodes in the cluster. The TMSCP server can also make HSG and HSV tapes accessible to OpenVMS Cluster satellites.

2.9.2. Enabling the TMSCP Server

The TMSCP server is controlled by the TMSCP_LOAD system parameter. The value of this parameter is set initially by answers to questions asked during the OpenVMS installation procedure (described in Section 4.2.3) or during the CLUSTER_CONFIG.COM procedure (described in Section 8.4). By default, the setting of the TMSCP_LOAD parameter does not load the TMSCP server and does not serve any tapes.

2.10. Queue Availability

The **distributed queue manager** makes queues available across the cluster to achieve the following:

Function	Description
Permit users on any OpenVMS Cluster computer to submit batch and print jobs to queues that execute on any computer in the OpenVMS Cluster	Users can submit jobs to any queue in the cluster, provided that the necessary mass storage volumes and peripheral devices are accessible to the computer on which the job executes.
Distribute the batch and print processing work load over OpenVMS Cluster nodes	System managers can set up generic batch and print queues that distribute processing work loads among computers. The distributed queue manager directs batch and print jobs either to the execution queue with the lowest ratio of jobs-to-queue limit or to the next available printer.

The distributed queue manager uses the distributed lock manager to signal other computers in the OpenVMS Cluster to examine the batch and print queue jobs to be processed.

2.10.1. Controlling Queues

To control queues, you use one or several queue managers to maintain a clusterwide queue database that stores information about queues and jobs.

Reference: For detailed information about setting up OpenVMS Cluster queues, see Chapter 7.

Chapter 3. OpenVMS Cluster Interconnect Configurations

This chapter provides an overview of various types of OpenVMS Cluster configurations and the ways they are interconnected.

For definitive information about supported OpenVMS Cluster configurations, see:

- *VSI OpenVMS Cluster Software Product Description*
- *Guidelines for OpenVMS Cluster Configurations*

3.1. Overview

Every node in an OpenVMS Cluster must have direct connections to all other nodes. Sites can choose to use one or more of the following interconnects:

- LANs
 - Ethernet (Fast Ethernet, Gigabit Ethernet, 10 Gigabit Ethernet)
- Internet Protocol (IP)
 - Ethernet (Fast Ethernet, Gigabit Ethernet, 10 Gigabit Ethernet)
- MEMORY CHANNEL (Alpha only)
- SMCI (Shared memory CI) (Alpha only) in OpenVMS Galaxy configurations, as described in the *VSI OpenVMS Alpha Partitioning and Galaxy Guide*
- SCSI (supported only as a node-to-storage interconnect, requires a second interconnect for node-to-node (SCS) communications for limited configurations)
- Fibre Channel (supported only as a node-to-storage interconnect, requires a second interconnect for node-to-node (SCS) communications)
- SAS (supported only as a node-to-storage interconnect, requires a second interconnect for node-to-node (SCS) communications for limited configurations) (Integrity servers only)

Processing needs and available hardware resources determine how individual OpenVMS Cluster systems are configured. The configuration discussions in this chapter are based on these physical interconnects.

You can use bridges or switches to connect the OpenVMS Integrity server nodes Fast Ethernet/Gigabit Ethernet NIC(s) to any intersite interconnect the WAN supplier provides, such as [D]WDM, Gigabit Ethernet, Fibre Channel or others.

Note

Multihost shared storage on a SCSI interconnect, commonly known as SCSI clusters, is not supported. It is also not supported on OpenVMS Alpha systems for newer SCSI adapters. However, multihost shared storage on industry-standard Fibre Channel is supported.

Locally attached storage, on both OpenVMS Alpha systems (FC or SCSI storage) and OpenVMS Integrity server systems (Fibre Channel, SAS, or SCSI storage), can be served to any other member of the cluster.

3.2. OpenVMS Cluster Systems Interconnected by LANs

All Ethernet interconnects are industry-standard local area networks that are generally shared by a wide variety of network consumers. When OpenVMS Cluster systems are based on LAN, cluster communications are carried out by a port driver (PEDRIVER) that emulates CI port functions.

3.2.1. Design

The OpenVMS Cluster software is designed to use the Ethernet and interconnects simultaneously with the DECnet, TCP/IP, and SCSI protocols. This is accomplished by allowing LAN data link software to control the hardware port. This software provides a multiplexing function so that the cluster protocols are simply another user of a shared hardware resource. See Figure 2.1 for an illustration of this concept.

3.2.1.1. PEDRIVER Fast Path Support

PEdriver, the software that enables OpenVMS Cluster communications over a LAN, also provides Fast Path support. This PEdriver feature provides the following benefits:

- Improves SMP performance scalability.
- Reduces the contention for the SCS/IOLOCK8 spinlock. PEdriver uses a private port mainline spinlock to synchronize its internal operation.
- Allows PEdriver to perform cluster communications processing on a secondary CPU, thus offloading the primary CPU.
- Allows PEdriver to process cluster communications using a single CPU.
- Reduces CPU cost by providing a Fast Path streamlined code path for DSA and served blocked data operations.

For more detailed information, see the *VSI OpenVMS I/O User's Reference Manual*, the *VSI OpenVMS System Manager's Manual*, and the *VSI OpenVMS System Management Utilities Reference Manual*.

3.2.2. Cluster Group Numbers and Cluster Passwords

A single LAN can support multiple LAN-based OpenVMS Cluster systems. Each OpenVMS Cluster is identified and secured by a unique cluster group number and a cluster password. Chapter 2 describes cluster group numbers and cluster passwords in detail.

3.2.3. Servers

OpenVMS Cluster computers interconnected by a LAN are generally configured as either servers or satellites. The following table describes servers.

Server Type	Description
MOP servers	Downline load the OpenVMS boot driver to satellites by means of the Maintenance Operations Protocol (MOP).
Disk servers	Use MSCP server software to make their locally connected disks available to satellites over the LAN.
Tape servers	Use TMSCP server software to make their locally connected tapes available to satellite nodes over the LAN.
Boot servers	A combination of a MOP server and a disk server that serves one or more Alpha system disks. Boot and disk servers make user and application data disks available across the cluster. These servers must be the most powerful computers in the OpenVMS Cluster and must use the highest-bandwidth LAN adapters in the cluster. Boot servers must always run the MSCP server software.

3.2.4. Satellites

Satellites are computers without a local system disk. Generally, satellites are consumers of cluster resources, although they can also provide facilities for disk serving, tape serving, and batch processing. If satellites are equipped with local disks, they can enhance performance by using such local disks for paging and swapping.

Satellites are booted remotely from a boot server (or from a MOP server and a disk server) serving the system disk. Section 3.2.5 describes MOP and disk server functions during satellite booting.

3.2.5. Satellite Booting (Alpha)

When a satellite requests an operating system load, a MOP server for the appropriate OpenVMS Alpha operating system sends a bootstrap image to the satellite that allows the satellite to load the rest of the operating system from a disk server and join the cluster. The sequence of actions during booting is described in Table 3.1.

Table 3.1. Satellite Booting Process

Step	Action	Comments
1	Satellite requests MOP service.	This is the original boot request that a satellite sends out across the network. Any node in the OpenVMS Cluster that has MOP service enabled and has the LAN address of the particular satellite node in its database can become the MOP server for the satellite.
2	MOP server loads the Alpha system.	The MOP server responds to an Alpha satellite boot request by downline loading the SYS\$SYSTEM:APB.EXE program along with the required parameters. For Alpha computers, Some of these parameters include: <ul style="list-style-type: none"> • System disk name • Root number of the satellite
3	Satellite finds additional parameters located on	OpenVMS Cluster system parameters, such as SCSSYSTEMID, SCSNODE, an NISCS_CONV_BOOT. The satellite also finds the cluster group code and password.

Step	Action	Comments
	the system disk and root.	
4	Satellite executes the load program	The program establishes an SCS connection to a disk server for the satellite system disk and loads the SYSBOOT.EXE program.

Configuring and starting a satellite booting service for Alpha computers is described in detail in Section 4.5.

3.2.6. Satellite Booting (Integrity servers)

Configuring and starting a satellite booting service for Integrity server systems is described in detail in Section 4.5.

3.2.7. Configuring Multiple LAN Adapters

LAN support for multiple adapters allows PEDRIVER (the port driver for the LAN) to establish more than one channel between the local and remote cluster nodes. A **channel** is a network path between two nodes that is represented by a pair of LAN adapters.

3.2.7.1. System Characteristics

OpenVMS Cluster systems with multiple LAN adapters have the following characteristics:

- At boot time, *all* Ethernet adapters are automatically configured for local area OpenVMS Cluster use.
- PEDRIVER automatically detects and creates a new channel between the local node and each remote cluster node for each unique pair of LAN adapters.
- Channel viability is monitored continuously.
- In many cases, channel failure does not interfere with node-to-node (virtual circuit)communications as long as there is at least one remaining functioning channel between the nodes.

3.2.7.2. System Requirements

Configurations for OpenVMS Cluster systems with multiple LAN adapters must meet the following requirements:

- The MOP server and the system disk server for a given satellite must be connected to the same extended LAN segment. (LANs can be extended using bridges that manage traffic between two or more local LANs).
- All nodes must have a direct path to all other nodes. A direct path can be a bridged or a nonbridged LAN segment.

Rule: For each node, DECnet for OpenVMS (Phase IV) and MOP serving (Alpha or VAX, as appropriate) can be performed by only one adapter per extended LAN to prevent LAN address duplication.

3.2.7.3. Guidelines

The following guidelines are for configuring OpenVMS Cluster systems with multiple LAN adapters. If you configure these systems according to the guidelines, server nodes (nodes serving disks, tape, and lock traffic) can typically use some of the additional bandwidth provided by the added LAN adapters and increase the overall performance of the cluster. However, the performance increase depends on the configuration of your cluster and the applications it supports.

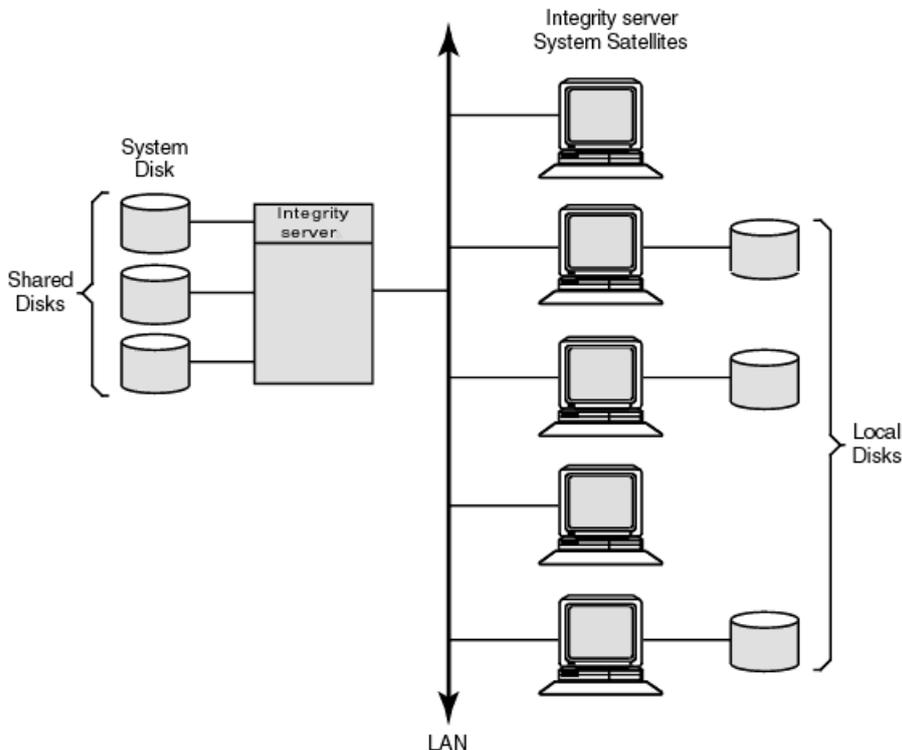
Configurations with multiple LAN adapters should follow these guidelines:

- Connect each LAN adapter to a separate LAN segment. A LAN segment can be bridged or nonbridged. Doing this can help provide higher performance and availability in the cluster. The LAN segments can be Ethernet segments.
- Distribute satellites equally among the LAN segments. Doing this can help to distribute the cluster load more equally across all of the LAN segments.
- Systems providing MOP service should be distributed among the LAN segments to ensure that LAN failures do not prevent satellite booting. Systems should be bridged to multiple LAN segments for performance and availability.
- For the number of LAN adapters supported per node, refer to the OpenVMS Cluster Software SPD.

3.2.8. LAN Examples

Figure 3.1 shows an OpenVMS Cluster system based on a LAN interconnect with a single Alpha server node and a single Alpha system disk.

Figure 3.1. LAN OpenVMS Cluster System with Single Server Node and System Disk



VM-0666A-AI

In Figure 3.1, the server node (and its system disk) is a single point of failure. If the server node fails, the satellite nodes cannot access any of the shared disks including the system disk. Note that some of the satellite nodes have locally connected disks. If you convert one or more of these into system disks, satellite nodes can boot from their own local system disk.

3.2.9. Fast Path for LAN Devices

With OpenVMS Version 7.3-2, further enhancements have been made to Fast Path for LAN devices, which will continue to help streamline I/O processing and improve symmetric-multiprocessing (SMP) performance scalability on newer Alpha Server systems. Enhancements include:

- Reduced contention for the SCS/IOLOCK8 spinlock. The LAN drivers now synchronize using a LAN port-specific spinlock where possible.
- Offload of the primary CPU. The LAN drivers may be assigned to a secondary CPU so that I/O processing can be initiated and completed on the secondary CPU. This offloads the primary CPU and reduces cache contention between processors.

These features enhance the Fast Path functionality that already exist in LAN drivers. The enhanced functionality includes additional optimizations, preallocating of resources, and providing an optimized code path for mainline code.

For more information, see the *VSI OpenVMS I/O User's Reference Manual*

3.2.10. LAN Bridge Failover Process

The following table describes how the bridge parameter settings can affect the failover process.

Option	Comments
Decreasing the LISTEN_TIME value allows the bridge to detect topology changes more quickly.	If you reduce the LISTEN_TIME parameter value, you should also decrease the value for the HELLO_INTERVAL bridge parameter according to the bridge-specific guidelines. However, note that decreasing the value for the HELLO_INTERVAL parameter causes an increase in network traffic.
Decreasing the FORWARDING_DELAY value can cause the bridge to forward packets unnecessarily to the other LAN segment.	Unnecessary forwarding can temporarily cause more traffic on both LAN segments until the bridge software determines which LAN address is on each side of the bridge.

Note

If you change a parameter on one LAN bridge, you should change that parameter on all bridges to ensure that selection of a new root bridge does not change the value of the parameter. The actual parameter value the bridge uses is the value specified by the root bridge.

3.2.11. Virtual LAN Support in OpenVMS

Virtual LAN (VLAN) is a mechanism for segmenting a LAN broadcast domain into smaller sections. The IEEE 802.1Q specification defines the operation and behavior of a VLAN. The OpenVMS implementation adds IEEE 802.1Q support to selected OpenVMS LAN drivers so that OpenVMS can now route VLAN tagged packets to LAN applications using a single LAN adapter.

You can use VLAN to do the following:

- Segment specific LAN traffic on a network for the purposes of network security or traffic containment, or both.
- Use VLAN isolated networks to simplify address management.

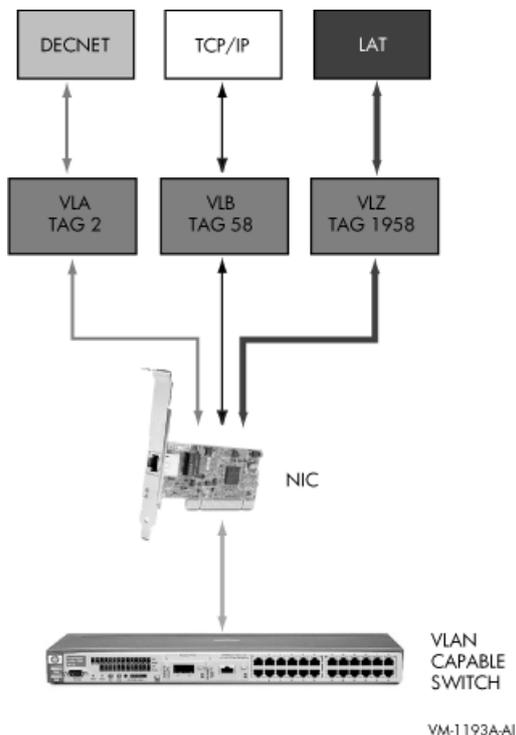
3.2.11.1. VLAN Design

In OpenVMS, VLAN presents a virtual LAN device to LAN applications. The virtual LAN device associates a single IEEE 802.1Q tag with communications over a physical LAN device. The virtual device provides the ability to run any LAN application (for example, SCA, DECnet, TCP/IP, or LAT) over a physical LAN device, allowing host-to-host communications as shown in Figure 3.2.

Note

DECnet-Plus and DECnet Phase IV can be configured to run over a VLAN device.

Figure 3.2. Virtual LAN



OpenVMS VLAN has been implemented through a new driver, `SY$VLANDRIVER.EXE`, which provides the virtual LAN devices. Also, existing LAN drivers have been updated to handle VLAN tags. `LANCP.EXE` and `LANACP.EXE` have been updated with the ability to create and deactivate VLAN devices and to display status and configuration information.

The OpenVMS VLAN subsystem was designed with particular attention to performance. Thus, the performance cost of using VLAN support is negligible.

When configuring VLAN devices, remember that VLAN devices share the same locking mechanism as the physical LAN device. For example, running OpenVMS cluster protocol on a VLAN device along with the underlying physical LAN device does not result in increased benefit and might, in fact, hinder performance.

3.2.11.2. VLAN Support Details

All supported Gigabit and 10-Gb (Integrity servers-only) LAN devices are capable of handling VLAN traffic on Alpha and Integrity server systems.

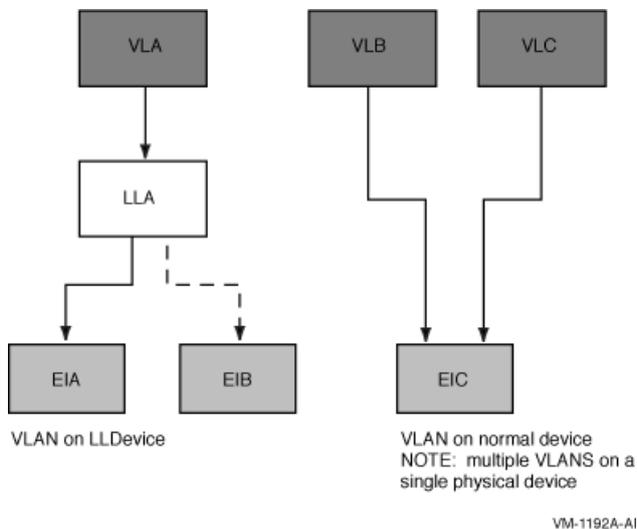
The following list describes additional details of VLAN-related support:

- Switch support

For VLAN configuration, the only requirement of a switch is conformance to the IEEE 802.1Q specification. The VLAN user interface to the switch is not standard; therefore, you must pay special attention when you configure a switch and especially when you configure VLANs across different switches.

- LAN Failover support Figure 3.3 illustrates LAN Failover support.

Figure 3.3. LAN Failover Support



You can create VLAN devices using a LAN Failover set as a source if all members of the set are VLAN-capable devices. However, you cannot build a Failover set using VLAN devices.

- Supported capabilities

VLAN devices inherit the capability of the underlying physical LAN device, including fast path, auto-negotiation, and jumbo frame setting. If a capability needs to be modified, you must modify the underlying physical LAN device.

- Restrictions

No support exists for satellite booting over a VLAN device. The OpenVMS LAN boot drivers do not include VLAN support; therefore, you cannot use a VLAN device to boot an OpenVMS system. Currently, no support exists in OpenVMS for automatic configuration of VLAN devices. You must create VLAN devices explicitly using LANCP commands.

3.3. Cluster over IP

OpenVMS Version 8.4 has been enhanced with the Cluster over IP (Internet Protocol) feature. Cluster over IP provides the ability to form clusters beyond a single LAN or VLAN segment using industry standard Internet Protocol.

System managers also have the ability to manage or monitor OpenVMS cluster that uses IP for cluster communication using SCACP management utility.

Cluster protocol (SCS also known as SCA) over LAN is provided by Port Emulator driver (PEDRIVER). PEDRIVER uses User Datagram Protocol (UDP) and IP in addition to directly using 802.3 interfacing with LAN for cluster communication as shown in Figure 1-0. The datagram characteristics of UDP combined with PEDRIVER's inbuilt reliable delivery mechanism is used for transporting cluster messages which is used by SYSAP (system level application) to communicate between two cluster nodes.

Cluster over IP is an optional feature that can be enabled in addition to the traditional LAN based communication. However, if both LAN and IP mode of communication exist between nodes in a cluster, PEDRIVER prefers LAN communication instead of IP.

Note

OpenVMS Cluster over IP and IP Cluster Interconnect (IPCI) terms are interchangeably used in the document and refers to using TCP/IP stack for cluster communication.

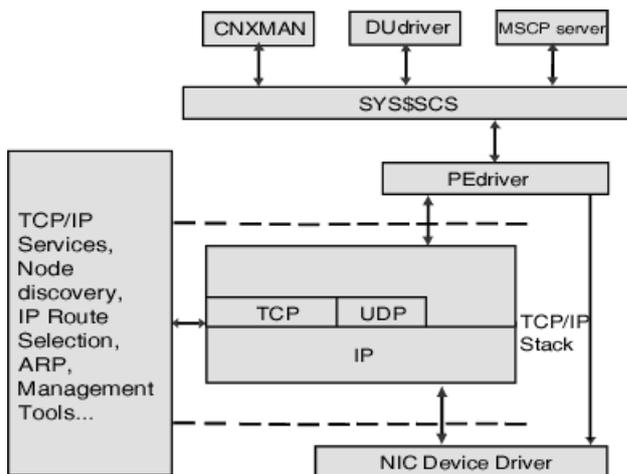
3.3.1. Design

Cluster over IP solution is an integration of the following:

- PEDRIVER support for UDP protocol
- TCP/IP Services boot time loading and initialization

Figure 3.4 shows the cluster over IP architecture.

Figure 3.4. Cluster Communication Design Using IP



3.3.1.1. PEDRIVER Support for UDP

This consists of enhancing PEdriver to use the IP UDP protocol. Some of the features of this solution include:

- The IP UDP service has the same packet delivery characteristics as 802 LANs. PEDRIVER implements the transport layer of NISCA which has inbuilt delay probing, reliable delivery for

sequenced messages (retransmission), implement datagram service and also variable buffer size for block transfers for I/O suitable for cluster traffic.

- The kernel VCI (KVCI) is a kernel mode. It acts as a highly efficient interface to the VSI OpenVMS TCP/IP Services stack. It is a variant of the VCI interface, which PEDriver uses to communicate with OpenVMS LAN drivers. PEDRIVER interfaces to UDP similar to a LAN device.
- Only the lowest layer of PEDRIVER is extended to support UDP. The PEDRIVER changes are transparent to PEDRIVER's upper layers.
- Providing management interface ability to control and configure IP interfaces to PEDRIVER.

3.3.1.2. TCP/IP Services Boot Time Loading and Initialization

To ensure that cluster communication is available in an IP only network environment, it is essential to have TCP/IP stack loaded when the cluster formation starts. This also retains the existing functionality of cluster formation of OpenVMS clusters. Normal booting sequence includes loading of LAN drivers followed by PEDRIVER. TCP/IP drivers are loaded when TCP/IP services are started. If cluster over IP is enabled, LAN, TCP/IP exelets, and PEDRIVER are loaded sequentially. Once the system comes up, TCP/IP services can be started to use other TCP/IP components, such as TELNET, FTP and so on.

Note

Ensure that the TCP/IP software is configured before configuring cluster over IP. To ensure that network and TCP/IP is configured properly, use the PING utility and ping the node from outside the subnet.

3.3.2. Availability

The ability to create a logical LAN failover set using IP for cluster communication provides high availability systems. The nodes will be able to resume if a local LAN card fails, as it will switchover to another interface configured in the logical LAN failover set. For a complete description of creating a logical LAN failover set, see *Guidelines for OpenVMS Cluster Configurations*. The hardware dependency on the LAN bridge is also overcome by GbE switches or routers used for transmission and forwarding the information.

3.3.3. System Characteristics

The existing functionalities of OpenVMS Clusters continue to exist with IP interconnect. Cluster over IP has the following characteristics:

- Cluster over IP does not require any new hardware to use TCP/IP stack as interconnect.
- UDP protocol is used for cluster communication.
- The PEDRIVER includes delay probing technique that helps reduce latency in the IP network by selecting a path with the least latency.
- The OpenVMS Cluster feature of rolling upgrades to the new version without a cluster reboot is retained.

- Provides interoperability with servers running earlier versions of OpenVMS Clusters that are LAN based. Cluster over IP is available only with OpenVMS Version 8.4. Hence, if the node requires IP interconnect to be a part of the cluster, then all the nodes of the cluster must be running OpenVMS Version 8.4 and TCP/IP Services for OpenVMS, Version 5.7.
- At the boot time, LAN, TCP/IP, and PEDRIVER are started sequentially.
- PEDRIVER automatically detects and creates an IP channel for communication between two nodes.
- Cluster over IP feature can be optionally enabled by running the CLUSTER_CONFIG_LAN.COM.
- IP address used for cluster communication must be primary static address of the interface.

3.3.4. Software Requirements

The following software is required to support Clusters over IP interconnect:

- OpenVMS Version 8.4 for Integrity servers or OpenVMS Alpha Version 8.4
- TCP/IP Services for OpenVMS Version 5.7

Note

Ensure that the TCP/IP software is configured before configuring Cluster over IP. To ensure that network and TCP/IP is configured properly, use the PING utility and ping the node from outside the subnet.

3.3.5. Configuration Overview

IP Multicast Address

PEDRIVER uses 802 multicast for discovering cluster members in a LAN. IP multicast maps 1:1 onto the existing LAN discovery, and hence, has been selected as the preferred mechanism to discover nodes in a cluster. Every cluster using IP multicast will have one IP multicast address unique for that cluster. Multicast address is also used for keep-alive mechanism. Administratively scoped IP multicast address is used for cluster communication.

IP Unicast Address

Unicast address can be used if IP multicast is not enabled in a network. Remote node IP address must be present in the local node configuration files to allow the remote node to join the cluster. As a best practice, include all IP addresses and maintain one copy of the file throughout the cluster. \$MC SCACP RELOAD can be used to refresh IP unicast list on a live system.

NISCS_USE_UDP SYSGEN Parameter

This parameter is set to enable the Cluster over IP functionality. PEDRIVER will use the UDP protocol in addition to IEEE 802.3 for cluster communication. CLUSTER_CONFIG_LAN is used to enable cluster over IP which will set this SYSGEN parameter.

UDP Port Number

UDP port number can be configured using `CLUSTER_CONFIG_LAN` and is constant in all nodes of a cluster.

Note

Standard internet practice such as firewall could be applied based on the port number that is selected for cluster.

3.3.5.1. Configuration Files

`SYSSYSTEM:PE$IP_CONFIG.DAT` and `SYSSYSTEM:TCPIP$CLUSTER.DAT` are the two configuration files. These files are loaded during the boot process and provide the necessary configuration details for Cluster over IP. Both these files are generated when a node is configured to be a member of the cluster and if cluster over IP is enabled during the configuration.

`SYSSYSTEM:PE$IP_CONFIG.DAT` includes the optional IP multicast and IP unicast addresses of the nodes of the cluster. IP multicast messages are used for discovering a node within the same IP multicast domain. Remote nodes in a different IP multicast domain can use the IP unicast messaging technique to join the cluster. `SYSSYSTEM:PE$IP_CONFIG.DAT` can be common for all the nodes of a cluster.

`SYSSYSTEM:TCPIP$CLUSTER.DAT` contains the IP interface name and IP addresses on which cluster communication is enabled. It also includes the TCP/IP route information. `SYSSYSTEM:TCPIP$CLUSTER.DAT` is unique for each node in a cluster.

3.3.6. Satellite Node Support

Integrity server satellite node support

The Integrity server satellite node must be in the same LAN on which the boot server resides. The Alpha satellite node must be in the same LAN as its disk server.

Alpha satellite node support

The Alpha console uses the MOP protocol for network load of satellite systems. Because the MOP protocol is non-routable, the satellite boot server or servers and all satellites booting from them must reside in the same LAN. In addition, the boot server must have at least one LAN device enabled for cluster communications to permit the Alpha satellite nodes to access the system disk.

3.3.7. High Availability Configuration using Logical LAN

The ability to create a logical LAN failover set and using IP for cluster communication with the logical LAN failover set provides high availability and can withstand NIC failure to provide high availability configuration. The nodes will be able to continue to communicate even if a local LAN card fails, as it will switchover to another interface configured in the logical LAN failover set. For a complete description of creating a logical LAN failover set and using it for Cluster over IP, see Guidelines for OpenVMS Cluster Configurations. For an example on how to create and configure a Logical LAN failover, refer to Scenario 5: Configuring an Integrity server Node Using a Logical LAN Failover set.

3.3.8. Performance Guidelines

The TCP/IP stack overhead is considered to be in μ s because of the additional layer used for cluster communication. As distance increases this overhead becomes negligible compared to the latency of speed of light. Multi site cluster can leverage from Cluster over IP feature. FASTPATH CPU configuration is recommended for better performance. LAN, TCP/IP and PE device must be on a single CPU. Ensure that there is headroom in the CPU and the CPU is not saturated.

Note

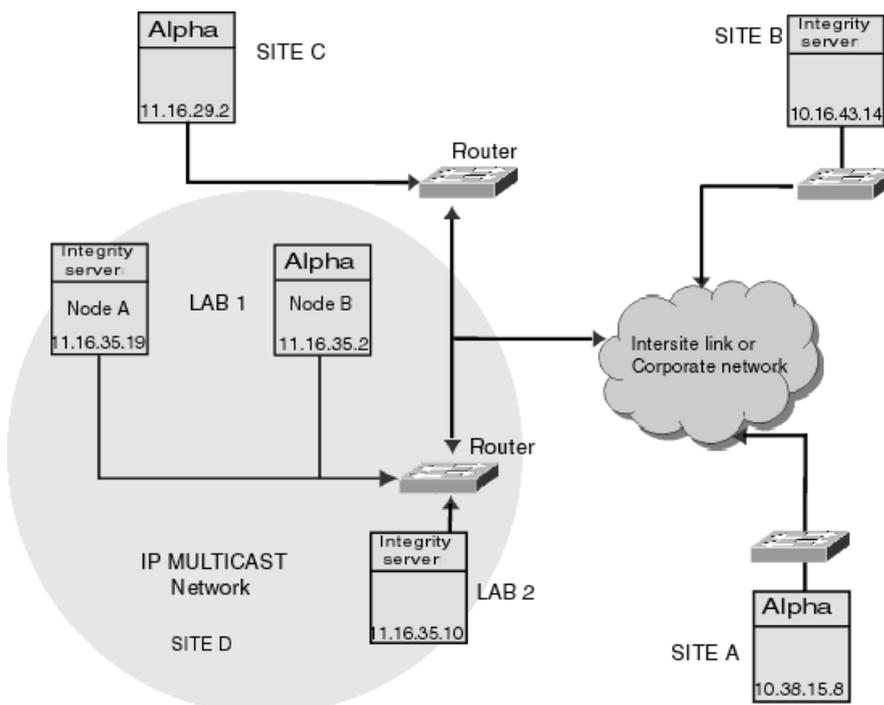
Fast path configuration is not applicable for BG devices when (Packet Processing Engine) PPE is enabled. BG device always takes the primary CPU when cluster over IP is configured and if TCP/IP stack is loaded. It is required to move the BG device to an appropriate CPU using the \$SET DEVICE/PREFERRED command.

3.3.9. Example

Figure 3.5 illustrates an OpenVMS Cluster system based on IP as interconnect. Cluster over IP enables you to connect nodes that are located across various geographical locations. IP multicast is used to locate nodes in the same domain and IP unicast is used to locate nodes in different sites or domains. Cluster over IP supports mixed-architecture, that is, a combination of Integrity server systems and Alpha systems. Lab A and Lab B have the same IP multicast address, and are connected using different LANs.

Node A and Node B are located in the same LAN and use LAN for cluster communication. However, these nodes use IP for cluster communication with all other nodes that are geographically distributed in different sites.

Figure 3.5. OpenVMS Cluster Configuration Based on IP



3.4. OpenVMS Cluster Systems Interconnected by MEMORYCHANNEL (Alpha Only)

MEMORY CHANNEL is a high-performance cluster interconnect technology for PCI-based Alpha systems. With the benefits of very low latency, high bandwidth, and direct memory access, MEMORY CHANNEL complements and extends the ability of OpenVMS Clusters to work as a single virtual system. MEMORYCHANNEL is used for node-to-node cluster communications only. You use it in combination with another interconnect, such as Fibre Channel, SCSI, CI, or DSSI, that is dedicated to storage traffic.

3.4.1. Design

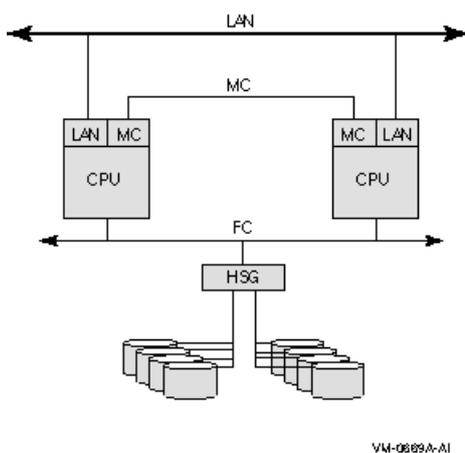
A node requires the following three hardware components to support a MEMORYCHANNEL connection:

- PCI-to MEMORY CHANNEL adapter
- Link cable (3 m or 10 feet long)
- Port in a MEMORY CHANNEL hub (except for a two-node configuration in which the cable connects just two PCI adapters)

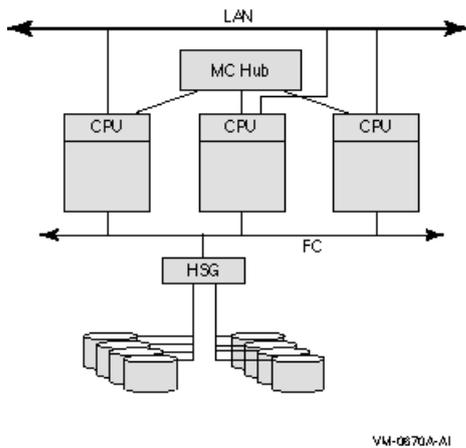
3.4.2. Examples

Figure 3.6 shows a two-node MEMORY CHANNEL cluster with shared access to Fibre Channel storage and a LAN interconnect for failover.

Figure 3.6. Two-Node MEMORY CHANNEL OpenVMS Cluster Configuration



A three-node MEMORY CHANNEL cluster connected by a MEMORY CHANNEL hub and also by a LAN interconnect is shown in Figure 3.7. The three nodes share access to the Fibre Channel storage. The LAN interconnect enables failover if the MEMORY CHANNEL interconnect fails.

Figure 3.7. Three-Node MEMORY CHANNEL OpenVMS Cluster Configuration

3.5. Mixed-Interconnect OpenVMS Cluster Systems

A mixed-interconnect OpenVMS Cluster system is any OpenVMS Cluster system that uses more than one interconnect for SCS communication. You can use mixed interconnects to combine the advantages of each type and to expand your OpenVMS Cluster system. For example, an Ethernet cluster that requires more storage can expand with the addition of Fibre Channel, SCSI, or SAS connections.

Note

If any one node in a cluster requires IP for cluster communication, all the other members in the cluster must be enabled for IP cluster communication.

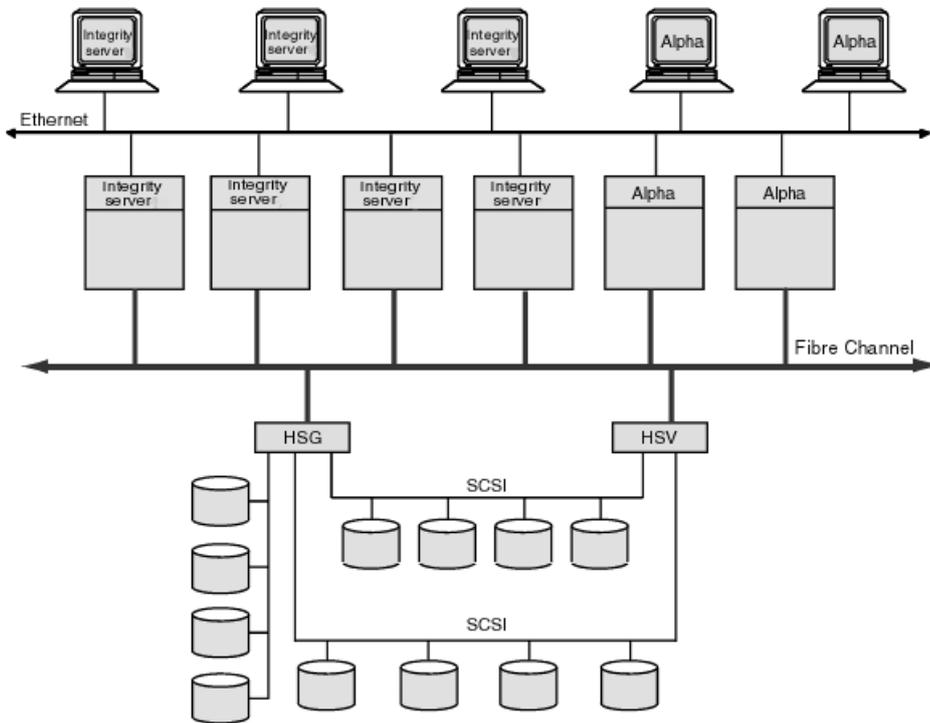
3.5.1. Availability

OpenVMS Cluster systems using a mix of interconnects provide maximum flexibility in combining CPUs, storage, and workstations into highly available configurations.

3.5.2. Examples

Figure 3.8 shows a mixed-interconnect OpenVMS Cluster system using both FC and Ethernet interconnects.

The computers based on the FC can serve HSG or HSV disks to the satellite nodes by means of MSCP server software and drivers; therefore, satellites can access the large amount of storage that is available through HSG and HSV subsystems.

Figure 3.8. OpenVMS Cluster System Using FC and Ethernet Interconnects

ZK5946AGE

3.6. Multihost SCSI OpenVMS Cluster Systems

OpenVMS Cluster systems support the SCSI as a storage interconnect. A SCSI interconnect, also called a SCSI bus, is an industry-standard interconnect that supports one or more computers, peripheral devices, and interconnecting components.

Beginning with OpenVMS Alpha Version 6.2, multiple Alpha computers using the KZPBA SCSI host-based adapter, can simultaneously access SCSI disks over a SCSI interconnect. Another interconnect, for example, a local area network, is required for host-to-host OpenVMS cluster communications. On Alpha computers, this support is limited to the KZPBA adapter. Newer SCSI host-based adapters for Alpha computers support only directly attached SCSI storage.

Beginning with OpenVMS Version 8.2-1, support is available for shared SCSI storage in a two-node OpenVMS Integrity server systems configuration using the MSA30-MI storage shelf.

Shared SCSI storage in an OpenVMS Cluster system enables computers connected to a single SCSI bus to share access to SCSI storage devices directly. This capability makes it possible to build highly available servers using shared access to SCSI storage.

3.6.1. Design for OpenVMS Alpha Configurations

Beginning with OpenVMS Alpha Version 6.2-1H3, OpenVMS Alpha supports up to three nodes on a shared SCSI bus as the storage interconnect. A quorum disk can be used on the SCSI bus to improve

the availability of two-node configurations. Host-based RAID (including host-based shadowing) and the MSCP server are supported for shared SCSI storage devices.

Using the SCSI hub DWZZH-05, four nodes can be supported in a SCSI multihost OpenVMS Cluster system. In order to support four nodes, the hub's fair arbitration feature must be enabled.

For a complete description of these configurations, see *Guidelines for OpenVMS Cluster Configurations*.

3.6.2. Design for OpenVMS Integrity server Configurations

Shared SCSI storage in an OpenVMS Integrity server Cluster system is subject to the following restrictions:

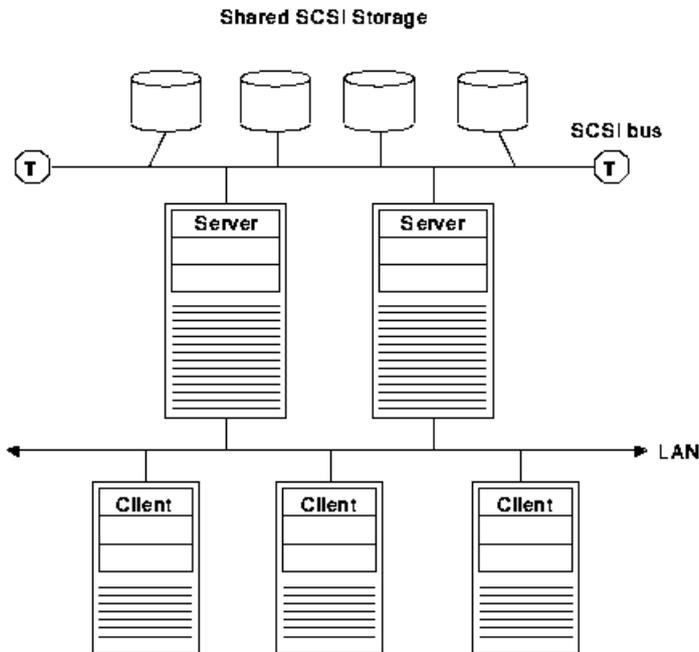
- Maximum of two OpenVMS Integrity server systems connected to a single SCSI bus.
- Maximum of four shared-SCSI buses connected to each system.
- rx1600 and rx2600 family systems are supported.
- A7173A HBA is the only supported HBA.
- MSA30-MI storage enclosure is the only supported SCSI storage type.
- Ultra320 SCSI disk family is the only supported disk family.

In Figure 3.10 the SCSI IDs of 6 and 7, are required in this configuration. One of the systems must have a SCSI ID of 6 for each A7173A adapter port connected to a shared SCSI bus, instead of the factory-set default of 7. You can use the U320_SCSI pscsi.efi utility, included in the IPF Offline Diagnostics and Utilities CD, to change the SCSIID. The procedure for doing this is documented in the *HP A7173APCI-X Dual Channel Ultra 320 SCSI Host Bus Adapter Installation Guide*.

3.6.3. Examples

Figure 3.9 shows an OpenVMS Cluster configuration that uses a SCSI interconnect for shared access to SCSI devices. Note that another interconnect, a LAN in this example, is used for host-to-host communications.

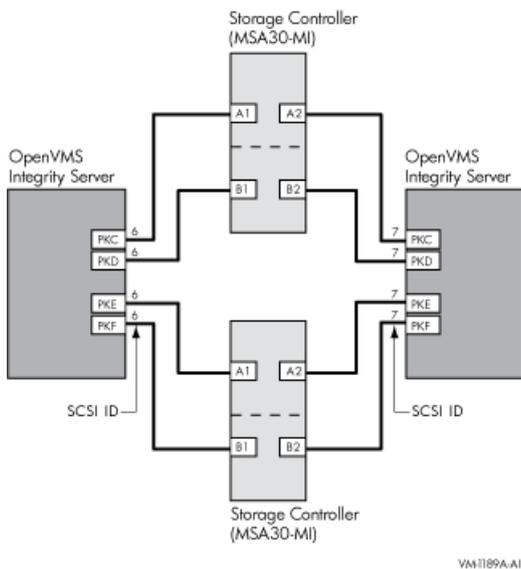
Figure 3.9. Three-Node OpenVMS Cluster Configuration Using a Shared SCSI Interconnect



ZK-7479A-GE

Figure 3.10 illustrates the two-node OpenVMS Integrity server configuration. Note that a second interconnect, a LAN, is required for host-to-host OpenVMS Cluster communications. (OpenVMS Cluster communications are also known as SCA (System Communications Architecture) communications).

Figure 3.10. Two-Node OpenVMS Integrity server Cluster System



VW1189A-A1

3.7. Serial Attached SCSI (SAS) (Integrity servers Only)

OpenVMS Cluster systems support SAS as a storage interconnect. SAS is a point-to-point architecture that transfers data to and from SCSI storage devices by using serial communication (one bit at a time). SAS uses the SAS devices and differential signaling method to achieve reliable, high-speed serial communication.

SAS combines high-end features from fiber channel (such as multi-initiator support and full duplex communication) and the physical interface leveraged from SATA (for better compatibility and investment protection), with the performance, reliability and ease of use of traditional SCSI technology.

3.8. Multihost Fibre Channel OpenVMS Cluster Systems

OpenVMS Cluster systems support FC interconnect as a storage interconnect. Fibre Channel is an ANSI standard network and storage interconnect that offers many advantages over other interconnects, including high-speed transmission and long interconnect distances. A second interconnect is required for node-to-node communications.

3.8.1. Design

OpenVMS Alpha supports the Fibre Channel SAN configurations described in the latest *HP Storage Works SAN Design Reference Guide* and in the Data Replication Manager (DRM) user documentation. This configuration support includes multiswitch Fibre Channel fabrics, up to 500 meters of multimode fiber, and up to 100 kilometers of single-mode fiber. In addition, DRM configurations provide long-distance intersite links (ISLs) through the use of the Open Systems Gateway and wave division multiplexors. OpenVMS supports sharing of the fabric and the HSG storage with non-OpenVMS systems.

OpenVMS provides support for the number of hosts, switches, and storage controllers specified in the Storage Works documentation. In general, the number of hosts and storage controllers is limited only by the number of available fabric connections.

Host-based RAID (including host-based shadowing) and the MSCP server are supported for shared Fibre Channel storage devices. Multipath support is available for these configurations.

For a complete description of these configurations, see *Guidelines for OpenVMS Cluster Configurations*.

Chapter 4. The OpenVMS Cluster Operating Environment

This chapter describes how to prepare the OpenVMS Cluster operating environment.

4.1. Preparing the Operating Environment

To prepare the cluster operating environment, there are a number of steps you perform on the first OpenVMS Cluster node before configuring other computers into the cluster. The following table describes these tasks.

Task	Section
Check all hardware connections to computer, interconnects, and devices.	Described in the appropriate hardware documentation.
Verify that all microcode and hardware is set to the correct revision levels.	Contact your support representative.
Install the OpenVMS operating system.	Section 4.2
Install all software licenses, including OpenVMS Cluster licenses.	Section 4.3
Install layered products.	Section 4.4
Configure and start LANCP or DECnet for satellite booting	Section 4.5

4.2. Installing the OpenVMS Operating System

Only one OpenVMS operating system version can exist on a system disk. Therefore, when installing or upgrading the OpenVMS operating systems ensure that you:

- Install the OpenVMS Integrity servers operating system on each Integrity system disk
- Install the OpenVMS Alpha operating system on each Alpha system disk

Note

Mixed architecture clusters of either OpenVMS Integrity server systems and OpenVMS Alpha systems are supported.

4.2.1. System Disks

A system disk is one of the few resources that cannot be shared between Integrity and Alpha systems.

Once booted, Integrity server systems and Alpha systems can share access to data on any disk in the OpenVMS Cluster, including system disks. For example, an Integrity server system can mount an

Alpha system disk as a data disk and an Alpha system can mount an Integrity server system disk as a data disk.

Note

An OpenVMS Cluster running both implementations of DECnet requires a system disk for DECnet for OpenVMS (Phase IV) and another system disk for DECnet-Plus (Phase V). For more information, see the DECnet-Plus documentation.

4.2.2. Where to Install

You might want to set up common system disks according to these guidelines:

IF you want the cluster to have...	THEN perform the installation or upgrade...
One common system disk for all computer members	Once on the cluster common system disk.
A combination of one or more common system disks and one or more local (individual) system disks	<ul style="list-style-type: none"> • Once for each system disk or <ul style="list-style-type: none"> • Once on a common system disk and then run the CLUSTER_CONFIG.COM procedure to create duplicate system disks (thus enabling systems to have their own local system disk)
<p>Note: If your cluster includes multiple common system disks, you must later coordinate system files to define the cluster operating environment, as described in Chapter 5.</p> <p>Reference: See Section 8.5 for information about creating a duplicate system disk.</p>	

Example: If your OpenVMS Cluster consists of 10 computers, four of which boot from a common Integrity server system disk, two of which boot from a second common Integrity system disk, two of which boot from a common Alpha system disk, and two of which boot from their own local system disk, you need to perform an installation five times.

4.2.3. Information Required

Table 4.1 table lists the questions that the OpenVMS operating system installation procedure prompts you with and describes how certain system parameters are affected by responses you provide. You will notice that two of the prompts vary, depending on whether the node is running DECnet. The table also provides an example of an installation procedure that is taking place on a node named JUPITR.

Important: Be sure you determine answers to the questions before you begin the installation.

Note about versions: Refer to the appropriate OpenVMS Release Notes document for the required version numbers of hardware and firmware. When mixing versions of the operating system in an OpenVMS Cluster, check the release notes for information about compatibility.

Reference: Refer to the appropriate OpenVMS upgrade and installation manual for complete installation instructions.

Table 4.1. Information Required to Perform an Installation

Prompt	Response			Parameter
Will this node be a cluster member (Y/N)?	WHEN you respond...	AND...	THEN the VAXcluster parameter is set to...	VAXCLUSTER
	N	CI and DSSI hardware is not present	0 — Node will not participate in the OpenVMS Cluster.	
	N	CI and DSSI hardware is present	1 — Node will automatically participate in the OpenVMS Cluster in the presence of CI or DSSI hardware.	
	Y		2 — Node will participate in the OpenVMS Cluster.	
What is the node's DECnet node name?	If the node is running DECnet, this prompt, the following prompt, and the SCSSYSTEMID prompt are displayed. Enter the DECnet node name or the DECnet-Plus node synonym (for example, JUPITR). If a node synonym is not defined, SCSNODE can be any name from 1 to 6 alphanumeric characters in length. The name cannot include dollar signs (\$) or underscores (_).			SCSNODE
What is the node's DECnet node address?	Enter the DECnet node address (for example, a valid address might be 2.211). If an address has not been assigned, enter 0 now and enter a valid address when you start DECnet (discussed later in this chapter). For DECnet-Plus, this question is asked when nodes are configured with a Phase IV compatible address. If a Phase IV compatible address is not configured, then the SCSSYSTEMID system parameter can be set to any value.			SCSSYSTEMID
What is the node's SCS node name?	If the node is not running DECnet, this prompt and the following prompt are displayed in place of the two previous prompts. Enter a name of 1 to 6 alphanumeric characters that uniquely names this node. At least 1 character must be a letter. The name cannot include dollar signs (\$) or underscores (_).			SCSNODE
What is the node's SCSSYSTEMID number?	This number must be unique within this cluster. SCSSYSTEMID is the low-order 32 bits of the 48-bit system identification number. If the node is running DECnet for OpenVMS, calculate the value from the DECnet address using the following formula: $\text{SCSSYSTEMID} = (\text{DECnet-area-number} * 1024) + (\text{DECnet-node-number})$			SCSSYSTEMID

Prompt	Response		Parameter
	Example: If the DECnet address is 2.211, calculate the value as follows: SCSSYSTEMID = (2 * 1024) + 211 = 2259		
Will the Ethernet be used for cluster communications (Y/N)?	IF you respond...	THEN the NISCS_LOAD_PEA0 parameter is set to...	NISCS_LOAD_PEA0
	N	0 — PEDRIVER is not loaded ¹ ; cluster communications does not use Ethernet or FDDI.	
	Y	1 — Loads PEDRIVER to enable cluster communications over Ethernet or FDDI.	
Will the IP interconnect be used for cluster communications (Y/N)?	IF you respond...	THEN the NISCS_USE_UDP parameter is set to...	NISCS_USE_UDP
	N	0 — Cluster over IP is disabled and uses the LAN interconnect for cluster communication	
	Y	1 — Cluster over IP is enabled and communicates using the TCP/IP stack. During the boot process, the TCP/IP driver and then the PEDRIVER authorization information is loaded for cluster communication. The hello packets are transmitted using IP multicast and unicast.	
Enter this cluster's group number:	Enter a number in the range of 1 to 4095 or 61440 to 65535 (see Section 2.5). This value is stored in the CLUSTER_AUTHORIZE.DAT file in the SYS\$COMMON:[SYSEXE] directory.		Not applicable
Enter this cluster's password:	Enter the cluster password. The password must be from 1 to 31 alphanumeric characters in length and can include dollar signs (\$) and underscores (_) (see Section 2.5). This value is stored in scrambled form in the CLUSTER_AUTHORIZE.DAT file in the SYS\$COMMON:[SYSEXE] directory.		Not applicable
Reenter this cluster's password for verification:	Reenter the password.		Not applicable
Will JUPITR be a disk server (Y/N)?	IF you respond...	THEN the MSCP_LOAD parameter is set to...	MSCP_LOAD

Prompt	Response	Parameter	
	N	0 — The MSCP server will not be loaded. This is the correct setting for configurations in which all OpenVMS Cluster nodes can directly access all shared storage and do not require LAN failover.	
	Y	1 — Loads the MSCP server with attributes specified by the MSCP_SERVE_ALL parameter, using the default CPU load capacity.	
Will JUPITR serve HSC or RF disks (Y/N)?	IF you respond...	THEN the MSCP_SERVE_ALL parameter is set to...	MSCP_SERVE_ALL
	Y	1 — Serves all available disks.	
	N	2 — Serves only locally connected (not HSC, HSJ, or RF) disks.	
Enter a value for JUPITR's ALLOCLASS parameter. ²	<p>The value is dependent on the system configuration:</p> <ul style="list-style-type: none"> If the system disk is connected to a dual-pathed disk, enter a value from 1 to 255 that will be used on both storage controllers. If the system is connected to a shared SCSI or SAS bus (it shares storage on that bus with another system) and if it does not use port allocation classes for naming the SCSI or SAS disks, enter a value from 1 to 255. This value must be used by all the systems and disks connected to the SCSI or SAS bus. <p>Reference: For complete information about portal location classes, see Section 6.2.1.</p> <ul style="list-style-type: none"> If the system will use Volume Shadowing for OpenVMS, enter a value from 1 to 255. <p>Reference: For more information, see the <i>VSI OpenVMS Volume Shadowing Guide</i>.</p> <ul style="list-style-type: none"> If none of the above are true, enter 0 (zero). 	ALLOCLASS	
Does this cluster contain a quorum disk [N]?	Enter Y or N, depending on your configuration. If you enter Y, the procedure prompts for the name of the quorum disk. Enter the device name of the	DISK_QUORUM	

Prompt	Response	Parameter
	quorum disk. (Quorum disks are discussed in Chapter 2).	

¹PEDRIVER is the LAN port emulator driver that implements the NISCA protocol and controls communications between local and remote LAN ports.

²Refer to Section 6.2 for complete information about device naming conventions.

4.3. Installing Software Licenses

While rebooting at the end of the installation procedure, the system displays messages warning that you must install the operating system software and the OpenVMS Cluster software license. The OpenVMS Cluster software supports the OpenVMS License Management Facility (LMF). License units for clustered systems are allocated on an unlimited system-use basis.

4.3.1. Guidelines

Be sure to install all OpenVMS Cluster licenses and all licenses for layered products and DECnet as soon as the system is available. Procedures for installing licenses are described in the release notes distributed with the software kit and in the *VSI OpenVMS License Management Utility Manual*. Additional licensing information is described in the respective SPDs.

Use the following guidelines when you install software licenses:

- Install an OpenVMS Cluster Software for Alpha license for each Alpha processor in the OpenVMS Cluster.
- Install an OpenVMS Cluster Software for Integrity server system license for each Integrity server processor in the OpenVMS Cluster.
- Install or upgrade licenses for layered products that runs on all nodes in an OpenVMS Cluster system.
- OpenVMS Product Authorization Keys (PAKs) that have the Alpha option can be loaded and used only on Alpha processors. PAKs that have the Integrity servers option can be loaded and used only on Integrity server processors. However, PAKs can be located in a **license database (LDB)** that is shared by all processors(Integrity servers and Alpha).
- PAK types, such as Activity PAKs (also known as concurrent or n-user PAKs)and Personal Use PAKs (identified by the RESERVE_UNITS option) work on Alpha systems.
- PAK types, such as PCL PAKs (per core licensing) are only supported on Integrity servers.
- License management commands can be issued from every node in the cluster.

4.4. Installing Layered Products

By installing layered products before other nodes are added to the OpenVMS Cluster, the software is installed automatically on new members when they are added to the OpenVMS Cluster system.

Note: For clusters with multiple system disks (Integrity servers or Alpha) you must perform a separate installation for each system disk.

4.4.1. Procedure

Table 4.2 describes the actions you take to install layered product on a common system disk.

Table 4.2. Installing Layered Products on a Common System Disk

Phase	Action
Before installation	<p>Perform one or more of the following steps, as necessary for your system.</p> <ol style="list-style-type: none"> 1. Check each node's system parameters and modify the values, if necessary. Refer to the layered-product installation guide or release notes for information about adjusting system parameter values. 2. If necessary, disable logins on each node that boots from the disk using the DCL command SET LOGINS/INTERACTIVE=0. Send a broadcast message to notify users about the installation.
Installation	<p>Refer to the appropriate layered-product documentation for product-specific installation information. Perform the installation once for each system disk.</p>
After installation	<p>Perform one or more of the following steps, as necessary for your system.</p> <ol style="list-style-type: none"> 1. If necessary, create product-specific files in the SYSSPECIFIC directory on each node. (The installation utility describes whether or not you need to create a directory in SYSSPECIFIC.) When creating files and directories, be careful to specify exactly where you want the file to be located: <ul style="list-style-type: none"> • Use SYSSPECIFIC or SYS\$COMMON instead of SYS\$SYSROOT. • Use SYSSPECIFIC:[SYSEXE] or SYS\$COMMON:[SYSEXE] instead of SYS\$SYSTEM. <p>Reference: Section 5.3 describes directory structures in more detail.</p> 2. Modify files in SYSSPECIFIC if the installation procedure tells you to do so. Modify files on each node that boots from this system disk. 3. Reboot each node to ensure that: <ul style="list-style-type: none"> • The node is set up to run the layered product correctly. • The node is running the latest version of the layered product. 4. Manually run the installation verification procedure (IVP) if you did not run it during the layered product installation. Run the IVP from at least one node in the OpenVMS Cluster, but preferably from all nodes that boot from this system disk.

4.5. Configuring and Starting a Satellite Booting Service

After you have installed the operating system and the required licenses on the first OpenVMS Cluster computer, you can configure and start a satellite booting service. You can use the LANCP utility, or DECnet software, or both.

VSI recommends LANCP for booting OpenVMS Cluster satellites. LANCP has shipped with the OpenVMS operating system since Version 6.2. It provides a general-purpose MOP booting service that can be used for booting satellites into an OpenVMS Cluster. (LANCP can service all types of MOP downline load requests, including those from terminal servers, LAN resident printers, and X terminals, and can be used to customize your LAN environment).

DECnet provides a MOP booting service for booting OpenVMS Cluster satellites, as well as other local and wide area network services, including task-to-task communications for applications.

Note

If you plan to use LANCP in place of DECnet, and you also plan to move from DECnet PhaseIV to DECnet-Plus, VSI recommends the following order:

1. Replace DECnet with LANCP for satellite booting (MOP downline load service) using LAN \$POPULATE.COM.
 2. Migrate from DECnet Phase IV to DECnet-Plus.
-

There are two cluster configuration command procedures, `CLUSTER_CONFIG_LAN.COM` and `CLUSTER_CONFIG.COM`. `CLUSTER_CONFIG_LAN.COM` uses LANCP to provide MOP services to boot satellites; `CLUSTER_CONFIG.COM` uses DECnet for the same purpose.

Before choosing LANCP, DECnet, or both, consider the following factors:

- Applications you will be running on your cluster

DECnet task-to-task communications is a method commonly used for communication between programs that run on different nodes in a cluster or a network. If you are running a program with that dependency, you need to run DECnet. If you are not running any programs with that dependency, you do not need to run DECnet.

- Limiting applications that require DECnet to certain nodes in your cluster

If you are running applications that require DECnet task-to-task communications, you can run those applications on a subset of the nodes in your cluster and restrict DECnet usage to those nodes. You can use LANCP software on the remaining nodes and use a different network, such as TCP/IP Services for OpenVMS, for other network services.

- Managing two types of software for the same purpose

If you are already using DECnet for booting satellites, you may not want to introduce another type of software for that purpose. Introducing any new software requires time to learn and manage it.

- LANCP MOP services can coexist with DECnet MOP services in an OpenVMS Cluster in the following ways:

- Running on different systems

For example, DECnet MOP service is enabled on some of the systems on the LAN and LAN MOP is enabled on other systems.

- Running on different LAN devices on the same system

For example, DECnet MOP service is enabled on a subset of the available LAN devices on the system and LAN MOP is enabled on the remainder.

- Running on the same LAN device on the same system but targeting a different set of nodes for service

For example, both DECnet MOP and LAN MOP are enabled but LAN MOP has limited the nodes to which it will respond. This allows DECnet MOP to respond to the remaining nodes.

Instructions for configuring both LANCP and DECnet are provided in this section.

4.5.1. Configuring and Starting the LANCP Utility

You can use the LAN Control Program (LANCP) utility to configure a local area network (LAN). You can also use the LANCP utility, in place of DECnet or in addition to DECnet, to provide support for booting satellites in an OpenVMS Cluster and for servicing all types of MOP downline load requests, including those from terminal servers, LAN resident printers, and X terminals.

Reference: For more information about using the LANCP utility to configure a LAN, see the *VSI OpenVMS System Manager's Manual* and the *VSI OpenVMS System Management Utilities Reference Manual*.

4.5.2. Booting Satellite Nodes with LANCP

The LANCP utility provides a general-purpose MOP booting service that can be used for booting satellites into an OpenVMS Cluster. It can also be used to service all types of MOP downline load requests, including those from terminal servers, LAN resident printers, and X terminals. To use LANCP for this purpose, all OpenVMS Cluster nodes must be running OpenVMS Version 6.2 or higher.

The CLUSTER_CONFIG_LAN.COM cluster configuration command procedure uses LANCP in place of DECnet to provide MOP services to boot satellites.

Note: If you plan to use LANCP in place of DECnet, and you also plan to move from DECnet for OpenVMS (Phase IV) to DECnet-Plus, VSI recommends the following order:

1. Replace DECnet with LANCP for satellite booting (MOP downline load service), using LAN \$POPULATE.COM.
2. Migrate from DECnet for OpenVMS to DECnet-Plus.

4.5.3. Data Files Used by LANCP

LANCP uses the following data files:

- SYSS\$SYSTEM:LAN\$DEVICE_DATABASE.DAT

This file maintains information about devices on the local node. By default, the file is created in SYSS\$SPECIFIC:[SYSEXE], and the system looks for the file in that location. However, you can modify the file name or location for this file by redefining the systemwide logical name LAN \$DEVICE_DATABASE.

- SYSS\$SYSTEM:LAN\$NODE_DATABASE.DAT

This file contains information about the nodes for which LANCP will supply boot service. This file must be shared among all nodes in the OpenVMS Cluster, including Integrity servers, Alpha, and VAX systems. By default, the file is created in SYSS\$COMMON:[SYSEXE], and the system

looks for the file in that location. However, you can modify the file name or location for this file by redefining the systemwide logical name LAN\$NODE_DATABASE.

4.5.4. Using LAN MOP Services in New Installations

To use LAN MOP services for satellite booting in new installations, follow these steps:

1. Add the startup command for LANCP.

You should start up LANCP as part of your system startup procedure. To do this, remove the comment from the line in SYSS\$MANAGER:SYSTARTUP_VMS.COM that runs the LAN \$STARTUP command procedure. If your OpenVMS Cluster system will have more than one system disk, see Section 4.5.3 for a description of logicals that can be defined for locating LANCP configuration files.

```
$ @SYS$STARTUP:LAN$STARTUP
```

You should now either reboot the system or invoke the preceding command procedure from the system manager's account to start LANCP.

2. Follow the steps in Chapter 8 for configuring an OpenVMS Cluster system and adding satellites. Use the CLUSTER_CONFIG_LAN.COM command procedure instead of CLUSTER_CONFIG.COM. If you invoke CLUSTER_CONFIG.COM, it gives you the option to switch to running CLUSTER_CONFIG_LAN.COM if the LANCP process has been started.

4.5.5. Using LAN MOP Services in Existing Installations

To migrate from DECnet MOP services to LAN MOP services for satellite booting, follow these steps:

1. Redefine the LANCP database logical names.

This step is optional. If you want to move the data files used by LANCP, LAN \$DEVICE_DATABASE and LAN\$NODE_DATABASE, off the system disk, redefine their systemwide logical names. Add the definitions to the system startup files.

2. Use LANCP to create the LAN\$DEVICE_DATABASE

The permanent LAN\$DEVICE_DATABASE is created when you issue the first LANCP DEVICE command. To create the database and get a list of available devices, enter the following commands:

```
$ MCR LANCP
LANCP> LIST DEVICE /MOPDLL
%LANCP-I-FNFDEV, File not found, LAN$DEVICE_DATABASE
%LANCP-I-CREATDEV, Created LAN$DEVICE_DATABASE file
```

Device Listing, permanent database:

```
--- MOP Downline Load Service Characteristics ---
Device      State      Access Mode      Client              Data Size
-----
ESA0        Disabled  NoExclusive      NoKnownClientsOnly 246 bytes
FCA0        Disabled  NoExclusive      NoKnownClientsOnly 246 bytes
```

3. Use LANCP to enable LAN devices for MOP booting.

By default, the LAN devices have MOP booting capability disabled. Determine the LAN devices for which you want to enable MOP booting. Then use the DEFINE command in the LANCP utility to enable these devices to service MOP boot requests in the permanent database, as shown in the following example:

```
LANCP> DEFINE DEVICE ESA0:/MOP=ENABLE
```

4. Run LAN\$POPULATE.COM (found in SYS\$EXAMPLES) to obtain MOP booting information and to produce LAN\$DEFINE and LAN\$DECNET_MOP_CLEANUP, which are site specific.

LAN\$POPULATE extracts all MOP booting information from a DECnet Phase IVNETNODE_REMOTE.DAT file or from the output of the DECnet-Plus NCL command SHOW MOP CLIENT * ALL.

For DECnet Phase IV sites, the LAN\$POPULATE procedure scans all DECnet areas (1–63) by default. If you MOP boot systems from only a single or a few DECnet areas, you can cause the LAN\$POPULATE procedure to operate on a single area at a time by providing the area number as the P1 parameter to the procedure, as shown in the following example (including log):

```
$ @SYS$EXAMPLES:LAN$POPULATE 15
LAN$POPULATE - V1.0
```

```
Do you want help (Y/N) <N>:
```

```
LAN$DEFINE.COM has been successfully created.
```

```
To apply the node definitions to the LANCP permanent database,
invoke the created LAN$DEFINE.COM command procedure.
```

```
VSI recommends that you review LAN$DEFINE.COM and remove any
obsolete entries prior to executing this command procedure.
```

```
A total of 2 MOP definitions were entered into LAN$DEFINE.COM
```

5. Run LAN\$DEFINE.COM to populate LAN\$NODE_DATABASE.

LAN\$DEFINE populates the LANCP downline loading information into the LAN node database, SYS\$COMMON:[SYSEVE]LAN\$NODE_DATABASE.DAT file. VSI recommends that you review LAN\$DEFINE.COM and remove any obsolete entries before executing it.

In the following sequence, the LAN\$DEFINE.COM procedure that was just created is displayed on the screen and then executed:

```
$ TYPE LAN$DEFINE.COM
$ !
$ ! This file was generated by LAN$POPULATE.COM on 16-DEC-1996
09:20:31
$ ! on node CLU21.
$ !
$ ! Only DECnet Area 15 was scanned.
$ !
$ MCR LANCP
Define Node PORK      /Address=08-00-2B-39-82-85 /File=APB.EXE -
                    /Root=$21$DKA300:<SYS11.> /Boot_type=Alpha_Satellite
Define Node JYPIG    /Address=08-00-2B-A2-1F-81 /File=APB.EXE -
                    /Root=$21$DKA300:<SYS10.> /Boot_type=Alpha_Satellite
```

```

EXIT
$ @LAN$DEFINE
%LANCP-I-FNFNOD, File not found, LAN$NODE_DATABASE
-LANCP-I-CREATNOD, Created LAN$NODE_DATABASE file
$

```

The following example shows a LAN\$DEFINE.COM command procedure that was generated by LAN\$POPULATE for migration from DECnet-Plus to LANCP.

```

$ ! LAN$DEFINE.COM - LAN MOP Client Setup
$ !
$ ! This file was generated by LAN$POPULATE.COM at 8-DEC-1996
14:28:43.31
$ ! on node BIGBOX.
$ !
$ SET NOON
$ WRITE SYS$OUTPUT "Setting up MOP DLL clients in LANCP..."
$ MCR LANCP
SET NODE SLIDER
/ADDRESS=08-00-2B-12-D8-72/ROOT=BIGBOX$DKB0:<SYS10.>/BOOT_TYP
E=VAX_satellite/FILE=NISCS_LOAD.EXE
DEFINE NODE SLIDER
/ADDRESS=08-00-2B-12-D8-72/ROOT=BIGBOX$DKB0:<SYS10.>/BOOT_TYP
E=VAX_satellite/FILE=NISCS_LOAD.EXE
EXIT
$ !
$ WRITE SYS$OUTPUT "DECnet Phase V to LAN MOPDLL client migration
complete!"
$ EXIT

```

6. Run LAN\$DECNET_MOP_CLEANUP.COM.

You can use LAN\$DECNET_MOP_CLEANUP.COM to remove the clients' MOP downline loading information from the DECnet database. VSI recommends that you review LAN\$DECNET_MOP_CLEANUP.COM and remove any obsolete entries before executing it.

The following example shows a LAN\$DECNET_MOP_CLEANUP.COM command procedure that was generated by LAN\$POPULATE for migration from DECnet-Plus to LANCP.

Note: When migrating from DECnet-Plus, additional cleanup is necessary. You must edit your NCL scripts (*.NCL) manually.

```

$ ! LAN$DECNET_MOP_CLEANUP.COM - DECnet MOP Client Cleanup
$ !
$ ! This file was generated by LAN$POPULATE.COM at 8-DEC-1995
14:28:43.47
$ ! on node BIGBOX.
$ !
$ SET NOON
$ WRITE SYS$OUTPUT "Removing MOP DLL clients from DECnet database..."
$ MCR NCL
DELETE NODE 0 MOP CLIENT SLIDER
EXIT
$ !
$ WRITE SYS$OUTPUT "DECnet Phase V MOPDLL client cleanup complete!"
$ EXIT

```

7. Start LANCP.

To start LANCP, execute the startup command procedure as follows:

```
$ @SYS$STARTUP:LAN$STARTUP
%RUN-S-PROC_ID, identification of created process is 2920009B
$
```

You should start up LANCP for all boot nodes as part of your system startup procedure. To do this, include the following line in your site-specific startup file (SYS\$MANAGER:SYSTARTUP_VMS.COM):

```
$ @SYS$STARTUP:LAN$STARTUP
```

If you have defined logicals for either LAN\$DEVICE_DATABASE or LAN\$NODE_DATABASE, be sure that these are defined in your startup files prior to starting up LANCP.

8. Disable DECnet MOP booting.

If you use LANCP for satellite booting, you may no longer need DECnet to handle MOP requests. If this is the case for your site, you can turn off this capability with the appropriate NCP command (DECnet for OpenVMS) or NCL commands (DECnet-Plus).

For more information about the LANCP utility, see the *VSI OpenVMS System Manager's Manual* and the *VSI OpenVMS System Management Utilities Reference Manual*.

4.5.6. Configuring DECnet

The process of configuring the DECnet network typically entails several operations, as shown in Table 4.3. An OpenVMS Cluster running both implementations of DECnet requires a system disk for DECnet for OpenVMS (Phase IV) and another system disk for DECnet-Plus (Phase V).

Note: DECnet for OpenVMS implements Phase IV of Digital Network Architecture (DNA). DECnet-Plus implements Phase V of DNA. The following discussions are specific to the DECnet for OpenVMS product.

Reference: Refer to the DECnet-Plus documentation for equivalent DECnet-Plus configuration information.

Table 4.3. Procedure for Configuring the DECnet Network

Step	Action
1	<p>Log in as system manager and execute the NETCONFIG.COM command procedure as shown. Enter information about your node when prompted. Note that DECnet-Plus nodes execute the NET\$CONFIGURE.COM command procedure.</p> <p>Reference: See the DECnet for OpenVMS or the DECnet-Plus documentation, as appropriate, for examples of these procedures.</p>
2	<p>When a node uses multiple LAN adapter connections to the same LAN and also uses DECnet for communications, you must <i>disable DECnet use</i> of all but one of the LAN devices.</p> <p>To do this, remove all but one of the lines and circuits associated with the adapters connected to the same LAN or extended LAN from the DECnet configuration database after the NETCONFIG.COM procedure is run.</p>

Step	Action
	<p>For example, issue the following commands to invoke NCP and disable DECnet use of the LAN device XQB0:</p> <pre>\$ RUN SYS\$SYSTEM:NCP NCP> PURGE CIRCUIT QNA-1 ALL NCP> DEFINE CIRCUIT QNA-1 STA OFF NCP> EXIT</pre> <p>References:</p> <p>See <i>Guidelines for OpenVMS Cluster Configurations</i> for more information about distributing connections to LAN segments in OpenVMS Cluster configurations.</p> <p>See the DECnet-Plus documentation for information about removing routing circuits associated with all but one LAN adapter. (Note that the LAN adapter issue is not a problem if the DECnet-Plus node uses extended addressing and does not have any Phase IV compatible addressing in use on any of the routing circuits).</p>
3	<p>Make remote node data available clusterwide. NETCONFIG.COM creates in the SYS\$SPECIFIC:[SYSEXE] directory the permanent remote-node database file NETNODE_REMOTE.DAT, in which remote-node data is maintained. To make this data available throughout the OpenVMS Cluster, you move the file to the SYS\$COMMON:[SYSEXE] directory.</p> <p>Example: Enter the following commands to make DECnet information available clusterwide:</p> <pre>\$ RENAME SYS\$SPECIFIC:[SYSEXE]NETNODE_REMOTE.DAT SYS\$COMMON:[SYSEXE]NETNODE_REMOTE.DAT</pre> <p>If your configuration includes multiple system disks, you can set up a common NETNODE_REMOTE.DAT file automatically by using the following command in SYLOGICALS.COM:</p> <pre>\$ DEFINE/SYSTEM/EXE NETNODE_REMOTE ddcu:[directory]NETNODE_REMOTE.DAT</pre> <p>Notes: VSI recommends that you set up a common NETOBJECT.DAT file clusterwide in the same manner.</p> <p>DECdns is used by DECnet-Plus nodes to manage node data (the namespace). For DECnet-Plus, Session Control Applications replace objects.</p>
4	<p>Designate and enable router nodes to support the use of a cluster alias. At least one node participating in a cluster alias must be configured as a level 1 router.</p> <p>On Integrity servers and Alpha systems, you might need to enable level 1 routing manually because the NETCONFIG.COM procedure does not prompt you with the routing question.</p> <p>Depending on whether the configuration includes a combination of Integrity sever nodes and Alpha nodes, you must enable level 1 routing manually (see the example below) on one of the Alpha nodes.</p> <p>Example: On Alpha systems, if you need to enable level 1 routing on Alpha node, invoke the NCP utility to do so. For example:</p>

Step	Action
	<pre>\$ RUN SYS\$SYSTEM:NCP NCP> DEFINE EXECUTOR TYPE ROUTING IV</pre> <p>Note: On Integrity servers and Alpha systems, level 1 routing is supported to enable cluster alias operations only.</p>
5	<p>Optionally, define a cluster alias. If you want to define a cluster alias, invoke the NCP utility to do so. The information you specify using these commands is entered in the DECnet permanent executor database and takes effect when you start the network.</p> <p>Example: The following NCP commands establish SOLAR as an alias:</p> <pre>\$ RUN SYS\$SYSTEM:NCP NCP> DEFINE NODE 2.1 NAME SOLAR NCP> DEFINE EXECUTOR ALIAS NODE SOLAR NCP> EXIT \$</pre> <p>Reference: Section 4.5.8 describes the cluster alias. Section 4.5.9 describes how to enable alias operations for other computers. See the DECnet-Plus documentation for information about setting up a cluster alias on DECnet-Plus nodes.</p> <p>Note: DECnet for OpenVMS nodes and DECnet-Plus nodes cannot share a cluster alias.</p>

4.5.7. Starting DECnet

If you are using DECnet-Plus, a separate step is not required to start the network. DECnet-Plus starts automatically on the next reboot after the node has been configured using the NET \$CONFIGURE.COM procedure.

If you are using DECnet for OpenVMS, at the system prompt, enter the following command to start the network:

```
$ @SYS$MANAGER:STARTNET.COM
```

To ensure that the network is started each time an OpenVMS Cluster computer boots, add that command line to the appropriate startup command file or files. (Startup command files are discussed in Section 5.5).

4.5.8. What is Cluster Alias?

The cluster alias acts as a single network node identifier for an OpenVMS Cluster system. When enabled, the cluster alias makes all the OpenVMS Cluster nodes appear to be one node from the point of view of the rest of the network.

Computers in the cluster can use the alias for communications with other computers in a DECnet network. For example, networked applications that use the services of an OpenVMS Cluster should use an alias name. Doing so ensures that the remote access will be successful when at least one OpenVMS Cluster member is available to process the client program's requests.

Rules:

- DECnet for OpenVMS (Phase IV) allows a maximum of 64 OpenVMS Cluster computers to participate in a cluster alias. If your cluster includes more than 64 computers, you must determine which 64 should participate in the alias and then define the alias on those computers.

At least one of the OpenVMS Cluster nodes that uses the alias node identifier must have level 1 routing enabled.

- On Integrity servers and Alpha nodes, routing between multiple circuits is not supported. However, routing is supported to allow cluster alias operations. Level 1 routing is supported only for enabling the use of a cluster alias. The DVNETEXT PAK must be used to enable this limited function.
- On Integrity servers, Alpha, and VAX systems, all cluster nodes sharing the same alias node address must be in the same area.
- DECnet-Plus allows a maximum of 96 OpenVMS Cluster computers to participate in the cluster alias.

DECnet-Plus does not require that a cluster member be a routing node, but an adjacent Phase V router is required to use a cluster alias for DECnet-Plus systems.

- A single cluster alias can include nodes running either DECnet for OpenVMS or DECnet-Plus, but not both.

4.5.9. Enabling Alias Operations

If you have defined a cluster alias and have enabled routing as shown in Section 4.5.6, you can enable alias operations for other computers *after the computers are up and running in the cluster*. To enable such operations (that is, to allow a computer to accept incoming connect requests directed toward the alias), follow these steps:

1. Log in as system manager and invoke the SYSMAN utility. For example:

```
$ RUN SYS$SYSTEM:SYSMAN
SYSMAN>
```

2. At the SYSMAN> prompt, enter the following commands:

```
SYSMAN> SET ENVIRONMENT/CLUSTER
%SYSMAN-I-ENV, current command environment:
      Clusterwide on local cluster
      Username SYSTEM will be used on nonlocal nodes
SYSMAN> SET PROFILE/PRIVILEGES=(OPER,SYSRV)
SYSMAN> DO MCR NCP SET EXECUTOR STATE OFF
%SYSMAN-I-OUTPUT, command execution on node X...
.
.
.
SYSMAN> DO MCR NCP DEFINE EXECUTOR ALIAS INCOMING ENABLED
%SYSMAN-I-OUTPUT, command execution on node X...
.
.
.
SYSMAN> DO @SYS$MANAGER:STARTNET.COM
%SYSMAN-I-OUTPUT, command execution on node X...
.
.
.
```

Note: VSI does not recommend enabling alias operations for satellite nodes.

Reference: For more details about DECnet for OpenVMS networking and cluster alias, see the *VSI OpenVMS DECnet Networking Manual* and *VSI OpenVMS DECnet Network Management Utilities*. For equivalent information about DECnet-Plus, see the DECnet-Plus documentation.

4.5.10. Configuring TCP/IP

For information on how to configure and start TCP/IP, see the *VSI TCP/IP Services for OpenVMS Installation and Configuration Guide* and the *HP TCP/IP Services for OpenVMS Version 5.7 Release Notes*.

Chapter 5. Preparing a Shared Environment

In any OpenVMS Cluster environment, it is best to share resources as much as possible. Resource sharing facilitates workload balancing because work can be distributed across the cluster.

5.1. Shareable Resources

Most, but not all, resources can be shared across nodes in an OpenVMS Cluster. The following table describes resources that can be shared.

Shareable Resources	Description
System disks	All members of the same architecture ¹ can share a single system disk, each member can have its own system disk, or members can use a combination of both methods.
Data disks	All members can share any data disks. For local disks, access is limited to the local node unless you explicitly setup the disks to be cluster accessible by means of the MSCP server.
Tape drives	All members can share tape drives. (Note that this does not imply that all members can have simultaneous access.) For local tape drives, access is limited to the local node unless you explicitly set up the tapes to be cluster accessible by means of the TMSCP server. Only DSA tapes can be served to all OpenVMS Cluster members.
Batch and print queues	Users can submit batch jobs to any queue in the OpenVMS Cluster, regardless of the processor on which the job will actually execute. Generic queues can balance the load among the available processors.
Applications	Most applications work in an OpenVMS Cluster just as they do on a single system. Application designers can also create applications that run simultaneously on multiple OpenVMS Cluster nodes, which share data in a file.
User authorization files	All nodes can use either a common user authorization file (UAF) for the same access on all systems or multiple UAFs to enable node-specific quotas. If a common UAF is used, all user passwords, directories, limits, quotas, and privileges are the same on all systems.

¹Data on system disks can be shared between Integrity servers and Alpha computers. However, Integrity server nodes cannot boot from an Alpha system disk, and Alpha nodes cannot boot from an Integrity server system disk.

5.1.1. Local Resources

The following table lists resources that are accessible only to the local node.

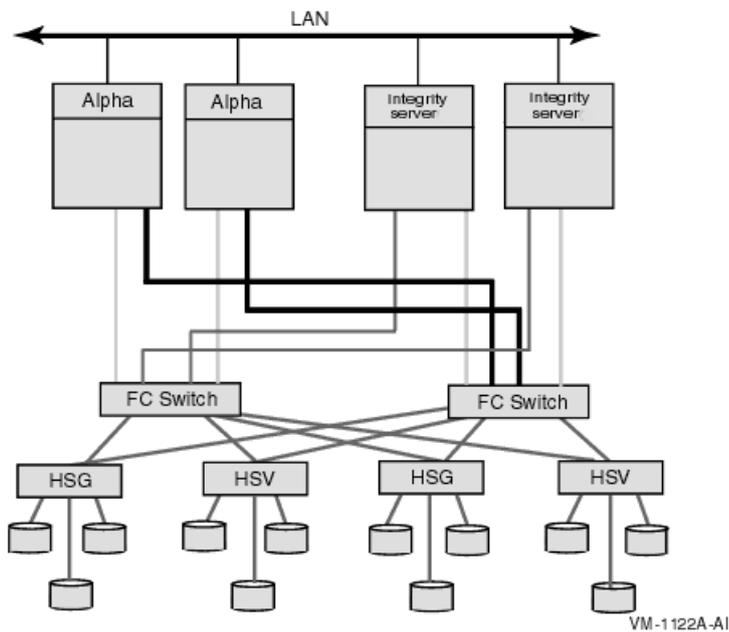
Nonshareable Resources	Description
Memory	Each OpenVMS Cluster member maintains its own memory.
User processes	When a user process is created on an OpenVMS Cluster member, the process must complete on that computer, using local memory.

Nonshareable Resources	Description
Printers	A printer that does not accept input through queues is used only by the OpenVMS Cluster member to which it is attached. A printer that accepts input through queues is accessible by any OpenVMS Cluster member.

5.1.2. Sample Configuration

Figure 5.1 shows an OpenVMS Cluster system that shares FC SAN storage between the Integrity servers and Alpha systems. Each architecture has its own system disk.

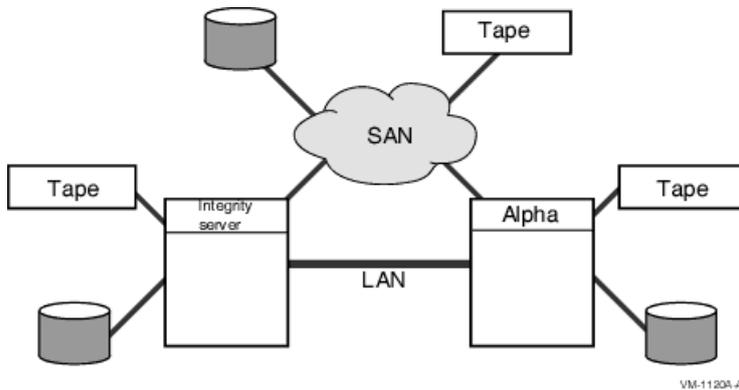
Figure 5.1. Resource Sharing in Mixed-Architecture Cluster System (Integrity servers and Alpha)



5.1.3. Storage in a Mixed-Architecture Cluster

This section describes the rules pertaining to storage, including system disks, in a mixed-architecture cluster consisting of OpenVMS Integrity servers and OpenVMS Alpha systems.

Figure 5.2 is a simplified version of a mixed-architecture cluster of OpenVMS Integrity servers and OpenVMS Alpha systems with locally attached storage and a shared Storage Area Network (SAN).

Figure 5.2. Resource Sharing in Mixed-Architecture Cluster System (Integrity servers and Alpha)

Integrity server systems in a mixed-architecture OpenVMS Cluster system:

- Must have an Integrity server system disk, either a local disk or a shared Fibre Channel disk.
- Can use served Alpha disks and served Alpha tapes.
- Can use SAN disks and tapes.
- Can share the same SAN data disk with Alpha systems.
- Can serve disks and tapes to other cluster members, both Integrity servers and Alpha systems.

Alpha systems in a mixed-architecture OpenVMS Cluster system:

- Must have an Alpha system disk, which can be shared with other clustered Alpha systems.
- Can use locally attached tapes and disks.
- Can serve disks and tapes to both Integrity servers and Alpha systems.
- Can use Integrity servers served data disks.
- Can use SAN disks and tapes.
- Can share the same SAN data disk with Integrity server systems.

5.2. Common-Environment and Multiple-Environment Clusters

Depending on your processing needs, you can prepare either an environment in which all environmental files are shared clusterwide or an environment in which some files are shared clusterwide while others are accessible only by certain computers.

The following table describes the characteristics of common- and multiple-environment clusters.

Cluster Type	Characteristics	Advantages
Common environment		

Cluster Type	Characteristics	Advantages
Operating environment is identical on all nodes in the OpenVMS Cluster.	<p>The environment is set up so that:</p> <ul style="list-style-type: none"> • All nodes run the same programs, applications, and utilities. • All users have the same type of user accounts, and the same logical names are defined. • All users can have common access to storage devices and queues.(Note that access is subject to how access control list [ACL] protection is set up for each user.) • All users can log in to any node in the configuration and work in the same environment as all other users. 	Easier to manage because you use a common version of each system file.
Multiple environment		
Operating environment can vary from node to node.	<p>An individual processor or a subset of processors are set up to:</p> <ul style="list-style-type: none"> • Provide multiple access according to the type of tasks users perform and the resources they use. • Share a set of resources that are not available on other nodes. • Perform specialized functions using restricted resources while other processors perform general time sharing work. • Allow users to work in environments that are specific to the node where they are logged in. 	Effective when you want to share <i>some</i> data among computers but you also want certain computers to serve specialized needs.

5.3. Directory Structure on Common System Disks

The installation or upgrade procedure for your operating system generates a **common system disk**, on which most operating system and optional product files are stored in a system root directory.

5.3.1. Directory Roots

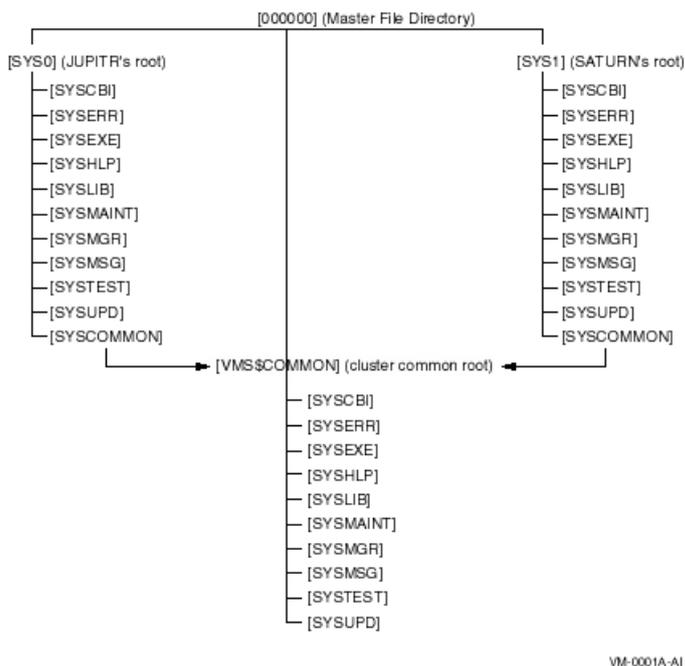
The system disk directory structure is the same on Integrity servers and Alpha systems. Whether the system disk is for an Integrity server system or Alpha, the entire directory structure—that is, the **common root** plus each computer's **local root** is stored on the same disk. After the installation or upgrade completes, you use the CLUSTER_CONFIG.COM or CLUSTER_CONFIG_LAN.COM command procedure described in Chapter 8 to create a local root for each new computer to use when booting into the cluster.

In addition to the usual system directories, each local root contains a [SYS *n*.SYSCOMMON] directory that is a directory alias for [VMS\$COMMON], the cluster common root directory in which cluster common files actually reside. When you add a computer to the cluster, the com procedure defines the common root directory alias.

5.3.2. Directory Structure Example

Figure 5.3 illustrates the directory structure set up for computers JUPITR and SATURN, which are run from a common system disk. The disk's master file directory (MFD) contains the local roots (SYS0 for JUPITR, SYS1 for SATURN) and the cluster common root directory, [VMS\$COMMON].

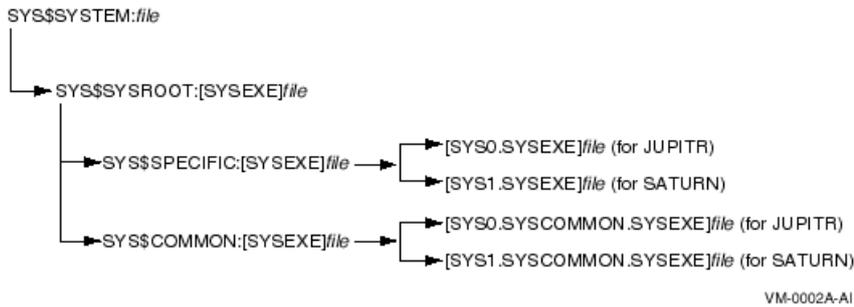
Figure 5.3. Directory Structure on a Common System Disk



5.3.3. Search Order

The logical name SYSSYSROOT is defined as a search list that points first to a local root (SYS\$SYSDEVICE:[SYS0.SYSEXE]) and then to the common root (SYS\$COMMON:[SYSEXE]). Thus, the logical names for the system directories (SYS\$SYSTEM, SYS\$LIBRARY, SYS\$MANAGER, and so forth) point to two directories.

Figure 5.4 shows how directories on a common system disk are searched when the logical name SYS\$SYSTEM is used in file specifications.

Figure 5.4. File Search Order on Common System Disk

Important: Keep this search order in mind when you manipulate system files on a common system disk. Computer-specific files must always reside and be updated in the appropriate computer's system subdirectory.

Examples

1. MODPARAMS.DAT must reside in SYS\$SPECIFIC:[SYSEXE], which is [SYS0.SYSEXE] on JUPITR, and in [SYS1.SYSEXE] on SATURN. Thus, to create a new MODPARAMS.DAT file for JUPITR when logged in on JUPITR, enter the following command:

```
$ EDIT SYS$SPECIFIC:[SYSEXE]MODPARAMS.DAT
```

Once the file is created, you can use the following command to modify it when logged on to JUPITR:

```
$ EDIT SYS$SYSTEM:MODPARAMS.DAT
```

Note that if a MODPARAMS.DAT file does not exist in JUPITR's SYS\$SPECIFIC:[SYSEXE] directory when you enter this command, but there is a MODPARAMS.DAT file in the directory SYS\$COMMON:[SYSEXE], the command edits the MODPARAMS.DAT file in the common directory. If there is no MODPARAMS.DAT file in either directory, the command creates the file in JUPITR's SYS\$SPECIFIC:[SYSEXE] directory.

2. To modify JUPITR's MODPARAMS.DAT when logged in on any other computer that boots from the same common system disk, enter the following command:

```
$ EDIT SYS$SYSDEVICE:[SYS0.SYSEXE]MODPARAMS.DAT
```

3. To modify records in the cluster common system authorization file in a cluster with a single, cluster-common system disk, enter the following commands on any computer:

```
$ SET DEFAULT SYS$COMMON:[SYSEXE]$ RUN SYS$SYSTEM:AUTHORIZE
```

4. To modify records in a computer-specific system authorization file when logged in to another computer that boots from the same cluster common system disk, you must set your default directory to the specific computer. For example, if you have set up a computer-specific system authorization file (SYSUAF.DAT) for computer JUPITR, you must set your default directory to JUPITR's computer-specific [SYSEXE] directory before invoking AUTHORIZE, as follows:

```
$ SET DEFAULT SYS$SYSDEVICE:[SYS0.SYSEXE]
$ RUN SYS$SYSTEM:AUTHORIZE
```

5.4. Clusterwide Logical Names

Clusterwide logical names, introduced in OpenVMS Version 7.2, extend the convenience and ease-of-use features of shareable logical names to OpenVMS Cluster systems. Clusterwide logical names are available on OpenVMS Integrity servers and OpenVMS Alpha systems, in a single or a mixed architecture OpenVMS Cluster.

Existing applications can take advantage of clusterwide logical names without any changes to the application code. Only a minor modification to the logical name tables referenced by the application (directly or indirectly) is required.

New logical names are local by default. Clusterwide is an attribute of a logical name table. In order for a new logical name to be clusterwide, it must be created in a clusterwide logical name table.

Some of the most important features of clusterwide logical names are:

- When a new node joins the cluster, it automatically receives the current set of clusterwide logical names.
- When a clusterwide logical name or name table is created, modified, or deleted, the change is automatically propagated to every other node in the cluster running OpenVMS Version 7.2 or later. Modifications include security profile changes to a clusterwide table.
- Translations are done locally so there is minimal performance degradation for clusterwide name translations.
- Because LNM\$CLUSTER_TABLE and LNM\$SYSCLUSTER_TABLE exist on all systems running OpenVMS Version 7.2 or later, the programs and command procedures that use clusterwide logical names can be developed, tested, and run on nonclustered systems.

5.4.1. Default Clusterwide Logical Name Tables

To support clusterwide logical names, the operating system creates two clusterwide logical name tables and their logical names at system startup, as shown in Table 5.1. These logical name tables and logical names are in addition to the ones supplied for the process, job, group, and system logical name tables. The names of the clusterwide logical name tables are contained in the system logical name directory, LNM\$SYSTEM_DIRECTORY.

Table 5.1. Default Clusterwide Logical Name Tables and Logical Names

Name	Purpose
LNM\$SYSCLUSTER_TABLE	The default table for clusterwide system logical names. It is empty when shipped. This table is provided for system managers who want to use clusterwide logical names to customize their environments. The names in this table are available to anyone translating a logical name using SHOW LOGICAL/SYSTEM, specifying a table name of LNM\$SYSTEM, or LNM\$DCL_LOGICAL (DCL's default table search list), or LNM\$FILE_DEV (system and RMS default).
LNM\$SYSCLUSTER	The logical name for LNM\$SYSCLUSTER_TABLE. It is provided for convenience in referencing LNM\$SYSCLUSTER_TABLE. It is consistent in format with LNM\$SYSTEM_TABLE and its logical name, LNM\$SYSTEM.

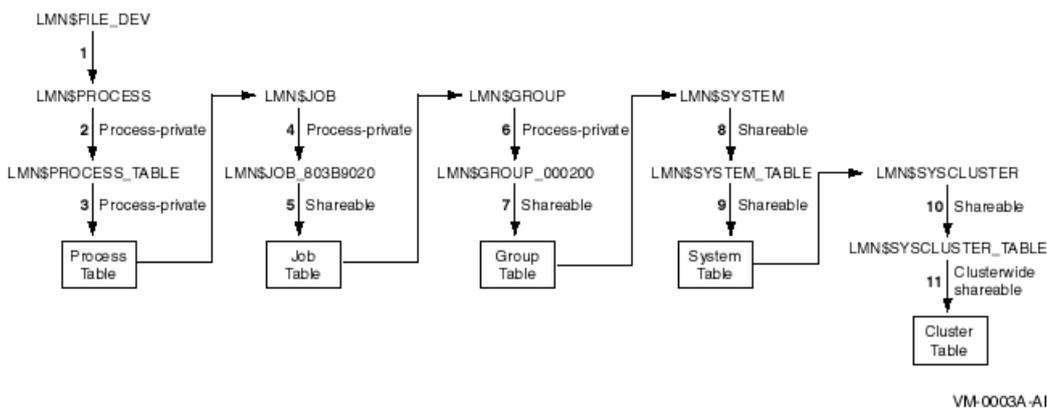
Name	Purpose
LNMS\$CLUSTER_TABLE	The parent table for all clusterwide logical name tables, including LNM\$SYSCLUSTER_TABLE. When you create a new table using LNM\$CLUSTER_TABLE as the parent table, the new table will be available clusterwide.
LNMS\$CLUSTER	The logical name for LNM\$CLUSTER_TABLE. It is provided for convenience in referencing LNM\$CLUSTER_TABLE.

5.4.2. Translation Order

The definition of LNM\$SYSTEM has been expanded to include LNM\$SYSCLUSTER. When a system logical name is translated, the search order is LNM\$SYSTEM_TABLE, LNM\$SYSCLUSTER_TABLE. Because the definitions for the system default table names, LNM\$FILE_DEV and LNM\$DCL_LOGICALS, include LNM\$SYSTEM, translations using those default tables include definitions in LNM\$SYSCLUSTER.

The current precedence order for resolving logical names is preserved. Clusterwide logical names that are translated against LNM\$FILE_DEV are resolved last, after system logical names. The precedence order, from first to last, is process → job → group → system → cluster, as shown in Figure 5.5.

Figure 5.5. Translation Order Specified by LNM\$FILE_DEV



5.4.3. Creating Clusterwide Logical Name Tables

You might want to create additional clusterwide logical name tables for the following purposes:

- For a multiprocess clusterwide application to use
- For members of a UIC group to share

To create a clusterwide logical name table, you must have create (C) access to the parent table and write (W) access to LNM\$SYSTEM_DIRECTORY, or the SYSPRV(system) privilege.

A shareable logical name table has UIC-based protection. Each class of user(system (S), owner (O), group (G), and world (W)) can be granted four types of access: read (R), write (W), create (C), or delete (D).

You can create additional clusterwide logical name tables in the same way that you can create additional process, job, and group logical name tables – with the CREATE/NAME_TABLE command

or with the \$CRELNT system service. When creating a clusterwide logical name table, you must specify the /PARENT_TABLE qualifier and provide a value for the qualifier that is a clusterwide table name. Any existing clusterwide table used as the parent table will make the new table clusterwide.

The following example shows how to create a clusterwide logical name table:

```
$ CREATE/NAME_TABLE/PARENT_TABLE=LNMS$CLUSTER_TABLE -
_$ new-clusterwide-logical-name-table
```

5.4.4. Alias Collisions Involving Clusterwide Logical Name Tables

Alias collisions involving clusterwide logical name tables are treated differently from alias collisions of other types of logical name tables. Table 5.2 describes the types of collisions and their outcomes.

Table 5.2. Alias Collisions and Outcomes

Collision Type	Outcome
Creating a local table with same name and access mode as an existing clusterwide table	New local table is not created. The condition value <code>SS\$_NORMAL</code> is returned, which means that the service completed successfully but the logical name table already exists. The existing clusterwide table and its names on all nodes remain in effect.
Creating a clusterwide table with same name and access mode as an existing local table	New clusterwide table is created. The condition value <code>SS\$_LNMCREATED</code> is returned, which means that the logical name table was created. The local table and its names are deleted. If the clusterwide table was created with the DCL command <code>DEFINE</code> , a message is displayed: <code>DCL-I-TABSUPER, previous table <i>table_name</i> has been superseded</code> If the clusterwide table was created with the \$CRELNT system service, \$CRELNT returns the condition value: <code>SS\$_SUPERSEDE</code> .
Creating a clusterwide table with same name and access mode as an existing clusterwide table	New clusterwide table is not created. The condition value <code>SS\$_NORMAL</code> is returned, which means that the service completed successfully but the logical name table already exists. The existing table and all its names remain in effect, regardless of the setting of the \$CRELNT system service's <code>CREATE-IF</code> attribute. This prevents surprise implicit deletions of existing table names from other nodes.

5.4.5. Creating Clusterwide Logical Names

To create a clusterwide logical name, you must have write (W) access to the table in which the logical name is to be entered, or `SYSNAM` privilege if you are creating clusterwide logical names only in `LNMS$SYSCLUSTER`. Unless you specify an access mode (user, supervisor, and so on), the access mode of the logical name you create defaults to the access mode from which the name was created. If you created the name with a DCL command, the access mode defaults to supervisor mode. If you created the name with a program, the access mode typically defaults to user mode.

When you create a clusterwide logical name, you must include the name of a clusterwide logical name table in the definition of the logical name. You can create clusterwide logical names by using DCL commands or with the \$CRELNM system service.

The following example shows how to create a clusterwide logical name in the default clusterwide logical name table, LNM\$CLUSTER_TABLE, using the DEFINE command:

```
$ DEFINE/TABLE=LNMS$CLUSTER_TABLE logical-name equivalence-string
```

To create clusterwide logical names that will reside in a clusterwide logical name table you created, you define the new clusterwide logical name with the DEFINE command, specifying your new clusterwide table's name with the /TABLE qualifier, as shown in the following example:

```
$ DEFINE/TABLE=new-clusterwide-logical-name-table logical-name -  
_ $ equivalence-string
```

Note

If you attempt to create a new clusterwide logical name with the same access mode and identical equivalence names and attributes as an existing clusterwide logical name, the existing name is *not* deleted, and no messages are sent to remote nodes. This behavior differs from similar attempts for other types of logical names, which delete the existing name and create the new one. For clusterwide logical names, this difference is a performance enhancement.

The condition value SS\$_NORMAL is returned. The service completed successfully, but the new logical name was not created.

5.4.6. Management Guidelines

When using clusterwide logical names, observe the following guidelines:

1. Do not use certain logical names clusterwide.

The following logical names are not valid for clusterwide use:

- Mailbox names, because mailbox devices are local to a node.
- SYSS\$NODE and SYSS\$NODE_FULLNAME must be in LNM\$SYSTEM_TABLE and are node specific.
- LMF\$LICENSE_TABLE.

2. Do not redefine LNM\$SYSTEM.

LNM\$SYSTEM is now defined as LNM\$SYSTEM_TABLE, LNM\$SYSCLUSTER_TABLE. Do not reverse the order of these two tables. If you do, then any names created using the /SYSTEM qualifier or in LNM\$SYSTEM would go in LNM\$SYSCLUSTER_TABLE and be clusterwide. Various system failures would result. For example, the MOUNT/SYSTEM command would attempt to create a clusterwide logical name for a mounted volume, which would result in an error.

3. Keep LNM\$SYSTEM contents in LNM\$SYSTEM.

Do not merge the logical names in LNM\$SYSTEM into LNM\$SYSCLUSTER. Many system logical names in LNM\$SYSTEM contain system roots and either node-specific devices, or node-specific directories, or both.

4. Adopt naming conventions for logical names used at your site.

To avoid confusion and name conflicts, develop one naming convention for system-specific logical names and another for clusterwide logical names.

5. Avoid using the dollar sign (\$) in your own site's logical names, because OpenVMS software uses it in its names.
6. Be aware that clusterwide logical name operations will stall when the clusterwide logical name database is not consistent.

This can occur during system initialization when the system's clusterwide logical name database is not completely initialized. It can also occur when the cluster server process has not finished updating the clusterwide logical name database, or during resynchronization after nodes enter or leave the cluster. As soon as consistency is reestablished, the processing of clusterwide logical name operations resumes.

5.4.7. Using Clusterwide Logical Names in Applications

The \$TRNLNM system service and the \$GETSYI system service provide attributes that are specific to clusterwide logical names. This section describes those attributes. It also describes the use of \$CRELNT as it pertains to creating a clusterwide table. For more information about using logical names in applications, refer to the *VSI OpenVMS Programming Concepts Manual*.

5.4.7.1. Clusterwide Attributes for \$TRNLNM System Service

Two clusterwide attributes are available in the \$TRNLNM system service:

- LNM\$V_CLUSTERWIDE
- LNM\$M_INTERLOCKED

LNM\$V_CLUSTERWIDE is an output attribute to be returned in the item list if you asked for the LNM\$_ATTRIBUTES item for a logical name that is clusterwide.

LNM\$M_INTERLOCKED is an **attr** argument bit that can be set to ensure that any clusterwide logical name modifications in progress are completed before the name is translated. LNM\$M_INTERLOCKED is not set by default. If your application requires translation using the most recent definition of a clusterwide logical name, use this attribute to ensure that the translation is stalled until all pending modifications have been made.

On a single system, when one process modifies the shareable part of the logical name database, the change is visible immediately to other processes on that node. Moreover, while the modification is in progress, no other process can translate or modify shareable logical names.

In contrast, when one process modifies the clusterwide logical name database, the change is visible immediately on that node, but it takes a short time for the change to be propagated to other nodes. By default, translations of clusterwide logical names are not stalled. Therefore, it is possible for processes on different nodes to translate a logical name and get different equivalence names when modifications are in progress.

The use of LNM\$M_INTERLOCKED guarantees that your application will receive the most recent definition of a clusterwide logical name.

5.4.7.2. Clusterwide Attribute for \$GETSYI System Service

The clusterwide attribute, SYI\$_CWLOGICALS, has been added to the \$GETSYI system service. When you specify SYI\$_CWLOGICALS, \$GETSYI returns the value 1 if the clusterwide logical name database has been initialized on the CPU, or the value 0 if it has not been initialized. Because

this number is a Boolean value (1 or 0), the buffer length field in the item descriptor should specify 1 (byte). On a nonclustered system, the value of SYIS_CWLOGICALS is always 0.

5.4.7.3. Creating Clusterwide Tables with the \$CRELNT System Service

When creating a clusterwide table, the \$CRELNT requester must supply a table name. OpenVMS does not supply a default name for clusterwide tables because the use of default names enables a process without the SYSPRV privilege to create a shareable table.

5.4.8. Defining and Accessing Clusterwide Logical Names

Initializing the clusterwide logical name database on a booting node requires sending a message to another node and having its CLUSTER_SERVER process reply with one or messages containing a description of the database. The CLUSTER_SERVER process on the booting node requests system services to create the equivalent names and tables. How long this initialization takes varies with conditions such as the size of the clusterwide logical name database, the speed of the cluster interconnect, and the responsiveness of the CLUSTER_SERVER process on the responding node.

Until a booting node's copy of the clusterwide logical name database is consistent with the logical name databases of the rest of the cluster, any attempt on the booting node to create or delete clusterwide names or tables is stalled transparently. Because translations are not stalled by default, any attempt to translate a clusterwide name before the database is consistent may fail or succeed, depending on timing. To stall a translation until the database is consistent, specify the F\$TRNLNM CASE argument as INTERLOCKED.

5.4.8.1. Defining Clusterwide Logical Names in SYSTARTUP_VMS.COM

In general, system managers edit the SYLOGICALS.COM command procedure to define site-specific logical names that take effect at system startup. However, VSI recommends that, if possible, clusterwide logical names be defined in the SYSTARTUP_VMS.COM command procedure instead with the exception of those logical names discussed in Section 5.4.8.2. The reason for defining clusterwide logical names in SYSTARTUP_VMS.COM is that SYSTARTUP_VMS.COM is run at a much later stage in the booting process than SYLOGICALS.COM.

OpenVMS startup is single streamed and synchronous except for actions taken by created processes, such as the CLUSTER_SERVER process. Although the CLUSTER_SERVER process is created very early in startup, it is possible that when SYLOGICALS.COM is executed, the booting node's copy of the clusterwide logical name database has not been fully initialized. In such a case, a clusterwide definition in SYLOGICALS.COM would stall startup and increase the time it takes for the system to become operational.

OpenVMS will ensure that the clusterwide database has been initialized before SYSTARTUP_VMS.COM is executed.

5.4.8.2. Defining Certain Logical Names in SYLOGICALS.COM

To be effective, certain logical names, such as LMF\$LICENSE, NET\$PROXY, and VMS\$OBJECTS must be defined earlier in startup than when SYSTARTUP_VMS.COM is invoked. Most such names are defined in SYLOGICALS.COM, with the exception of VMS\$OBJECTS, which is defined in SYSECURITY.COM, and any names defined in SYCONFIG.COM.

Although VSI recommends defining clusterwide logical names in SYSTARTUP_VMS.COM, to define these names to be clusterwide, you must do so in SYLOGICALS.COM or SYSECURITY.COM. Note that doing this may increase startup time.

Alternatively, you can take the traditional approach and define these names as systemwide logical names with the same definition on every node.

5.4.8.3. Using Conditional Definitions for Startup Command Procedures

For clusterwide definitions in any startup command procedure that is common to all cluster nodes, VSI recommends that you use a conditional definition. For example:

```
$ IF F$STRNLNM("CLUSTER_APPS") .EQS. "" THEN -
_ $ DEFINE/TABLE=LNM$SYSCLUSTER/EXEC CLUSTER_APPS -
_ $ $1$DKA500:[COMMON_APPS]
```

A conditional definition can prevent unpleasant surprises. For example, suppose a system manager redefines a name that is also defined in SYSTARTUP_VMS.COM but does not edit SYSTARTUP_VMS.COM because the new definition is temporary. If a new node joins the cluster, the new node would initially receive the new definition. However, when the new node executes SYSTARTUP_VMS.COM, it will cause all the nodes in the cluster, including itself, to revert to the original value.

If you include a conditional definition in SYLOGICALS.COM or SYSECURITY.COM, specify the F\$STRNLNM CASE argument as INTERLOCKED to ensure that clusterwide logical names have been fully initialized before the translation completes. An example of a conditional definition with the argument specified follows:

```
$ IF F$STRNLNM("CLUSTER_APPS",,,, "INTERLOCKED") .EQS.
"" THEN - _ $ DEFINE/TABLE=LNM$SYSCLUSTER/EXEC CLUSTER_APPS -
_ $ $1$DKA500:[COMMON_APPS]
```

Note

F\$GETSYI("CWLOGICALS") always returns a value of FALSE on a noncluster system. Procedures that are designed to run in both clustered and nonclustered environments should first determine whether they are in a cluster and, if so, then determine whether clusterwide logical names are initialized.

5.4.9. Displaying Clusterwide Logical Names

The /CLUSTER qualifier was added to the SHOW LOGICAL DCL command in OpenVMS Version 8.2. When the SHOW LOGICAL/CLUSTER command is specified, all clusterwide logical names are displayed, as shown in the following example:

```
$ SHOW LOGICAL/CLUSTER

(LNM$CLUSTER_TABLE)

(LNM$SYSCLUSTER_TABLE)

"MSCPMOUNT$_AMALFI_LAST" = "2005-10-10 14:25:03.74"
"MSCPMOUNT$_AMALFI_LOGINTIM" = " 8-OCT-2005 01:02:22.17"
"MSCPMOUNT$_AMALFI_NEXT" = "2005-10-10 14:40:03.74"
"MSCPMOUNT$_AMALFI_PID" = "26200462"
```

```

.
.
.
"MSCPMOUNT$_ETNA_LAST" = "2005-10-10 14:25:18.78"
"MSCPMOUNT$_ETNA_LOGINTIM" = " 8-OCT-2005 07:44:37.89"
"MSCPMOUNT$_ETNA_NEXT" = "2005-10-10 14:40:18.79"
"MSCPMOUNT$_ETNA_PID" = "26A0044E"
.
.
.
"MSCPMOUNT$_MILAN_LAST" = "2005-10-10 14:25:19.64"
"MSCPMOUNT$_MILAN_LOGINTIM" = " 8-OCT-2005 07:22:08.05"
"MSCPMOUNT$_MILAN_NEXT" = "2005-10-10 14:40:19.64"
"MSCPMOUNT$_MILAN_PID" = "26600458"
.
.
.
"MSCPMOUNT$_ORVIET_LAST" = "2005-10-10 14:29:25.94"
"MSCPMOUNT$_ORVIET_LOGINTIM" = "30-SEP-2005 09:38:27.38"
"MSCPMOUNT$_ORVIET_NEXT" = "2005-10-10 14:44:26.61"
"MSCPMOUNT$_ORVIET_PID" = "25600139"
.
.
.
"MSCPMOUNT$_TURIN_LAST" = "2005-10-10 14:39:59.59"
"MSCPMOUNT$_TURIN_LOGINTIM" = "10-OCT-2005 09:22:48.46"
"MSCPMOUNT$_TURIN_NEXT" = "2005-10-10 14:54:59.59"
"MSCPMOUNT$_TURIN_PID" = "2760012C"
"PREPOPULATE_NEXT_STREAM$IGNORE_BUILD_MASTER_944" = "1"

(CLU$ICC_ORBS_AMALFI)

"ICC$ORB_ICC$PID_26200450_U" = "T"
= "M\.v....k.....æ...æ...þ...þ....AMALFI::ICC
$PID_26200450_U....."
"ICC$ORB_REG$SERVER_E" = "T"
= "p.O<....e.....æ...æ...þ...þ....AMALFI::REG
$SERVER_E044....."
"ICC$ORB_REG$SERVER_K" = "T"
= "p.O<....e.....æ...æ...þ...þ....AMALFI::REG
$SERVER_K044....."
"ICC$ORB_REG$SERVER_U" = "T"
= "p.O<....e.....æ...æ...þ...þ....AMALFI::REG
$SERVER_U044....."

(CLU$ICC_ORBS_ETNA)

(CLU$ICC_ORBS_MILAN)

(CLU$ICC_ORBS_ORVIET)

"ICC$ORB_ICC$PID_26000450_U" = "T"
= "VQ.p....k.....æ...æ...þ...þ....ETNA::ICC
$PID_26000450_U....."

(CLU$ICC_ORBS_TURIN)
.

```

```
.
.
( ICC$REGISTRY_TABLE )
```

5.5. Coordinating Startup Command Procedures

Immediately after a computer boots, it runs the site-independent command procedure SYS\$SYSTEM:STARTUP.COM to start up the system and control the sequence of startup events. The STARTUP.COM procedure calls a number of other startup command procedures that perform cluster-specific and node-specific tasks.

The following sections describe how, by setting up appropriate cluster-specific startup command procedures and other system files, you can prepare the OpenVMS Cluster operating environment on the first installed computer before adding other computers to the cluster.

Reference: See also the *VSI OpenVMS System Manager's Manual* for more information about startup command procedures.

5.5.1. OpenVMS Startup Procedures

Several startup command procedures are distributed as part of the OpenVMS operating system. The SYSS\$SYSTEM:STARTUP.COM command procedure executes immediately after OpenVMS is booted and invokes the site-specific startup command procedures described in the following table.

Procedure Name	Invoked by	Function
SYSS\$MANAGER: YPAGSWPFILES.COM	SYSS\$SYSTEM: STARTUP.COM	A file to which you add commands to install page and swap files (other than the primary page and swap files that are installed automatically).
SYSS\$MANAGER: SYCONFIG.COM	SYSS\$SYSTEM: STARTUP.COM	Connects special devices and loads device I/O drivers.
SYSS\$MANAGER: SYSECURITY.COM	SYSS\$SYSTEM: STARTUP.COM	Defines the location of the security audit and archive files before it starts the security audit server.
SYSS\$MANAGER: SYLOGICALS.COM	SYSS\$SYSTEM: STARTUP.COM	Creates systemwide logical names, and defines system components as executive-mode logical names. (Clusterwide logical names should be defined in SYSTARTUP_VMS.COM.) Cluster common disks can be mounted at the end of this procedure.
SYSS\$MANAGER: SYSTARTUP_VMS.COM	SYSS\$SYSTEM: STARTUP.COM	Performs many of the following startup and login functions: <ul style="list-style-type: none"> • Mounts all volumes except the system disk. • Sets device characteristics. • Defines clusterwide logical names • Initializes and starts batch and print queues.

Procedure Name	Invoked by	Function
		<ul style="list-style-type: none"> • Installs known images. • Starts layered products. • Starts the DECnet software. • Analyzes most recent system failure. • Purges old operator log files. • Starts the LAT network (if used). • Defines the maximum number of interactive users. • Announces that the system is up and running. • Allows users to log in.

The directory SYSSCOMMON:[SYSMGR] contains a template file for each command procedure that you can edit. Use the command procedure templates (in SYSSCOMMON:[SYSMGR]*.TEMPLATE) as examples for customization of your system's startup and login characteristics.

5.5.2. Building Startup Procedures

The first step in preparing an OpenVMS Cluster shared environment is to build a SYSTARTUP_VMS command procedure. Each computer executes the procedure at startup time to define the operating environment.

Prepare the SYSTARTUP_VMS.COM procedure as follows:

Step	Action
1	<p>In each computer's SYSSSPECIFIC:[SYSMGR] directory, edit the SYSTARTUP_VMS.TEMPLATE file to set up a SYSTARTUP_VMS.COM procedure that:</p> <ul style="list-style-type: none"> • Performs computer-specific startup functions such as the following: <ul style="list-style-type: none"> • Setting up dual-ported and local disks • Loading device drivers • Setting up local terminals and terminal server access • Invoking the common startup procedure (described next).
2	<p>Build a common command procedure that includes startup commands that you want to be common to all computers. The common procedure might contain commands that:</p> <ul style="list-style-type: none"> • Install images • Define logical names • Set up queues

Step	Action
	<ul style="list-style-type: none"> • Set up and mount physically accessible mass storage devices • Perform any other common startup functions <p>Note: You might choose to build these commands into individual command procedures that are invoked from the common procedure. For example, the MSCPMOUNT.COM file in the SYS\$EXAMPLES directory is a sample common command procedure that contains commands typically used to mount cluster disks. The example includes comments explaining each phase of the procedure.</p>
3	<p>Place the common procedure in the SYSS\$COMMON:[SYSMGR] directory on a common system disk or other cluster-accessible disk.</p> <p>Important: The common procedure is usually located in the SYSS\$COMMON:[SYSMGR] directory on a common system disk but can reside on any disk, provided that the disk is cluster accessible and is mounted when the procedure is invoked. If you create a copy of the common procedure for each computer, you must remember to update each copy whenever you make changes.</p>

5.5.3. Combining Existing Procedures

To build startup procedures for an OpenVMS Cluster system in which existing computers are to be combined, you should compare both the computer-specific SYSTARTUP_VMS and the common startup command procedures on each computer and make any adjustments required. For example, you can compare the procedures from each computer and include commands that define the same logical names in your common SYSTARTUP_VMS command procedure.

After you have chosen which commands to make common, you can build the common procedures on one of the OpenVMS Cluster computers.

5.5.4. Using Multiple Startup Procedures

To define a multiple-environment cluster, you set up computer-specific versions of one or more system files. For example, if you want to give users larger working set quotas on URANUS, you would create a computer-specific version of SYSUAF.DAT and place that file in system's root directory. That directory can be located in URANUS's root on a common system disk or on an individual system disk that you have set up on URANUS.

Follow these steps to build SYSTARTUP and SYLOGIN command files for a multiple-environment OpenVMS Cluster:

Step	Action
1	Include in SYSTARTUP_VMS.COM elements that you want to remain unique to a computer, such as commands to define computer-specific logical names and symbols.
2	Place these files in the SYSS\$SPECIFIC root on each computer.

Example: Consider a three-member cluster consisting of computers JUPITR, SATURN, and PLUTO. The time sharing environments on JUPITR and SATURN are the same. However, PLUTO runs applications for a specific user group. In this cluster, you would create a common SYSTARTUP_VMS command procedure for JUPITR and SATURN that defines identical environments on these computers. But the command procedure for PLUTO would be different; it would include commands to define PLUTO's special application environment.

5.6. Providing OpenVMS Cluster System Security

The OpenVMS security subsystem ensures that all authorization information and object security profiles are consistent across all nodes in the cluster. The OpenVMS operating system does not support multiple security domains because the operating system cannot enforce a level of separation needed to support different security domains on separate cluster members.

5.6.1. Security Checks

In an OpenVMS Cluster system, individual nodes use a common set of authorizations to mediate access control that, in effect, ensures that a security check results in the same answer from any node in the cluster. The following list outlines how the OpenVMS operating system provides a basic level of protection:

- Authorized users can have processes executing on any OpenVMS Cluster member.
- A process, acting on behalf of an authorized individual, requests access to a cluster object.
- A coordinating node determines the outcome by comparing its copy of the common authorization database with the security profile for the object being accessed.

The OpenVMS operating system provides the same strategy for the protection of files and queues, and further incorporates all other cluster-visible objects, such as devices, volumes, and lock resource domains.

Starting with OpenVMS Version 7.3, the operating system provides clusterwide intrusion detection, which extends protection against attacks of all types throughout the cluster. The intrusion data and information from each system is integrated to protect the cluster as a whole. Prior to Version 7.3, each system was protected individually.

The SECURITY_POLICY system parameter controls whether a local or a clusterwide intrusion database is maintained for each system. The default setting is for a clusterwide database, which contains all unauthorized attempts and the state of any intrusion events for all cluster members that are using this setting. Cluster members using the clusterwide intrusion database are made aware if a cluster member is under attack or has any intrusion events recorded. Events recorded on one system can cause another system in the cluster to take restrictive action. (For example, the person attempting to log in is monitored more closely and limited to a certain number of login retries within a limited period of time. Once a person exceeds either the retry or time limitation, he or she cannot log in).

Actions of the cluster manager in setting up an OpenVMS Cluster system can affect the security operations of the system. You can facilitate OpenVMS Cluster security management using the suggestions discussed in the following sections.

The easiest way to ensure a single security domain is to maintain a single copy of each of the following files on one or more disks that are accessible from anywhere in the OpenVMS Cluster system. When a cluster is configured with multiple system disks, you can use system logical names (as shown in Section 5.8) to ensure that only a single copy of each file exists.

```
SY$MANAGER:VM$AUDIT_SERVER.DAT
SY$SYSTEM:NETOBJECT.DAT
SY$SYSTEM:NETPROXY.DAT
TCPIP$PROXY.DAT
SY$SYSTEM:PE$IP_CONFIG.DAT
```

```

SYSS$SYSTEM:QMAN$MASTER.DAT
SYSS$SYSTEM:RIGHTSLIST.DAT
SYSS$SYSTEM:SYSALF.DAT
SYSS$SYSTEM:SYSUAF.DAT
SYSS$SYSTEM:SYSUAFALT.DAT
SYSS$SYSTEM:VMSPASSWORD_HISTORY.DATA
SYSS$SYSTEM:VMSMAIL_PROFILE.DATA
SYSS$LIBRARY:VMSPASSWORD_DICTIONARY.DATA
SYSS$LIBRARY:VMSPASSWORD_POLICY.EXE

```

Note: Using shared files is not the only way of achieving a single security domain. You may need to use multiple copies of one or more of these files on different nodes in a cluster. For example, on Alpha nodes you may choose to deploy system-specific user authorization files (SYSUAFs) to allow for different memory management working-set quotas among different nodes. Such configurations are fully supported as long as the security information available to each node in the cluster is identical.

5.6.2. Files Relevant to OpenVMS Cluster Security

Table 5.3 describes the security-relevant portions of the files that must be common across all cluster members to ensure that a single security domain exists.

Notes:

- Some of these files are created only on request and may not exist in all configurations.
- A file can be absent on one node only if it is absent on all nodes.
- As soon as a required file is created on one node, it must be created or commonly referenced on all remaining cluster nodes.

The following table describes designations for the files in Table 5.3.

Table Keyword	Meaning
Required	The file contains some data that must be kept common across all cluster members to ensure that a single security environment exists.
Recommended	The file contains data that should be kept common at the discretion of the site security administrator or system manager. Nonetheless, VSI recommends that you synchronize the recommended files.

Table 5.3. Security Files

File Name	Contains
CLUSTER_AUTHORIZE.DAT	The cluster authorization file, SYSS\$COMMON:[SYSEXE]CLUSTER_AUTHORIZE.DAT, contains the cluster group number in a disorderly form and the cluster password. The CLUSTER_AUTHORIZE.DAT file is accessible only to users with the SYSPRV privilege.
PE\$IP_CONFIG.DAT [recommended]	For cluster over IP configurations, which are using IP unicast, the remote node IP address should be present in the existing cluster members file in the SYSS\$SYSTEM:PE\$IP_CONFIG.DAT file. Remote nodes in a different IP multicast domain can use the IP unicast messaging technique to join the Cluster.

File Name	Contains
VM\$AUDIT_SERVER.DAT [recommended]	<p>Information related to security auditing. Among the information contained is the list of enabled security auditing events and the destination of the system security audit journal file. When more than one copy of this file exists, all copies should be update dafter any SET AUDIT command.</p> <p>OpenVMS Cluster system managers should ensure that the name assigned to the security audit journal file resolves to the following location:</p> <pre>SY\$COMMON: [SY\$MGR] SECURITY . AUDIT\$JOURNAL</pre> <p>Rule: If you need to relocate the audit journal file somewhere other than the system disk (or if you have multiple system disks), you should redirect the audit journal uniformly across all nodes in the cluster. Use the command SETAUDIT/JOURNAL=SECURITY/DESTINATION=<i>file-name</i>, specifying a file name that resolves to the same file throughout the cluster.</p> <p>Changes are automatically made in the audit server database, SY\$MANAGER:VM\$AUDIT_SERVER.DAT. This database also identifies which events are enabled and how to monitor the audit system's use of resources, and restores audit system settings each time the system is rebooted.</p> <p>Caution: Failure to synchronize multiple copies of this file properly may result in partitioned auditing domains.</p> <p>Reference: For more information, see the <i>VSI OpenVMS Guide to System Security</i>.</p>
NETOBJECT.DAT [required]	<p>The DECnet object database. Among the information contained in this file is the list of known DECnet server accounts and passwords. When more than one copy of this file exists, all copies must be updated after every use of the NCP commands SET OBJECT or DEFINE OBJECT.</p> <p>Caution: Failure to synchronize multiple copies of this file properly may result in unexplained network login failures and unauthorized network access. For instructions on maintaining a single copy, refer to Section 5.8.1.</p> <p>Reference: Refer to the DECnet-Plus documentation for equivalent NCL command information.</p>
NETPROXY.DAT and NET\$PROXY.DAT [required]	<p>The network proxy database. It is maintained by the OpenVMS Authorize utility. When more than one copy of this file exists, all copies must be updated after any UAF proxy command.</p> <p>Note: The NET\$PROXY.DAT and NETPROXY.DAT files are equivalent; NET\$PROXY.DAT is for DECnet-Plus</p>

File Name	Contains
	<p>implementations and NETPROXY.DAT is for DECnet for OpenVMS implementations.</p> <p>Caution: Failure to synchronize multiple copies of this file properly may result in unexplained network login failures and unauthorized network access. For instructions on maintaining a single copy, refer to Section 5.8.1.</p> <p>Reference: Appendix B discusses how to consolidate several NETPROXY.DAT and RIGHTSLIST.DAT files.</p>
TCPIP\$PROXY.DAT	<p>This database provides OpenVMS identities for remote NFS clients and UNIX-style identifiers for local NFS client users; provides proxy accounts for remote processes. For more information about this file, see the <i>VSI TCP/IP Services for OpenVMS Management Manual</i>.</p>
QMAN\$MASTER.DAT [required]	<p>The master queue manager database. This file contains the security information for all shared batch and print queues.</p> <p>Rule: If two or more nodes are to participate in a shared queuing system, a single copy of this file must be maintained on a shared disk. For instructions on maintaining a single copy, refer to Section 5.8.1.</p>
RIGHTSLIST.DAT [required]	<p>The rights identifier database. It is maintained by the OpenVMS Authorize utility and by various rights identifier system services. When more than one copy of this file exists, all copies must be updated after any change to any identifier or holder records.</p> <p>Caution: Failure to synchronize multiple copies of this file properly may result in unauthorized system access and unauthorized access to protected objects. For instructions on maintaining a single copy, refer to Section 5.8.1.</p> <p>Reference: Appendix B discusses how to consolidate several NETPROXY.DAT and RIGHTSLIST.DAT files.</p>
SYSALF.DAT [required]	<p>The system Auto login facility database. It is maintained by the OpenVMS SYSMAN utility. When more than one copy of this file exists, all copies must be updated after any SYSMAN ALF command.</p> <p>Note: This file may not exist in all configurations.</p> <p>Caution: Failure to synchronize multiple copies of this file properly may result in unexplained login failures and unauthorized system access. For instructions on maintaining a single copy, refer to Section 5.8.1.</p>
SYSUAF.DAT [required]	<p>The system user authorization file. It is maintained by the OpenVMS Authorize utility and is modifiable via the \$SETUAI system service. When more than one copy of this file exists, you must ensure that the SYSUAF and</p>

File Name	Contains	
	associated \$SETUAI item codes are synchronized for each user record. The following table shows the fields in SYSUAF and their associated \$SETUAI item codes.	
	Internal Field Name	\$SETUAI Item Code
	UAF\$R_DEF_CLASS	UAI\$_DEF_CLASS
	UAF\$Q_DEF_PRIV	UAI\$_DEF_PRIV
	UAF \$B_DIALUP_ACCESS_P	UAI\$_LOCAL_ACCESS_P
	UAF \$B_LOCAL_ACCESS_S	UAI\$_DIALUP_ACCESS_S
	UAF\$B_ENCRYPT	UAI\$_ENCRYPT
	UAF\$B_ENCRYPT2	UAI\$_ENCRYPT2
	UAF\$Q_EXPIRATION	UAI\$_EXPIRATION
	UAF\$L_FLAGS	UAI\$_FLAGS
	UAF \$B_LOCAL_ACCESS_P	UAI\$_LOCAL_ACCESS_P
	UAF \$B_LOCAL_ACCESS_S	UAI\$_LOCAL_ACCESS_S
	UAF \$B_NETWORK_ACCESS_P	UAI \$_NETWORK_ACCESS_P
	UAF \$B_NETWORK_ACCESS_S	UAI \$_NETWORK_ACCESS_S
	UAF\$B_PRIME_DAYS	UAI\$_PRIMEDAYS
	UAF\$Q_PRIV	UAI\$_PRIV
	UAF\$Q_PWD	UAI\$_PWD
	UAF\$Q_PWD2	UAI\$_PWD2
	UAF\$Q_PWD_DATE	UAI\$_PWD_DATE
	UAF\$Q_PWD2_DATE	UAI\$_PWD2_DATE
	UAF\$B_PWD_LENGTH	UAI\$_PWD_LENGTH
	UAF\$Q_PWD_LIFETIME	UAI\$_PWD_LIFETIME
	UAF \$B_REMOTE_ACCESS_P	UAI \$_REMOTE_ACCESS_P
	UAF \$B_REMOTE_ACCESS_S	UAI \$_REMOTE_ACCESS_S
	UAF\$R_MAX_CLASS	UAI\$_MAX_CLASS
	UAF\$R_MIN_CLASS	UAI\$_MIN_CLASS
	UAF\$W_SALT	UAI\$_SALT
	UAF\$L_UIC	Not applicable
	Caution: Failure to synchronize multiple copies of the SYSUAF files properly may result in unexplained login	

File Name	Contains
	<p>failures and unauthorized system access. For instructions on maintaining a single copy, refer to Section 5.8.1.</p> <p>Reference: Appendix B discusses creation and management of the various elements of an OpenVMS Cluster common SYSUAF.DAT authorization database.</p>
SYSUAFALT.DAT [required]	<p>The system alternate user authorization file. This file serves as a backup to SYSUAF.DAT and is enabled via the SYSUAFALT system parameter. When more than one copy of this file exists, all copies must be updated after any change to any authorization records in this file.</p> <p>Note: This file may not exist in all configurations.</p> <p>Caution: Failure to synchronize multiple copies of this file properly may result in unexplained login failures and unauthorized system access.</p>
VMS\$PASSWORD_HISTORY.DATA [recommended]	<p>The system password history database. It is maintained by the system password change facility. When more than one copy of this file exists, all copies should be updated after any password change.</p> <p>Caution: Failure to synchronize multiple copies of this file properly may result in a violation of the system password policy.</p>
VMSMAIL_PROFILE.DATA [recommended]	<p>The system mail database. This file is maintained by the OpenVMS Mail utility and contains mail profiles for all system users. Among the information contained in this file is the list of all mail forwarding addresses in use on the system. When more than one copy of this file exists, all copies should be updated after any changes to mail forwarding.</p> <p>Caution: Failure to synchronize multiple copies of this file properly may result in unauthorized disclosure of information.</p>
VMS\$PASSWORD_DICTIONARY.DATA [recommended]	<p>The system password dictionary. The system password dictionary is a list of English language words and phrases that are not legal for use as account passwords. When more than one copy of this file exists, all copies should be updated after any site-specific additions.</p> <p>Caution: Failure to synchronize multiple copies of this file properly may result in a violation of the system password policy.</p>
VMS\$PASSWORD_POLICY.EXE [recommended]	<p>Any site-specific password filters. It is created and installed by the site-security administrator or system manager. When more than one copy of this file exists, all copies should be identical.</p>

File Name	Contains
	<p>Caution: Failure to synchronize multiple copies of this file properly may result in a violation of the system password policy.</p> <p>Note: System managers can create this file as an image to enforce their local password policy. This is an architecture-specific image file that cannot be shared among different architecture types.</p>

5.7. Network Security

Network security must promote interoperability and uniform security approaches throughout networks. The following list shows three major areas of network security:

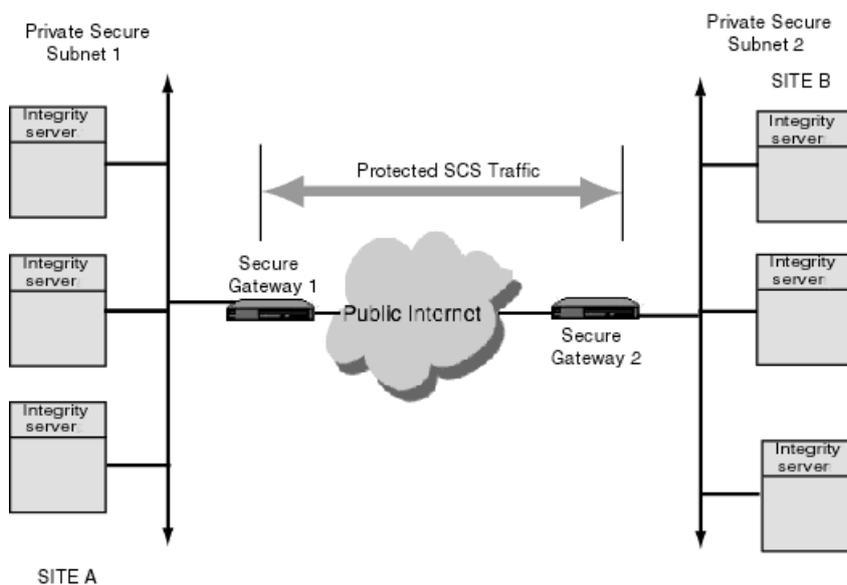
- User authentication

On Cluster systems connected using IP, ensure that the cluster communications over insecure WAN links are encrypted and authenticated.

- OpenVMS Cluster membership management

On Cluster systems connected using IP, isolate IP subnets that are used for cluster communication from the public internet using a secure gateway as shown in Figure 5.6.

Figure 5.6. Virtual Private Network for Protecting Cluster Traffic



- Using a security audit log file

OpenVMS Cluster system managers must also ensure consistency in the use of DECnet software for intracluster communication.

5.7.1. Mechanisms

Depending on the level of network security required, you might also want to consider how other security mechanisms, such as protocol encryption and decryption, can promote additional security protection across the cluster.

Reference: See the *VSI OpenVMS Guide to System Security*.

5.8. Coordinating System Files

Follow these guidelines to coordinate system files:

IF you are setting up...	THEN follow the procedures in...
A common-environment OpenVMS Cluster that consists of newly installed systems	<i>VSI OpenVMS System Manager's Manual</i> to build these files. Because the files on new operating systems are empty except for the Digital-supplied accounts, very little coordination is necessary.
An OpenVMS Cluster that will combine one or more computers that have been running with computer-specific files	Appendix B to create common copies of the files from the computer-specific files.

5.8.1. Procedure

In a common-environment cluster with one common system disk, you use a common copy of each system file and place the files in the SY\$COMMON:[SYSEXE] directory on the common system disk or on a disk that is mounted by all cluster nodes. No further action is required.

To prepare a common user environment for an OpenVMS Cluster system that includes more than one common OpenVMS Integrity server system disk or more than one common OpenVMS Alpha system disk, you must coordinate the system files on those disks.

Rules: The following rules apply to the procedures described in Table 5.4:

- Disks holding common resources must be mounted early in the system startup procedure, such as in the SYLOGICALS.COM procedure.
- You must ensure that the disks are mounted with each OpenVMS Cluster reboot.

Table 5.4. Procedure for Coordinating Files

Step	Action
1	Decide where to locate the SYSUAF.DAT and NETPROXY.DAT files. In a cluster with multiple system disks, system management is much easier if the common system files are located on a single disk that is not a system disk.
2	Copy SY\$SYSTEM:SYSUAF.DAT and SY\$SYSTEM:NETPROXY.DAT to a location other than the system disk.
3	Copy SY\$SYSTEM:RIGHTSLIST.DAT and SY\$SYSTEM:VMSMAIL_PROFILE.DATA to the same directory in which SYSUAF.DAT and NETPROXY.DAT reside.
4	Edit the file SY\$COMMON:[SYSMGR]SYLOGICALS.COM <i>on each system disk</i> and define logical names that specify the location of the cluster common files.

Step	Action
	<p>Example: If the files will be located on \$1\$DGA16, define logical names as follows:</p> <pre> \$ DEFINE/SYSTEM/EXEC SYSUAF - \$1\$DGA16:[VMS\$COMMON.SYSEXE]SYSUAF.DAT \$ DEFINE/SYSTEM/EXEC NETPROXY - \$1\$DGA16:[VMS\$COMMON.SYSEXE]NETPROXY.DAT \$ DEFINE/SYSTEM/EXEC RIGHTSLIST - \$1\$DGA16:[VMS\$COMMON.SYSEXE]RIGHTSLIST.DAT \$ DEFINE/SYSTEM/EXEC VMSMAIL_PROFILE - \$1\$DGA16:[VMS\$COMMON.SYSEXE]VMSMAIL_PROFILE.DATA \$ DEFINE/SYSTEM/EXEC NETNODE_REMOTE - \$1\$DGA16:[VMS\$COMMON.SYSEXE]NETNODE_REMOTE.DAT \$ DEFINE/SYSTEM/EXEC NETNODE_UPDATE - \$1\$DGA16:[VMS\$COMMON.SYSMGR]NETNODE_UPDATE.COM \$ DEFINE/SYSTEM/EXEC QMAN\$MASTER - \$1\$DGA16:[VMS\$COMMON.SYSEXE]</pre>
5	<p>To ensure that the system disks are mounted correctly with each reboot, follow these steps:</p> <ol style="list-style-type: none"> 1. Copy the SYS\$EXAMPLES:CLU_MOUNT_DISK.COM file to the [VMS \$COMMON.SYSMGR] directory, and edit it for your configuration. 2. Edit SYLOGICALS.COM and include commands to mount, with the appropriate volume label, the system disk containing the shared files. <p>Example: If the system disk is \$1\$DGA16, include the following command:</p> <pre> \$ @SYS\$SYSDEVICE:[VMS\$COMMON.SYSMGR]CLU_MOUNT_DISK.COM \$1\$DGA16: volume-label</pre>
6	<p>When you are ready to start the queuing system, be sure you have moved the queue and journal files to a cluster-available disk. Any cluster common disk is a good choice if the disk has sufficient space.</p> <p>Enter the following command:</p> <pre> \$ START/QUEUE/MANAGER \$1\$DGA16:[VMS\$COMMON.SYSEXE]</pre>

5.8.2. Network Database Files

In OpenVMS Cluster systems on the LAN and in mixed-interconnect clusters, you must also coordinate the SYS\$MANAGER:NETNODE_UPDATE.COM file, which is a file that contains all essential network configuration data for satellites. NETNODE_UPDATE.COM is updated each time you add or remove a satellite or change its Ethernet or FDDI hardware address. This file is discussed more thoroughly in Section 10.4.2.

In OpenVMS Cluster systems configured with DECnet for OpenVMS software, you must also coordinate NETNODE_REMOTE.DAT, which is the remote node network database.

5.9. System Time on the Cluster

When a computer joins the cluster, the cluster attempts to set the joining computer's system time to the current time on the cluster. Although it is likely that the system time will be similar on each cluster computer, there is no assurance that the time will be set. Also, no attempt is made to ensure that

the system times remain similar throughout the cluster. (For example, there is no protection against different computers having different clock rates).

An OpenVMS Cluster system spanning multiple time zones must use a single, clusterwide common time on all nodes. Use of a common time ensures timestamp consistency (for example, between applications, file-system instances) across the OpenVMS Cluster members.

5.9.1. Setting System Time

Use the SYSMAN command CONFIGURATION SET TIME to set the time across the cluster. This command issues warnings if the time on all nodes cannot be set within certain limits. Refer to the *VSI OpenVMS System Manager's Manual* for information about the SET TIME command.

Chapter 6. Cluster Storage Devices

One of the most important features of OpenVMS Cluster systems is the ability to provide access to devices and files across multiple systems.

In a traditional computing environment, a single system is directly attached to its storage subsystems. Even though the system may be networked with other systems, when the system is shut down, no other system on the network has access to its disks or any other devices attached to the system.

In an OpenVMS Cluster system, disks and tapes can be made accessible to one or more members. So, if one computer shuts down, the remaining computers still have access to the devices.

6.1. Data File Sharing

Cluster-accessible devices play a key role in OpenVMS Clusters because, when you place data files or applications on a cluster-accessible device, computer scan share a single copy of each common file. Data sharing is possible between Integrity server systems, between:

- Integrity servers
- Integrity servers and AlphaServer systems
- AlphaServer systems

In addition, multiple systems that are permitted in the same OpenVMS Cluster system can write to a shared disk file simultaneously. It is this ability that allows multiple systems in an OpenVMS Cluster to share a single system disk; multiple systems can boot from the same system disk and share operating system files and utilities to save disk space and simplify system management.

Note

Tapes do not allow multiple systems to access a tape file simultaneously.

6.1.1. Access Methods

Depending on your business needs, you may want to restrict access to a particular device to the users on the computer that are directly connected (local) to the device. Alternatively, you may decide to set up a disk or tape as a served device so that any user on any OpenVMS Cluster computer can allocate and use it.

Table 6.1 describes the various access methods.

Table 6.1. Device Access Methods

Method	Device Access	Comments	Illustrated in
Local	Restricted to the computer that is directly connected to the device.	Can be set up to be served to other systems.	Figure 6.3
Dual ported	Using either of two physical ports, each of which can	As long as one of the controllers is available, the	Figure 6.1

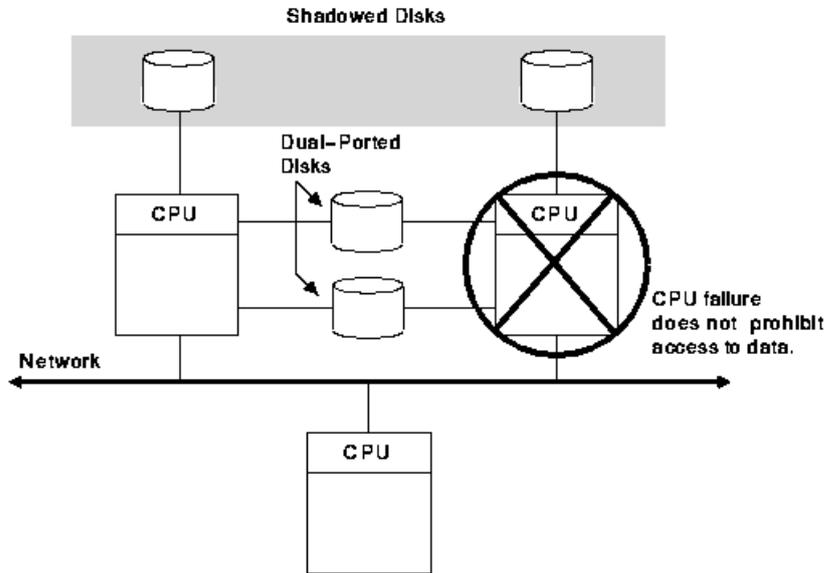
Method	Device Access	Comments	Illustrated in
	be connected to separate controllers. A dual-ported disk can survive the failure of a single controller by failing over to the other controller.	device is accessible by all systems in the cluster.	
Shared	Through a shared interconnect to multiple systems.	Can be set up to be served to systems that are not on the shared interconnect.	Figure 6.2
Served	Through a computer that has the MSCP or TMSCP server software loaded.	MSCP and TMSCP serving are discussed in Section 6.3.	Figures 6.2 and 6.3
Dual pathed	Possible through more than one path.	If one path fails, the device is accessed over the other path. Requires the use of allocation classes (described in Section 6.2.1 to provide a unique, path-independent name.)	Figure 6.2
Note: The path to an individual disk may appear to be local from some nodes and served from others.			

6.1.2. Examples

When storage subsystems are connected directly to a specific system, the availability of the subsystem is lower due to the reliance on the host system. To increase the availability of these configurations, OpenVMS Cluster systems support dual porting, dual pathing, and MSCP and TMSCP serving.

Figure 6.1 shows a dual-ported configuration, in which the disks have independent connections to two separate computers. As long as one of the computers is available, the disk is accessible by the other systems in the cluster.

Figure 6.1. Dual-Ported Disks



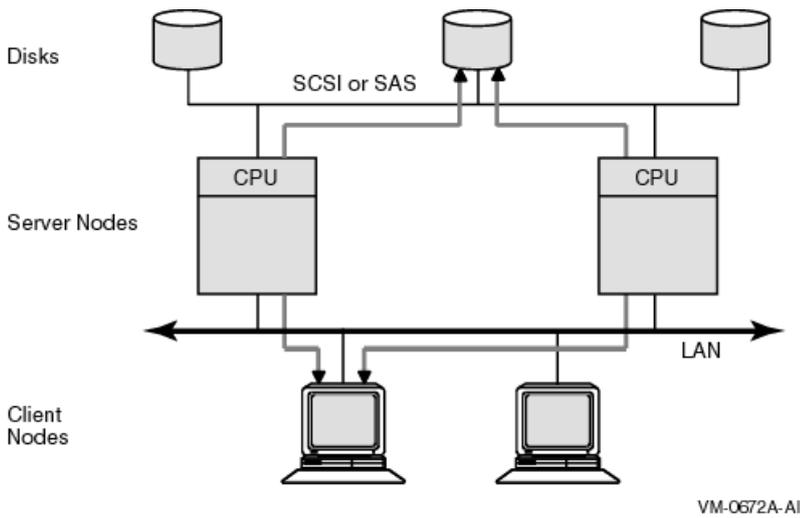
ZK-7017A-GE

Note

Disks can be shadowed using Volume Shadowing for OpenVMS. The automatic recovery from system failure provided by dual porting and shadowing is transparent to users and does not require any operator intervention.

Figure 6.2 shows a dual-pathed FC and Ethernet configuration. The disk devices, accessible through a shared SCSI interconnect, are MSCP served to the client nodes on the LAN.

Rule: A dual-pathed DSA disk cannot be used as a system disk for a directly connected CPU. Because a device can be on line to one controller at a time, only one of the server nodes can use its local connection to the device. The second server node accesses the device through the MSCP (or the TMSCP server). If the computer that is currently serving the device fails, the other computer detects the failure and fails the device over to its local connection. The device thereby remains available to the cluster.

Figure 6.2. Dual-Pathed Disks

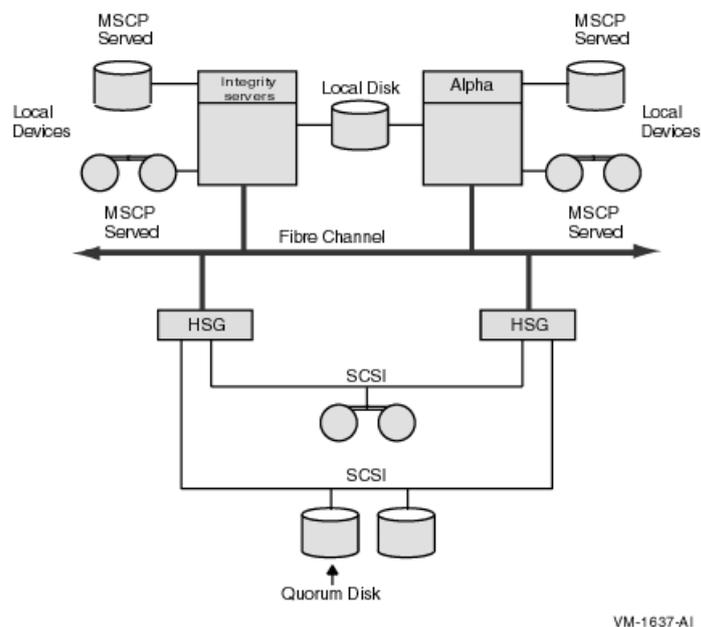
Dual-pathed disks or tapes can be failed over between two computers that serve the devices to the cluster, provided that:

- The same device controller letter is generated and the same allocation class is specified on each computer, with the result that the device has the same name on both systems. (Section 6.2.1 describes allocation classes).
- Both computers are running the MSCP server for disks, the TMSCP server for tapes, or both.

Caution

Failure to observe these requirements can endanger data integrity.

You can set up HSG or HSV storage devices to be dual ported between two storage subsystems, as shown in Figure 6.3.

Figure 6.3. Configuration with Cluster-Accessible Devices

By design, HSG and HSV disks and tapes are directly accessible by all OpenVMS Cluster nodes that are connected to the same star coupler. Therefore, if the devices are dual ported, they are automatically dual pathed. Computers connected by FC can access a dual-ported HSG or HSV device by way of a path through either subsystem connected to the device. If one subsystem fails, access fails over to the other subsystem.

Note

To control the path that is taken during failover, you can specify a preferred path to force access to disks over a specific path. Section 6.1.3 describes the preferred-path capability.

See the *Guidelines for OpenVMS Cluster Configurations*, for more information on FC storage devices.

6.1.3. Specifying a Preferred Path

The operating system supports specifying a preferred path for DSA disks, including RA series disks and disks that are accessed through the MSCP server. (This function is not available for tapes.) If a preferred path is specified for a disk, the MSCP disk class drivers use that path:

- For the first attempt to locate the disk and bring it on line with a DCL command MOUNT
- For failover of an already mounted disk

In addition, you can initiate failover of a mounted disk to force the disk to the preferred path or to use load-balancing information for disks accessed by MSCP servers.

You can specify the preferred path by using the SET PREFERRED_PATH DCL command or by using the \$QIO function (IO\$_SETPRFPATH), with the P1 parameter containing the address of a counted ASCII string(.ASCIC). This string is the node name of the HSG or HSV, or of the OpenVMS system that is to be the preferred path.

Rule: The node name must match an existing node running the MSCP server that is known to the local node.

Reference: For more information about the use of the SET PREFERRED_PATH DCL command, refer to the *VSI OpenVMS DCL Dictionary*.

For more information about the use of the IO\$_SETPRFPATH function, refer to the *VSI OpenVMS I/O User's Reference Manual*.

6.2. Naming OpenVMS Cluster Storage Devices

Note

The naming convention of Fibre Channel devices is documented in the *Guidelines for OpenVMS Cluster Configurations*. The naming of all other devices is described in this section.

In the OpenVMS operating system, a device name takes the form of *ddcu*, where:

- *dd* represents the predefined code for the device type
- *c* represents the predefined controller designation
- *u* represents the unit number

For SCSI, the controller letter is assigned by OpenVMS, based on the system configuration. The unit number is determined by the SCSI bus ID and the logical unit number (LUN) of the device.

Because device names must be unique in an OpenVMS Cluster, and because every cluster member must use the same name for the same device, OpenVMS adds a prefix to the device name, as follows:

- If a device is attached to a single computer, the device name is extended to include the name of that computer:

node\$ddcu

where *node* represents the SCS node name of the system on which the device resides.

- If a device is attached to multiple computers, the node name part of the device name is replaced by a dollar sign and a number (called a node or port allocation class, depending on usage), as follows:

\$allocation-class\$ddcu

- SAS disks follow the device naming similar to that of SCSI devices, that is, Target-LUN numbering. So a disk on SAS target ID 1 and LUN 0 will be named as DKA100. For SAS tapes you can use the Fibre channel naming convention, that is, DGAtxx: The SYSGEN parameter SAS_NAMING can be used to use SCSI numbering in tapes also.

6.2.1. Allocation Classes

The purpose of allocation classes is to provide unique and unchanging device names. The device name is used by the OpenVMS Cluster distributed lock manager in conjunction with OpenVMS facilities (such as RMS and the XQP) to uniquely identify shared devices, files, and data.

Allocation classes are required in OpenVMS Cluster configurations where storage devices are accessible through multiple paths. Without the use of allocation classes, device names that relied on node names would change as access paths to the devices change.

Prior to OpenVMS Version 7.1, only one type of allocation class existed, which was node based. It was named **allocation class**. OpenVMS Version 7.1 introduced a second type, **port allocation class**, which is specific to a single interconnect and is assigned to all devices attached to that interconnect. Port allocation classes were originally designed for naming SCSI devices. Their use has been expanded to include additional device types: floppy disks, PCI RAID controller disks, and IDE disks.

The use of port allocation classes is optional. They are designed to solve the device-naming and configuration conflicts that can occur in certain configurations, as described in Section 6.2.3.

To differentiate between the earlier node-based allocation class and the newer port allocation class, the term **node allocation class** was assigned to the earlier type.

Prior to OpenVMS Version 7.2, all nodes with direct access to the same multipathed device were required to use the same nonzero value for the node allocation class. OpenVMS Version 7.2 introduced the `MSCP_SERVE_ALL` system parameter, which can be set to serve all disks or to exclude those whose node allocation class differs.

Note

If SCSI devices are connected to multiple hosts and if port allocation classes are *not* used, then all nodes with direct access to the same multipathed devices must use the same nonzero node allocation class.

Multipathed MSCP controllers also have an allocation class parameter, which is set to match that of the connected nodes. (If the allocation class does not match, the devices attached to the nodes cannot be served).

6.2.2. Specifying Node Allocation Classes

A node allocation class can be assigned to computers, HSG or HSV controllers. The node allocation class is a numeric value from 1 to 255 that is assigned by the system manager.

The default node allocation class value is 0. A node allocation class value of 0 is appropriate only when serving a local, single-pathed disk. If a node allocation class of 0 is assigned, served devices are named using the *node-name\$device-name* syntax, that is, the device name prefix reverts to the node name.

The following rules apply to specifying node allocation class values:

1. When serving satellites, the same nonzero node allocation class value must be assigned to the serving computers and controllers.
2. All cluster-accessible devices on computers with a nonzero node allocation class value must have unique names throughout the cluster. For example, if two computers have the same node allocation class value, it is invalid for both computers to have a local disk named DGA0 or a tape named MUA0. This also applies to HSG and HSV subsystems.

System managers provide node allocation classes separately for disks and tapes. The node allocation class for disks and the node allocation class for tapes can be different.

The node allocation class names are constructed as follows:

```
$disk-allocation-class$device-name
$tape-allocation-class$device-name
```

Caution: Failure to set node allocation class values and device unit numbers correctly can endanger data integrity and cause locking conflicts that suspend normal cluster operations.

Figure 6.5 includes satellite nodes that access devices \$1\$DUA17 and \$1\$MUA12 through the JUPITR and NEPTUN computers. In this configuration, the computers JUPITR and NEPTUN require node allocation classes so that the satellite nodes are able to use consistent device names regardless of the access path to the devices.

Note: System management is usually simplified by using the same node allocation class value for all servers, HSG and HSV subsystems; you can arbitrarily choose a number between 1 and 255. Note, however, that to change a node allocation class value, you must shutdown and reboot the entire cluster (described in Section 8.6). If you use a common node allocation class for computers and controllers, ensure that all devices have unique unit numbers.

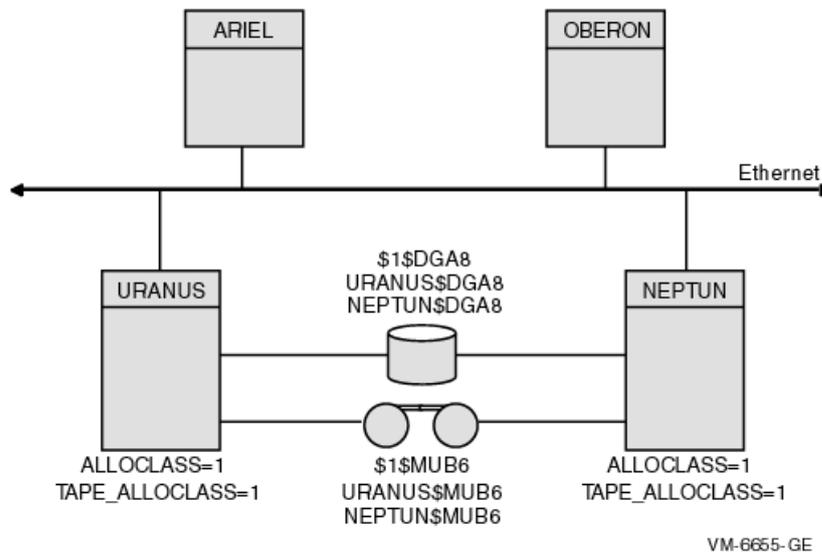
6.2.2.1. Assigning Node Allocation Class Values on Computers

There are two ways to assign a node allocation class: by using CLUSTER_CONFIG.COM or CLUSTER_CONFIG_LAN.COM, which is described in Section 8.4, or by using AUTOGEN, as shown in the following table.

Step	Action
1	<p>Edit the root directory [SYS n.SYSEXE]MODPARAMS.DAT on each node that boots from the system disk. The following example shows a MODPARAMS.DAT file. The entries are hypothetical and should be regarded as examples, not as suggestions for specific parameter settings.</p> <pre>! ! Site-specific AUTOGEN data file. In an OpenVMS Cluster ! where a common system disk is being used, this file ! should reside in SYS\$SPECIFIC:[SYSEXE], not a common ! system directory. ! ! Add modifications that you want to make to AUTOGEN's ! hardware configuration data, system parameter ! calculations, and page, swap, and dump file sizes ! to the bottom of this file. SCSNODE="NODE01" SCSSYSTEMID=99999 NISCS_LOAD_PEA0=1 VAXCLUSTER=2 MSCP_LOAD=1 MSCP_SERVE_ALL=1 ALLOCLASS=1 TAPE_ALLOCLASS=1</pre>
2	<p>Invoke AUTOGEN to set the system parameter values:</p> <pre>\$ @SYS\$UPDATE:AUTOGEN start-phase end-phase</pre>
3	<p>Shut down and reboot the entire cluster in order for the new values to take effect.</p>

6.2.2.2. Node Allocation Class Example With a DSA Disk and Tape

Figure 6.4 shows a DSA disk and tape that are dual pathed between two computers.

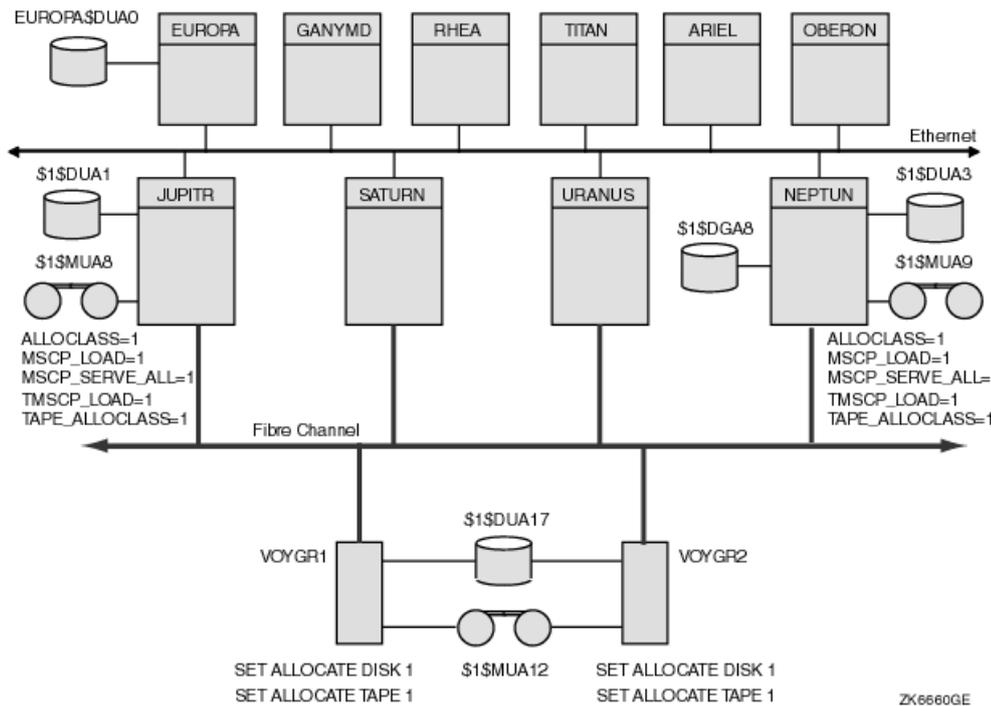
Figure 6.4. Disk and Tape Dual Pathed Between Computers

In this configuration:

- URANUS and NEPTUN access the disk either locally or through the other computer's MSCP server.
- When satellites ARIEL and OBERON access `1DGA8`, a path is made through either URANUS or NEPTUN.
- If, for example, the node URANUS has been shut down, the satellites can access the devices through NEPTUN. When URANUS reboots, access is available through either URANUS or NEPTUN.

6.2.2.3. Node Allocation Class Example With Mixed Interconnects

Figure 6.5 shows how device names are typically specified in a mixed-interconnect cluster. This figure also shows how relevant system parameter values are set for each FC computer.

Figure 6.5. Device Names in a Mixed-Interconnect Cluster

In this configuration:

- A disk and a tape are dual pathed to the HSG or HSV subsystems named VOYGR1 and VOYGR2; these subsystems are connected to JUPITR, SATURN, URANUS and NEPTUN through the star coupler.
- The MSCP and TMSCP servers are loaded on JUPITR and NEPTUN (MSCP_LOAD = 1, TMSCP_LOAD = 1) and the ALLOCLASS and TAPE_ALLOCLASS parameters are set to the same value (1) on these computers and on both HSG or HSV subsystems.

Note: For optimal availability, two or more FC connected computers can serve HSG or HSV devices to the cluster.

6.2.2.4. Node Allocation Classes and RAID Array 210 and 230 Devices

If you have RAID devices connected to StorageWorks RAID Array 210 or 230 subsystems, you might experience device-naming problems when running in a cluster environment if nonzero node allocation classes are used. In this case, the RAID devices will be named $n\$DRcu$, where n is the (nonzero) node allocation class, c is the controller letter, and u is the unit number.

If multiple nodes in the cluster have the same (nonzero) node allocation class and these same nodes have RAID controllers, then RAID devices that are distinct might be given the same name (for example, $\$1\$DRA0$). This problem can lead to data corruption.

To prevent such problems, use the DR_UNIT_BASE system parameter, which causes the DR devices to be numbered sequentially, starting with the DR_UNIT_BASE value that you specify. For example, if the node allocation class is $\$1$, the controller letter is A, and you set DR_UNIT_BASE on one cluster member to 10, the first device name generated by the RAID controller will be $\$1\$DRA10$, followed by $\$1\$DRA11$, $\$1\$DRA12$, and so forth.

To ensure unique DR device names, set the `DR_UNIT_BASE` number on each cluster member so that the resulting device numbers do not overlap. For example, you can set `DR_UNIT_BASE` on three cluster members to 10, 20, and 30 respectively. As long as each cluster member has 10 or fewer devices, the DR device numbers will be unique.

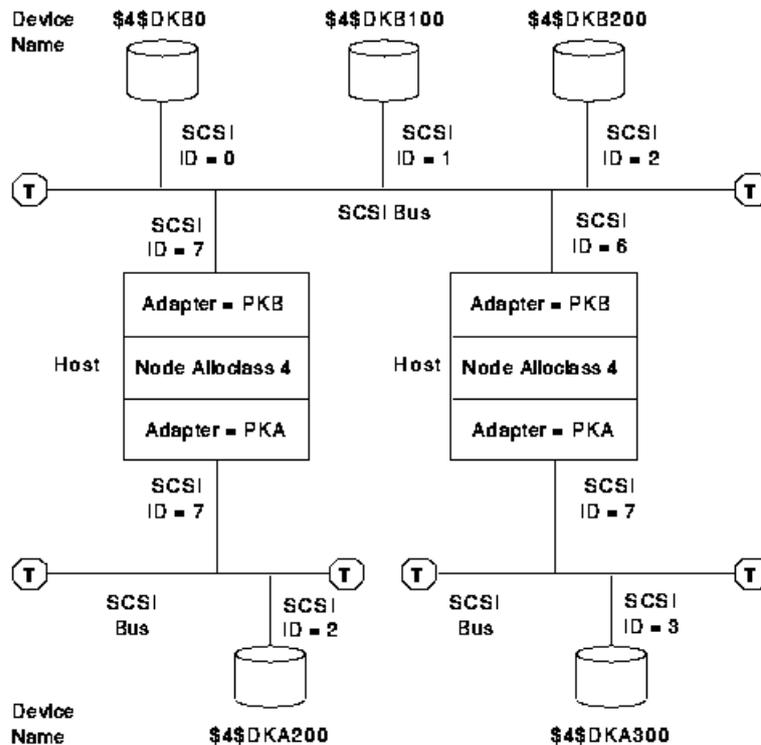
6.2.3. Reasons for Using Port Allocation Classes

When the node allocation class is nonzero, it becomes the device name prefix for all attached devices, whether the devices are on a shared interconnect or not. To ensure unique names within a cluster, it is necessary for the *ddcu* part of the disk device name (for example, DKB0) to be unique within an allocation class, even if the device is on a private bus.

This constraint is relatively easy to overcome for DIGITAL Storage Architecture(DSA) devices, because a system manager can select from a large unit number space to ensure uniqueness. The constraint is more difficult to manage for other device types, such as SCSI devices whose controller letter and unit number are determined by the hardware configuration.

For example, in the configuration shown in Figure 6.6, each system has a private SCSI bus with adapter letter A. To obtain unique names, the unit numbers must be different. This constrains the configuration to a maximum of 8 devices on the two buses (or 16 if wide addressing can be used on one or more of the buses). This can result in empty StorageWorks drive bays and in a reduction of the system's maximum storage capacity.

Figure 6.6. SCSI Device Names Using a Node Allocation Class



ZK-7489A-GE

6.2.3.1. Constraint of the SCSI Controller Letter in Device Names

The SCSI device name is determined in part by the SCSI controller through which the device is accessed (for example, B in DKB *n*). Therefore, to ensure that each node uses the same name for each device, all SCSI controllers attached to a shared SCSI bus must have the same OpenVMS device name. In Figure 6.6, each host is attached to the shared SCSI bus by controller PKB.

This requirement can make configuring a shared SCSI bus difficult, because a system manager has little or no control over the assignment of SCSI controller device names. It is particularly difficult to match controller letters on different system types when one or more of the systems have:

- Built-in SCSI controllers that are not supported in SCSI clusters
- Long internal cables that make some controllers inappropriate for SCSI clusters

6.2.3.2. Constraints Removed by Port Allocation Classes

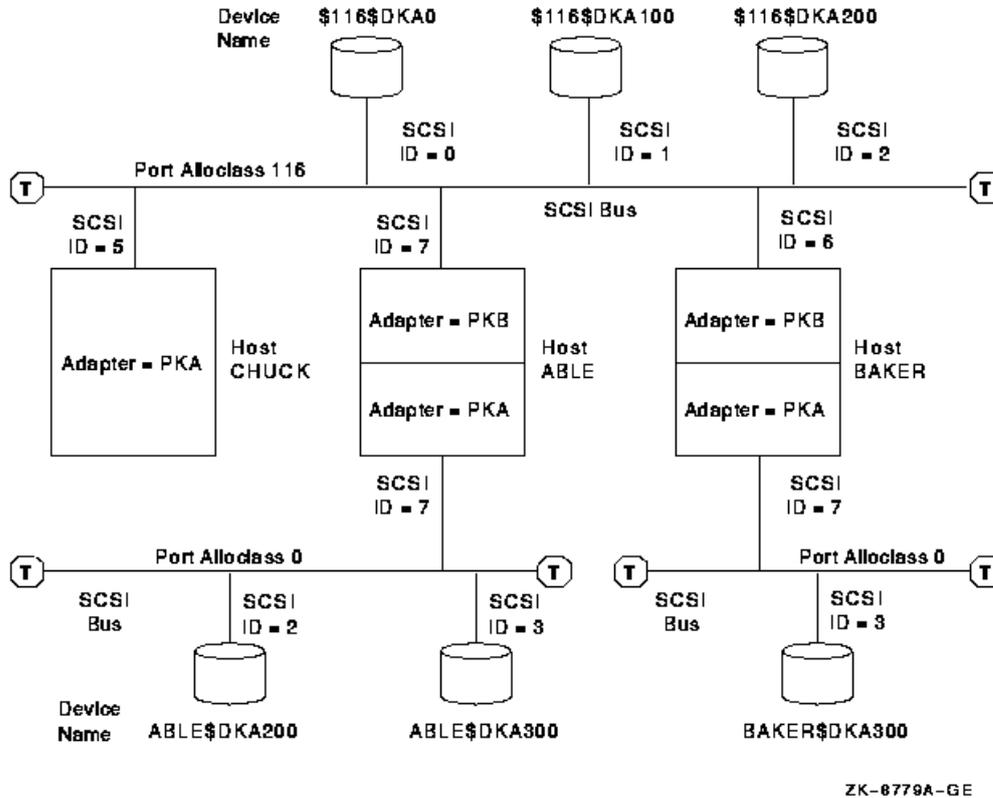
The **port allocation class** feature has two major benefits:

- A system manager can specify an allocation class value that is specific to a port rather than node-wide.
- When a port has a nonzero port allocation class, the controller letter in the device name that is accessed through that port is always the letter A.

Using port allocation classes for naming SCSI, IDE, floppy disk, and PCI RAID controller devices removes the configuration constraints described in Section 6.2.2.4, in Section 6.2.3, and in Section 6.2.3.1. You do not need to use the `DR_UNIT_BASE` system parameter recommended in Section 6.2.2.4. Furthermore, each bus can be given its own unique allocation class value, so the *ddcu* part of the disk device name (for example, DKB0) does not need to be unique across buses. Moreover, controllers with different device names can be attached to the same bus, because the disk device names no longer depend on the controller letter.

Figure 6.7 shows the same configuration as Figure 6.6, with two additions: a host named CHUCK and an additional disk attached to the lower left SCSI bus. Port allocation classes are used in the device names in this figure. A port allocation class of 116 is used for the SCSI interconnect that is shared, and port allocation class 0 is used for the SCSI interconnects that are not shared. By using port allocation classes in this configuration, you can do what was not allowed previously:

- Attach an adapter with a name (PKA) that differs from the name of the other adapters (PKB) attached to the shared SCSI interconnect, as long as that port has the same port allocation class (116 in this example).
- Use two disks with the same controller name and number (DKA300) because each disk is attached to a SCSI interconnect that is not shared.

Figure 6.7. Device Names Using Port Allocation Classes

6.2.4. Specifying Port Allocation Classes

A port allocation class is a designation for all ports attached to a single interconnect. It replaces the node allocation class in the device name.

The three types of port allocation classes are:

- Port allocation classes of 1 to 32767 for devices attached to a multihost interconnect or a single-host interconnect, if desired
- Port allocation class 0 for devices attached to a single-host interconnect
- Port allocation class -1 when no port allocation class is in effect

Each type has its own naming rules.

6.2.4.1. Port Allocation Classes for Devices Attached to a Multi-Host Interconnect

The following rules pertain to port allocation classes for devices attached to a multihost interconnect:

1. The valid range of port allocation classes is 1 through 32767.

- When using port allocation classes, the controller letter in the device name is always A, regardless of the actual controller letter. The \$GETDVI item code DVI\$_DISPLAY_DEVNAM displays the actual port name.

Note that it is now more important to use fully specified names (for example, \$101\$DKA100 or ABLE\$DKA100) rather than abbreviated names (such as DK100), because a system can have multiple DKA100 disks.

- Each port allocation class must be unique within a cluster.
- A port allocation class cannot duplicate the value of another node's tape or disk node allocation class.
- Each node for which MSCP serves a device should have the same nonzero allocation class value.

Examples of device names that use this type of port allocation class are shown in Table 6.2.

Table 6.2. Examples of Device Names with Port Allocation Classes 1-32767

Device Name	Description
\$101\$DKA0	The port allocation class is 101; DK represents the disk device category, A is the controller name, and 0 is the unit number.
\$147\$DKA0	The port allocation class is 147; DK represents the disk device category, A is the controller name, and 0 is the unit number.

6.2.4.2. Port Allocation Class 0 for Devices Attached to a Single-Host Interconnect

The following rules pertain to port allocation class 0 for devices attached to a single-host interconnect:

- Port allocation class 0 does not become part of the device name. Instead, the name of the node to which the device is attached becomes the first part of the device name.
- The controller letter in the device name remains the designation of the controller to which the device is attached. (It is not changed to A as it is for port allocation classes greater than zero.)

Examples of device names that use port allocation class 0 are shown in Table 6.3.

Table 6.3. Examples of Device Names With Port Allocation Class 0

Device Name	Description
ABLE\$DKD100	ABLE is the name of the node to which the device is attached. D is the designation of the controller to which it is attached, not A as it is for port allocation classes with a nonzero class. The unit number of this device is 100. The port allocation class of \$0\$ is not included in the device name.
BAKER\$DKC200	BAKER is the name of the node to which the device is attached, C is the designation of the controller to which it is attached, and 200 is the unit number. The port allocation class of \$0\$ is not included in the device name.

6.2.4.3. Port Allocation Class -1

The designation of port allocation class -1 means that a port allocation class is not being used. Instead, a node allocation class is used. The controller letter remains its predefined designation. (It is assigned by OpenVMS, based on the system configuration. It is not affected by a node allocation class).

6.2.4.4. How to Implement Port Allocation Classes

Port allocation classes were introduced in OpenVMS Alpha Version 7.1 with support in OpenVMS VAX. VAX computers can serve disks connected to Alpha systems that use port allocation classes in their names.

To implement port allocation classes, you must do the following:

- Enable the use of port allocation classes.
- Assign one or more port allocation classes.
- At a minimum, reboot the nodes on the shared SCSI bus.

Enabling the Use of Port Allocation Classes

To enable the use of port allocation classes, you must set a new SYSGEN parameter `DEVICE_NAMING` to 1. The default setting for this parameter is zero. In addition, the `SCSSYSTEMIDH` system parameter must be set to zero. Check to make sure that it is.

Assigning Port Allocation Classes

You can assign one or more port allocation classes with the OpenVMS Cluster configuration procedure, `CLUSTER_CONFIG.COM` (or `CLUSTER_CONFIG_LAN.COM`).

If it is not possible to use `CLUSTER_CONFIG.COM` or `CLUSTER_CONFIG_LAN.COM` to assign port allocation classes (for example, if you are booting a private system disk into an existing cluster), you can use the new `SYSBOOT SET/CLASS` command.

The following example shows how to use the new `SYSBOOT SET/CLASS` command to assign an existing port allocation class of 152 to port PKB.

```
SYSBOOT> SET/CLASS PKB 152
```

The `SYSINIT` process ensures that this new name is used in successive boots.

To deassign a port allocation class, enter the port name without a class number. For example:

```
SYSBOOT> SET/CLASS PKB
```

The mapping of ports to allocation classes is stored in `SYSS$SYSTEM:SYSS$DEVICES.DAT`, a standard text file. You use the `CLUSTER_CONFIG.COM` (or `CLUSTER_CONFIG_LAN.COM`) command procedure or, in special cases, `SYSBOOT` to change `SYSS$DEVICES.DAT`.

6.2.4.5. Clusterwide Reboot Requirements for SCSI Interconnects

Changing a device's allocation class changes the device name. A clusterwide reboot ensures that all nodes see the device under its new name, which in turn means that the normal device and file locks remain consistent.

Rebooting an entire cluster when a device name changes is not mandatory. You may be able to reboot only the nodes that share the SCSI bus, as described in the following steps. The conditions under which you can do this and the results that follow are also described.

1. Dismount the devices whose names have changed from all nodes.

This is not always possible. In particular, you cannot dismount a disk on nodes where it is the system disk. If the disk is not dismounted, a subsequent attempt to mount the same disk using the new device name will fail with the following error:

`%MOUNT-F-VOLALRMNT`, another volume of same label already mounted

Therefore, you must reboot any node that cannot dismount the disk.

2. Reboot all nodes connected to the SCSI bus.

Before you reboot any of these nodes, make sure the disks on the SCSI bus are dismounted on the nodes not rebooting.

Note

OpenVMS ensures that a node cannot boot if the result is a SCSI bus with naming different from another node already accessing the same bus. (This check is independent of the dismount check in step 1).

After the nodes that are connected to the SCSI bus reboot, the device exists with its new name.

3. Mount the devices systemwide or clusterwide.

If no other node has the disk mounted under the old name, you can mount the disk systemwide or clusterwide using its new name. The new device name will be seen on all nodes running compatible software, and these nodes can also mount the disk and access it normally.

Nodes that have not rebooted still see the old device name as well as the new device name.

However, the old device name cannot be used; the device, when accessed by the old name, is off line. The old name persists until the node reboots.

6.3. MSCP and TMSCP Served Disks and Tapes

The MSCP server and the TMSCP server make locally connected disks and tapes available to all cluster members. Locally connected disks and tapes are not automatically cluster accessible. Access to these devices is restricted to the local computer unless you explicitly set them up as cluster accessible using the MSCP server for disks or the TMSCP server for tapes.

6.3.1. Enabling Servers

To make a disk or tape accessible to all OpenVMS Cluster computers, the MSCP or TMSCP server must be:

- Loaded on the local computer, as described in Table 6.4
- Made functional by setting the MSCP and TMSCP system parameters, as described in Table 6.5

Table 6.4. MSCP_LOAD and TMSCP_LOAD Parameter Settings

Parameter	Value	Meaning
MSCP_LOAD	0	Do not load the MSCP_SERVER. This is the default.
	1	Load the MSCP server with attributes specified by the MSCP_SERVE_ALL parameter using the default CPU load capacity.

Parameter	Value	Meaning
	>1	Load the MSCP server with attributes specified by the MSCP_SERVE_ALL parameter. Use the MSCP_LOAD value as the CPU load capacity.
TMSCP_LOAD	0	Do not load the TMSCP server and do not serve any tapes (default value).
	1	Load the TMSCP server and serve all available tapes, including all local tapes and all multihost tapes with a matching TAPE_ALLOCLASS value.

Table 6.5 summarizes the system parameter values you can specify for MSCP_SERVE_ALL and TMSCP_SERVE_ALL to configure the MSCP and TMSCP servers. Initial values are determined by your responses when you execute the installation or upgrade procedure or when you execute the CLUSTER_CONFIG.COM command procedure described in Chapter 8 to set up your configuration.

Starting with OpenVMS Version 7.2, the serving types are implemented as a bit mask. To specify the type of serving your system will perform, locate the type you want in Table 6.5 and specify its value. For some systems, you may want to specify two serving types, such as serving the system disk and serving locally attached disks. To specify such a combination, add the values of each type, and specify the sum.

Note

In a mixed-version cluster that includes any systems running OpenVMS Version 7.1-*x* or earlier, serving all available disks is restricted to serving all disks whose allocation class matches the system's node allocation class (pre-Version 7.2 meaning). To specify this type of serving, use the value 9 (which sets bit 0 and bit 3).

Table 6.5. MSCP_SERVE_ALL and TMSCP_SERVE_ALL Parameter Settings

Parameter	Bit	Value When Set	Meaning
MSCP_SERVE_ALL	0	1	Serve all available disks (locally attached and those connected to HS <i>x</i> and DSSI controllers). Disks with allocation classes that differ from the system's allocation class (set by the ALLOCLASS parameter) are also served if bit 3 is not set.
	1	2	Serve locally attached (non-HS <i>x</i> and non-DSSI) disks. The server does not monitor its I/O traffic and does not participate in load balancing.
	2	4	Serve the system disk. This is the default setting. This setting is important when other nodes in the cluster rely on this system being able to serve its system disk. This setting prevents obscure contention problems that can occur when a system attempts to complete I/O to a remote system disk whose system has failed. For more information, see Section 6.3.1.1.
	3	8	Restrict the serving specified by bit 0. All disks except those with allocation classes that differ from the

Parameter	Bit	Value When Set	Meaning
			<p>system's allocation class (set by the ALLOCLASS parameter) are served.</p> <p>This is pre-Version 7.2 behavior. If your cluster includes systems running Open 7.1- <i>x</i> or earlier, and you want to serve all available disks, you must specify 9, the result of setting this bit and bit 0.</p>
	4	15	By default, the bit 4 is not set, hence the DUDRIVER will accept the devices with unit number greater than 9999. On the client side, if bit 4 is set (10000 binary) in the MSCP_SERVE_ALL parameter, the client will reject devices with unit number greater than 9999 and retains the earlier behavior.
TMSCP_SERVE_ALL	0	1	Serve all available tapes (locally attached and those connected to HS <i>x</i> and DSSI controllers). Tapes with allocation classes that differ from the system's allocation class (set by the ALLOCLASS parameter) are also served if bit 3 is not set.
	1	2	Serve locally attached (non-HS <i>x</i> and non-DSSI) tapes.
	3	8	<p>Restrict the serving specified by bit 0. Serve all tapes except those with allocation classes that differ from the system's allocation class (set by the ALLOCLASS parameter).</p> <p>This is pre-Version 7.2 behavior. If your cluster includes systems running OpenVMS Version 7.1- <i>x</i> or earlier, and you want to serve all available tapes, you must specify 9, the result of setting this bit and bit 0.</p>
	4	15	By default, the bit 4 is not set, hence the TUDRIVER will accept the devices with unit number greater than 9999. On the client side, if bit 4 is set (10000 binary) in the TMSCP_SERVE_ALL parameter, the client will reject devices with unit number greater than 9999 and retains the earlier behavior.

Although the serving types are now implemented as a bit mask, the values of 0, 1, and 2, specified by bit 0 and bit 1, retain their original meanings. These values are shown in the following table:

Value	Description
0	Do not serve any disks (tapes). This is the default.
1	Serve all available disks (tapes).
2	Serve only locally attached (non-HS <i>x</i> and non-DSSI) disks (tapes).

6.3.1.1. Serving the System Disk

Setting bit 2 of the MSCP_SERVE_ALL system parameter to serve the system disk is important when other nodes in the cluster rely on this system being able to serve its system disk. This setting

prevents obscure contention problems that can occur when a system attempts to complete I/O to a remote system disk whose system has failed.

The following sequence of events describes how a contention problem can occur if serving the system disk is disabled (that is, if bit 2 is not set):

- The MSCP_SERVE_ALL setting is changed to disable serving when the system reboots.
- The serving system crashes.
- The client system that was executing I/O to the serving system's system disk is holding locks on resources of that system disk.
- The client system starts mount verification.
- The serving system attempts to boot but cannot because of the locks held on its system disk by the client system.
- The client's mount verification process times out after a period of time set by the MVTIMEOUT system parameter, and the client system releases the locks. The time period could be several hours.
- The serving system is able to reboot.

6.3.1.2. Setting the MSCP and TMSCP System Parameters

Use either of the following methods to set these system parameters:

- Specify appropriate values for these parameters in a computer's MODPARAMS.DAT file and then run AUTOGEN.
- Run the CLUSTER_CONFIG.COM or the CLUSTER_CONFIG_LAN.COM procedure, as appropriate, and choose the CHANGE option to perform these operations for disks and tapes.

With either method, the served devices become accessible when the serving computer reboots. Further, the servers automatically serve any suitable device that is added to the system later. For example, if new drives are attached to an HSC subsystem, the devices are dynamically configured.

Note: The SCSI retention command modifier is not supported by the TMSCP server. Retention operations should be performed from the node serving the tape.

6.4. MSCP I/O Load Balancing

MSCP I/O load balancing offers the following advantages:

- Faster I/O response
- Balanced work load among the members of an OpenVMS Cluster

Two types of MSCP I/O load balancing are provided by OpenVMS Cluster software: static and dynamic. Static load balancing occurs on Integrity servers and Alpha systems and are based on the load capacity ratings of the server systems.

6.4.1. Load Capacity

The load capacity ratings for Integrity servers and Alpha systems are predetermined by VSI. These ratings are used in the calculation of the available serving capacity for MSCP static and dynamic load

balancing. You can override these default settings by specifying a different load capacity with the `MSCP_LOAD` parameter.

Note that the MSCP server load-capacity values (either the default value or the value you specify with `MSCP_LOAD`) are estimates used by the load-balancing feature. They cannot change the actual MSCP serving capacity of a system.

A system's MSCP serving capacity depends on many factors including its power, the performance of its LAN adapter, and the impact of other processing loads. The available serving capacity, which is calculated by each MSCP server as described in Section 6.4.2, is used solely to bias the selection process when a client system (for example, a satellite) chooses which server system to use when accessing a served disk.

6.4.2. Available Serving Capacity

The load-capacity ratings are used by each MSCP server to calculate its available serving capacity.

The **available serving capacity** is calculated in the following way:

Step	Calculation
1	Each MSCP server counts the read and write requests sent to it and periodically converts this value to requests per second.
2	Each MSCP server subtracts its requests per second from its load capacity to compute its available serving capacity.

6.4.3. Static Load Balancing

MSCP servers periodically send their available serving capacities to the MSCP class driver(DUDRIVER). When a disk is mounted or one fails over, DUDRIVER assigns the server with the highest available serving capacity to it. (TMSCP servers do not perform this monitoring function.) This initial assignment is called static load balancing.

6.4.4. Overriding MSCP I/O Load Balancing for Special Purposes

In some configurations, you may want to designate one or more systems in your cluster as the primary I/O servers and restrict I/O traffic on other systems. You can accomplish these goals by overriding the default load-capacity ratings used by the MSCP server. For example, if your cluster consists of two Alpha systems and one VAX 6000-400 system and you want to reduce the MSCP served I/O traffic to the VAX, you can assign the VAX a low `MSCP_LOAD` value, such as 50. Because the two Alpha systems each start with a load-capacity rating of 340 and the VAX now starts with a load-capacity rating of 50, the MSCP served satellites will direct most of the I/O traffic to the Alpha systems.

6.5. Managing Cluster Disks With the Mount Utility

For locally connected disks to be accessible to other nodes in the cluster, the MSCP server software must be loaded on the computer to which the disks are connected (see Section 6.3.1). Further, each disk must be mounted with the Mount utility, using the appropriate qualifier: `/CLUSTER, /`

SYSTEM, or /GROUP. Mounting multiple disks can be automated with command procedures; a sample command procedure, MSCPMOUNT.COM, is provided in the SYS\$EXAMPLES directory on your system.

The Mount utility also provides other qualifiers that determine whether a disk is automatically rebuilt during a remount operation. Different rebuilding techniques are recommended for data and system disks.

This section describes how to use the Mount utility for these purposes.

6.5.1. Mounting Cluster Disks

To mount disks that are to be shared among all computers, specify the MOUNT command as shown in the following table.

IF...	THEN...
At system startup	
The disk is attached to a single system and is to be made available to all other nodes in the cluster.	Use MOUNT/CLUSTER <i>device-name</i> on the computer to which the disk is to be mounted. The disk is mounted on every computer that is active in the cluster at the time the command executes. First, the disk is mounted locally. Then, if the mount operation succeeds, the disk is mounted on other nodes in the cluster.
The computer has no disks directly attached to it.	Use MOUNT/SYSTEM <i>device-name</i> on the computer for each disk the computer needs to access. The disks can be attached to a single system or shared disks that are accessed by an HS xcontroller. Then, if the mount operation succeeds, the disk is mounted on the computer joining the cluster.
When the system is running	
You want to add a disk.	Use MOUNT/CLUSTER <i>device-name</i> on the computer to which the disk is to be mounted. The disk is mounted on every computer that is active in the cluster at the time the command executes. First, the disk is mounted locally. Then, if the mount operation succeeds, the disk is mounted on other nodes in the cluster.

To ensure disks are mounted whenever possible, regardless of the sequence that systems in the cluster boot (or shut down), startup command procedures should use MOUNT/CLUSTER and MOUNT/SYSTEM as described in the preceding table.

Note: Only system or group disks can be mounted across the cluster or on a subset of the cluster members. If you specify MOUNT/CLUSTER without the /SYSTEM or /GROUP qualifier, /SYSTEM is assumed. Also note that each cluster disk mounted with the /SYSTEM or /GROUP qualifier must have a unique volume label.

6.5.2. Examples of Mounting Shared Disks

Suppose you want all the computers in a three-member cluster to share a disk named COMPANYDOCS. To share the disk, one of the three computers can mount COMPANYDOCS using the MOUNT/CLUSTER command, as follows:

```
$ MOUNT/CLUSTER/NOASSIST $1$DUA4: COMPANYDOCS
```

If you want just two of the three computers to share the disk, those two computers must both mount the disk with the same MOUNT command, as follows:

```
$ MOUNT/SYSTEM/NOASSIST $1$DUA4: COMPANYDOCS
```

To mount the disk at startup time, include the MOUNT command either in a common command procedure that is invoked at startup time or in the computer-specific startup command file.

Note: The /NOASSIST qualifier is used in command procedures that are designed to make several attempts to mount disks. The disks may be temporarily offline or otherwise not available for mounting. If, after several attempts, the disk cannot be mounted, the procedure continues. The /ASSIST qualifier, which is the default, causes a command procedure to stop and query the operator if a disk cannot be mounted immediately.

6.5.3. Mounting Cluster Disks With Command Procedures

To configure cluster disks, you can create command procedures to mount them. You may want to include commands that mount cluster disks in a separate command procedure file that is invoked by a site-specific SYSTARTUP procedure. Depending on your cluster environment, you can set up your command procedure in either of the following ways:

- As a separate file specific to each computer in the cluster by making copies of the common procedure and storing them as separate files
- As a common computer-independent file on a shared disk

With either method, each computer can invoke the common procedure from the site-specific SYSTARTUP procedure.

Example: The MSCPMOUNT.COM file in the SYS\$EXAMPLES directory on your system is a sample command procedure that contains commands typically used to mount cluster disks. The example includes comments explaining each phase of the procedure.

6.5.4. Disk Rebuild Operation

To minimize disk I/O operations (and thus improve performance) when files are created or extended, the OpenVMS file system maintains a cache of preallocated file headers and disk blocks.

If a disk is dismounted improperly—for example, if a system fails or is removed from a cluster without running SYSSYSTEM:SHUTDOWN.COM—this preallocated space becomes temporarily unavailable. When the disk is remounted, MOUNT scans the disk to recover the space. This is called a **disk rebuild operation**.

6.5.5. Rebuilding Cluster Disks

On a nonclustered computer, the MOUNT scan operation for recovering preallocated space merely prolongs the boot process. In an OpenVMS Cluster system, however, this operation can degrade response time for all user processes in the cluster. While the scan is in progress on a particular disk, most activity on that disk is blocked.

Note: User processes that attempt to read or write to files on the disk can experience delays of several minutes or longer, especially if the disk contains a large number of files or has many users.

Because the rebuild operation can delay access to disks during the startup of any OpenVMS Cluster computer, VSI recommends that procedures for mounting cluster disks use the /NOREBUILD qualifier. When MOUNT/NOREBUILD is specified, disks are not scanned to recover lost space, and users experience minimal delays while computers are mounting disks.

Reference: Section 6.5.6 provides information about rebuilding system disks. Section 9.7.1 provides more information about disk rebuilds and system-disk throughput techniques.

6.5.6. Rebuilding System Disks

Rebuilding system disks is especially critical because most system activity requires access to a system disk. When a system disk rebuild is in progress, very little activity is possible on any computer that uses that disk.

Unlike other disks, the system disk is automatically mounted early in the boot sequence. If a rebuild is necessary, and if the value of the system parameter ACP_REBLDSYSD is 1, the system disk is rebuilt during the boot sequence. (The default setting of 1 for the ACP_REBLDSYSD system parameter specifies that the system disk should be rebuilt.) Exceptions are as follows:

Setting	Comments
ACP_REBLDSYSD parameter should be set to 0 on satellites.	This setting prevents satellites from rebuilding a system disk when it is mounted early in the boot sequence and eliminates delays caused by such a rebuild when satellites join the cluster.
ACP_REBLDSYSD should be set to the default value of 1 on boot servers, and procedures that mount disks on the boot servers should use the /REBUILD qualifier.	While these measures can make boot server rebooting more noticeable, they ensure that system disk space is available after an unexpected shutdown.

Once the cluster is up and running, system managers can submit a batch procedure that executes SET VOLUME/REBUILD commands to recover lost disk space. Such procedures can run at a time when users would not be inconvenienced by the blocked access to disks (for example, between midnight and 6 a.m. each day). Because the SET VOLUME/REBUILD command determines whether a rebuild is needed, the procedures can execute the command for each disk that is usually mounted.

Suggestion: The procedures run more quickly and cause less delay in disk access if they are executed on:

- Powerful computers
- Computers that have direct access to the volume to be rebuilt

Moreover, several such procedures, each of which rebuilds a different set of disks, can be executed simultaneously.

Caution: If either or both of the following conditions are true when mounting disks, it is essential to run a procedure with SET VOLUME/REBUILD commands on a regular basis to rebuild the disks:

- Disks are mounted with the MOUNT/NOREBUILD command.

- The ACP_REBLDSYSD system parameter is set to 0.

Failure to rebuild disk volumes can result in a loss of free space and in subsequent failures of applications to create or extend files.

6.6. Shadowing Disks Across an OpenVMS Cluster

Volume shadowing (sometimes referred to as disk mirroring) achieves high data availability by duplicating data on multiple disks. If one disk fails, the remaining disk or disks can continue to service application and user I/O requests.

6.6.1. Purpose

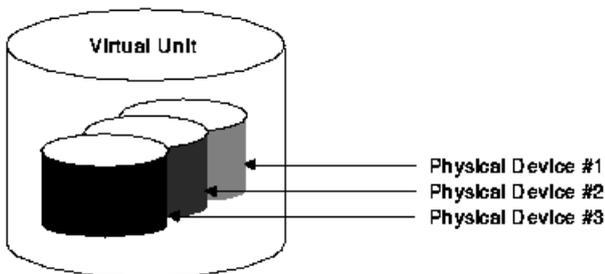
Volume Shadowing for OpenVMS software provides data availability across the full range of OpenVMS configurations—from single nodes to large OpenVMS Cluster systems—so you can provide data availability where you need it most.

Volume Shadowing for OpenVMS software is an implementation of RAID 1 (redundant arrays of independent disks) technology. Volume Shadowing for OpenVMS prevents a disk device failure from interrupting system and application operations. By duplicating data on multiple disks, volume shadowing transparently prevents your storage subsystems from becoming a single point of failure because of media deterioration, communication path failure, or controller or device failure.

6.6.2. Shadow Sets

You can mount up to six compatible disk volumes to form a **shadow set**. Figure 6.8 shows three compatible disk volumes used to form a shadow set. Each disk in the shadow set is known as a shadow set **member**. Volume Shadowing for OpenVMS logically binds the shadow set devices to gether and represents them as a single virtual device called a **virtual unit**. This means that the multiple members of the shadow set, represented by the virtual unit, appear to operating systems and users as a single, highly available disk.

Figure 6.8. Shadow Set With Three Members



ZK-5156A-GE

6.6.3. I/O Capabilities

Applications and users read and write data to and from a shadow set using the same commands and program language syntax and semantics that are used for nonshadowed I/O operations. System

managers manage and monitor shadow sets using the same commands and utilities they use for nonshadowed disks. The only difference is that access is through the virtual unit, not to individual devices.

Reference: *VSI OpenVMS Volume Shadowing Guide* describes the shadowing product capabilities in detail.

6.6.4. Supported Devices

For a single workstation or a large data centre, valid shadowing configurations include:

- All MSCP compliant DSA drives
- All SAS devices
- All StorageWorks SCSI disks and controllers, and some third-party SCSI devices that implement READL (read long) and WRITEL (write long) commands and use the SCSI disk driver (DKDRIVER)

Restriction: SCSI disks that do not support READL and WRITEL are restricted because these disks do not support the shadowing data repair (disk bad-block errors) capability. Thus, using unsupported SCSI disks can cause members to be removed from the shadow set.

You can shadow data disks and system disks. Thus, a system disk need not be a single point of failure for any system that boots from that disk. System disk shadowing becomes especially important for OpenVMS Cluster systems that use a *common* system disk from which multiple computers boot.

Volume Shadowing for OpenVMS does not support the shadowing of quorum disks. This is because volume shadowing makes use of the OpenVMS distributed lock manager, and the quorum disk must be utilized before locking is enabled.

There are no restrictions on the location of shadow set members beyond the valid disk configurations defined in the *Volume Shadowing for OpenVMS Software Product Description*.

6.6.5. Shadow Set Limits

You can mount a default maximum of 500 shadow sets (each having one to six members) in a standalone system or OpenVMS Cluster system. If more than 500 shadow sets are required, the SYSGEN parameter SHADOW_MAX_UNIT must be increased. The number of shadow sets supported is independent of controller and device types. The shadow sets can be mounted as public or private volumes.

For any changes to these limits, consult the *Volume Shadowing for OpenVMS Software Product Description*.

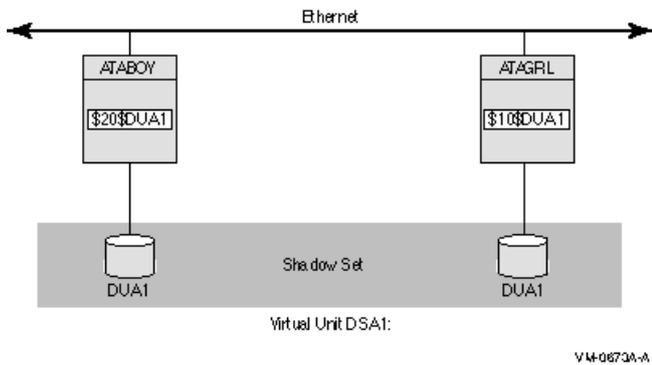
6.6.6. Distributing Shadowed Disks

The controller-independent design of shadowing allows you to manage shadow sets regardless of their controller connection or location in the OpenVMS Cluster system and helps provide improved data availability and very flexible configurations.

For clusterwide shadowing, members can be located anywhere in an OpenVMS Cluster system and served by MSCP servers across any supported OpenVMS Cluster interconnect.

Figure 6.9 shows how shadow set member units are on line to local controllers located on different nodes. In the figure, a disk volume is local to each of the nodes ATABOY and ATAGRL. The MSCP server provides access to the shadow set members over the LAN or IP network. Even though the disk volumes are local to different nodes, the disks are members of the same shadow set. A member unit that is local to one node can be accessed by the remote node over the MSCP server.

Figure 6.9. Shadow Sets Accessed Through the MSCP Server



For shadow sets that are mounted on an OpenVMS Cluster system, mounting or dismounting a shadow set on one node in the cluster does not affect applications or user functions executing on other nodes in the system. For example, you can dismount the virtual unit from one node in an OpenVMS Cluster system and leave the shadow set operational on the remaining nodes on which it is mounted.

Other shadowing notes:

- If an individual disk volume is already mounted as a member of an active shadow set, the disk volume cannot be mounted as a standalone disk on another node at the same time.
- System disks can be shadowed. All nodes booting from shadowed system disks must:
 - Have a Volume Shadowing for OpenVMS license.
 - Set shadowing system parameters to enable shadowing and specify the system disk virtual unit number.

Chapter 7. Setting Up and Managing Cluster Queues

This chapter discusses queuing topics specific to OpenVMS Cluster systems. Because queues in an OpenVMS Cluster system are established and controlled with the same commands used to manage queues on a standalone computer, the discussions in this chapter assume some knowledge of queue management on a standalone system, as described in the *VSI OpenVMS System Manager's Manual*. **Note:** See the *VSI OpenVMS System Manager's Manual* for information about queuing compatibility.

7.1. Introduction

Users can submit jobs to any queue in the OpenVMS Cluster system, regardless of the processor on which the job will actually execute. Generic queues can balance the work load among the available processors.

The system manager can use one or several queue managers to manage batch and print queues for an entire OpenVMS Cluster system. Although a single queue manager is sufficient for most systems, multiple queue managers can be useful for distributing the batch and print work load across nodes in the cluster.

Once the batch and print queue characteristics are set up, the system manager can rely on the distributed queue manager to make queues available across the cluster.

7.2. Controlling Queue Availability

The distributed queue manager prevents the queuing system from being affected when a node enters or leaves the cluster during cluster transitions. The following table describes how the distributed queue manager works.

WHEN...	THEN...	Comments
The node on which the queue manager is running leaves the OpenVMS Cluster system.	The queue manager automatically fails over to another node.	This failover occurs transparently to users on the system.
Nodes are added to the cluster.	The queue manager automatically serves the new nodes.	The system manager does not need to enter a command explicitly to start queuing on the new node.
The OpenVMS Cluster system reboots.	The queuing system automatically restarts by default.	Thus, you do not have to include commands in your startup command procedure for queuing.
	The operating system automatically restores the queuing system with the parameters defined in the queuing database.	This is because when you start the queuing system, the characteristics you define are stored in a queue database.

To control queues, the queue manager maintains a clusterwide queue database that stores information about queues and jobs. Whether you use one or several queue managers, only one queue database

is shared across the cluster. Keeping the information for all processes in one database allows jobs submitted from any computer to execute on any queue (provided that the necessary mass storage devices are accessible).

7.3. Starting a Queue Manager and Creating the Queue Database

You start up a queue manager using the `START/QUEUE/MANAGER` command as you would on a standalone computer. However, in an OpenVMS Cluster system, you can also provide a failover list and a unique name for the queue manager. The `/NEW_VERSION` qualifier creates a new queue database.

The following command example shows how to start a queue manager:

```
$ START/QUEUE/MANAGER/NEW_VERSION/ON=( GEM , STONE , * )
```

The following table explains the components of this sample command.

Command	Function
<code>START/QUEUE/MANAGER</code>	Creates a single, clusterwide queue manager named <code>SYS\$QUEUE_MANAGER</code> .
<code>/NEW_VERSION</code>	Creates a new queue database in <code>SYS\$COMMON:[SYSEXEXE]</code> that consists of the following three files: <ul style="list-style-type: none"> • <code>QMAN\$MASTER.DAT</code> (master file) • <code>SYS\$QUEUE_MANAGER.QMAN\$QUEUES</code> (queue file) • <code>SYS\$QUEUE_MANAGER.QMAN\$JOURNAL</code> (journal file) <p>Rule: Use the <code>/NEW_VERSION</code> qualifier only on the first invocation of the queue manager or if you want to create a new queue database.</p>
<code>/ON= (node-list) [optional]</code>	Specifies an ordered list of nodes that can claim the queue manager if the node running the queue manager should exit the cluster. In the example: <ul style="list-style-type: none"> • The queue manager process starts on node <code>GEM</code>. • If the queue manager is running on node <code>GEM</code> and <code>GEM</code> leaves the cluster, the queue manager fails over to node <code>STONE</code>. • The asterisk wildcard (*) is specified as the last node in the node list to indicate that any remaining, unlisted nodes can start the queue manager in any order. <p>Rules: Complete node names are required; you cannot specify the asterisk wildcard character as part of a node name.</p>

Command	Function
	If you want to exclude certain nodes from being eligible to run the queue manager, do not use the asterisk wildcard character in the node list.
/NAME_OF_MANAGER [optional]	Allows you to assign a unique name to the queue manager. Unique queue manager names are necessary if you run multiple queue managers. For example, using the /NAME_OF_MANAGER qualifier causes queue and journal files to be created using the queue manager name instead of the default name SYSS\$QUEUE_MANAGER. For example, adding the /NAME_OF_MANAGER=PRINT_MANAGER qualifier command creates these files: QMAN\$MASTER.DAT PRINT_MANAGER.QMAN\$QUEUES PRINT_MANAGER.QMAN\$JOURNAL
<p>Rules for OpenVMS Cluster systems with multiple system disks:</p> <ul style="list-style-type: none"> Specify the locations of both the master file and the queue and journal files for systems that do not boot from the system disk where the files are located. <p>Reference: If you want to locate the queue database files on other devices or directories, refer to the <i>VSI OpenVMS System Manager's Manual</i> for instructions.</p> <ul style="list-style-type: none"> Specify a device and directory that is accessible across the OpenVMS Cluster. Define the device and directory identically in the SYSS\$COMMON:SYLOGICALS.COM startup command procedure on every node. 	

7.4. Starting Additional Queue Managers

Running multiple queue managers balances the work load by distributing batch and print jobs across the cluster. For example, you might create separate queue managers for batch and print queues in clusters with CPU or memory shortages. This allows the batch queue manager to run on one node while the print queue manager runs on a different node.

7.4.1. Command Format

To start additional queue managers, include the /ADD and /NAME_OF_MANAGER qualifiers on the START/QUEUE/MANAGER command. Do not specify the /NEW_VERSION qualifier. For example:

```
$ START/QUEUE/MANAGER/ADD/NAME_OF_MANAGER=BATCH_MANAGER
```

7.4.2. Database Files

Multiple queue managers share one QMAN\$MASTER.DAT master file, but an additional queue file and journal file is created for each queue manager. The additional files are named in the following format, respectively:

- name_of_manager.QMAN\$QUEUES*

- `name_of_manager.QMAN$JOURNAL`

By default, the queue database and its files are located in `SYSS$COMMON:[SYSEXE]`. If you want to relocate the queue database files, refer to the instructions in Section 7.6.

7.5. Stopping the Queuing System

When you enter the `STOP/QUEUE/MANAGER/CLUSTER` command, the queue manager remains stopped, and requests for queuing are denied until you enter the `START/QUEUE/MANAGER` command (without the `/NEW_VERSION` qualifier).

The following command shows how to stop a queue manager named `PRINT_MANAGER`:

```
$ STOP/QUEUE/MANAGER/CLUSTER/NAME_OF_MANAGER=PRINT_MANAGER
```

Rule: You must include the `/CLUSTER` qualifier on the command line whether or not the queue manager is running on an OpenVMS Cluster system. If you omit the `/CLUSTER` qualifier, the command stops all queues on the default node without stopping the queue manager. (This has the same effect as entering the `STOP/QUEUE/ON_NODE` command).

7.6. Moving Queue Database Files

The files in the queue database can be relocated from the default location of `SYSS$COMMON:[SYSEXE]` to any disk that is mounted clusterwide or that is accessible to the computers participating in the clusterwide queue scheme. For example, you can enhance system performance by locating the database on a shared disk that has a low level of activity.

7.6.1. Location Guidelines

The master file `QMAN$MASTER` can be in a location separate from the queue and journal files, but the queue and journal files must be kept together in the same directory. The queue and journal files for one queue manager can be separate from those of other queue managers.

The directory you specify must be available to all nodes in the cluster. If the directory specification is a concealed logical name, it must be defined identically in the `SYSS$COMMON:SYLOGICALS.COM` startup command procedure on every node in the cluster.

Reference: The *VSI OpenVMS System Manager's Manual* contains complete information about creating or relocating the queue database files. See also Section 7.12 for a sample common procedure that sets up an OpenVMS Cluster batch and print system.

7.7. Setting Up Print Queues

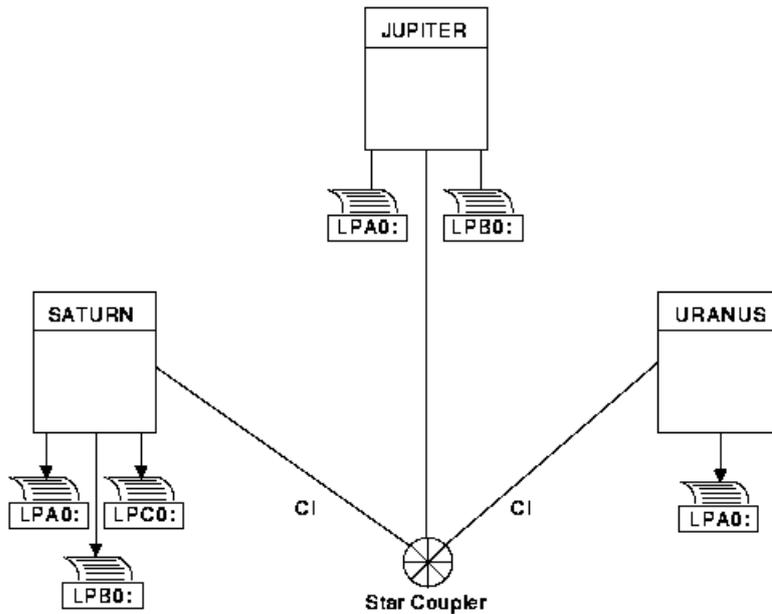
To establish print queues, you must determine the type of queue configuration that best suits your OpenVMS Cluster system. You have several alternatives that depend on the number and type of print devices you have on each computer and on how you want print jobs to be processed. For example, you need to decide:

- Which print queues you want to establish on each computer
- Whether to set up any clusterwide generic queues to distribute print job processing across the cluster

- Whether to set up auto start queues for availability or improved startup time

Once you determine the appropriate strategy for your cluster, you can create your queues. Figure 7.1 shows the printer configuration for a cluster consisting of the active computers JUPITER, SATURN, and URANUS.

Figure 7.1. Sample Printer Configuration



ZK-1631-GE

7.7.1. Creating a Queue

You set up OpenVMS Cluster print queues using the same method that you would use for a standalone computer. However, in an OpenVMS Cluster system, you must provide a unique name for each queue you create.

7.7.2. Command Format

You create and name a print queue by specifying the INITIALIZE/QUEUE command at the DCL prompt in the following format:

```
$ INITIALIZE/QUEUE/ON=node-name::device[/START][/NAME_OF_MANAGER=
name-of-manager] queue-name
```

Qualifier	Description
/ON	Specifies the computer and printer to which the queue is assigned. If you specify the /START qualifier, the queue is started.

Qualifier	Description
/NAME_OF_MANAGER	If you are running multiple queue managers, you should also specify the queue manager with the qualifier.

7.7.3. Ensuring Queue Availability

You can also use the auto start feature to simplify startup and ensure high availability of execution queues in an OpenVMS Cluster. If the node on which the autostart queue is running leaves the OpenVMS Cluster, the queue automatically fails over to the next available node on which autostart is enabled. Autostart is particularly useful on LAT queues. Because LAT printers are usually shared among users of multiple systems or in OpenVMS Cluster systems, many users are affected if a LAT queue is unavailable.

Format for creating autostart queues:

Create an autostart queue with a list of nodes on which the queue can run by specifying the DCL command INITIALIZE/QUEUE in the following format:

```
INITIALIZE/QUEUE/AUTOSTART_ON=(node-name::device:,node-name::device:, . . .
queue-name
```

When you use the /AUTOSTART_ON qualifier, you must initially activate the queue for autostart, either by specifying the /START qualifier with the INITIALIZE/QUEUE command or by entering a START/QUEUE command. However, the queue cannot begin processing jobs until the ENABLE AUTOSTART/QUEUES command is entered for a node on which the queue can run. Generic queues cannot be autostart queues.

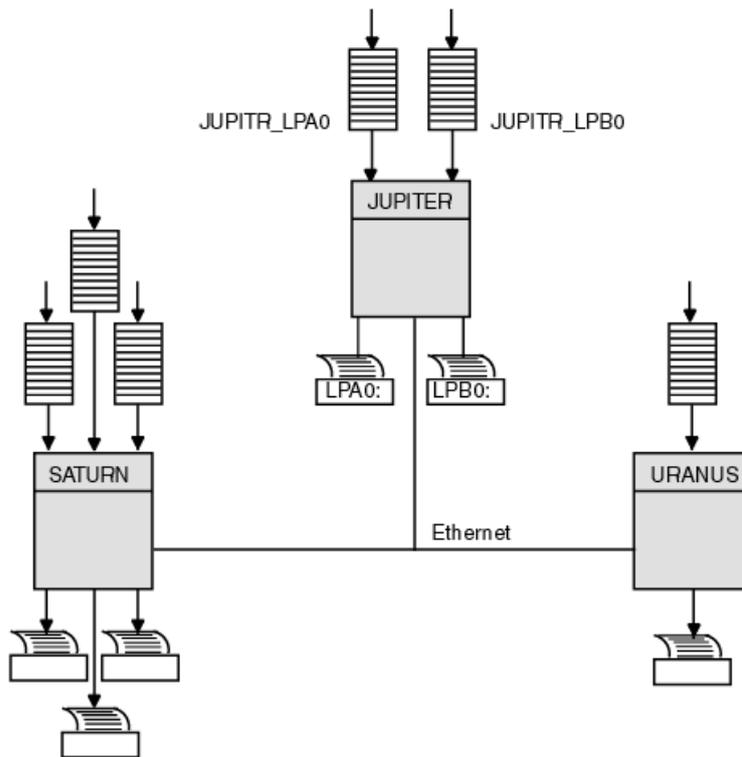
Rules: Generic queues cannot be autostart queues. Note that you cannot specify both /ON and /AUTOSTART_ON.

Reference: Refer to Section 7.13 for information about setting the time at which autostart is disabled.

7.7.4. Examples

The following commands make the local print queue assignments for JUPITR shown in Figure 7.2 and start the queues:

```
$ INITIALIZE/QUEUE/ON=JUPITR::LPA0/START/NAME_OF_MANAGER=PRINT_MANAGER
JUPITR_LPA0
$ INITIALIZE/QUEUE/ON=JUPITR::LPB0/START/NAME_OF_MANAGER=PRINT_MANAGER
JUPITR_LPB0
```

Figure 7.2. Print Queue Configuration

ZK1632GE

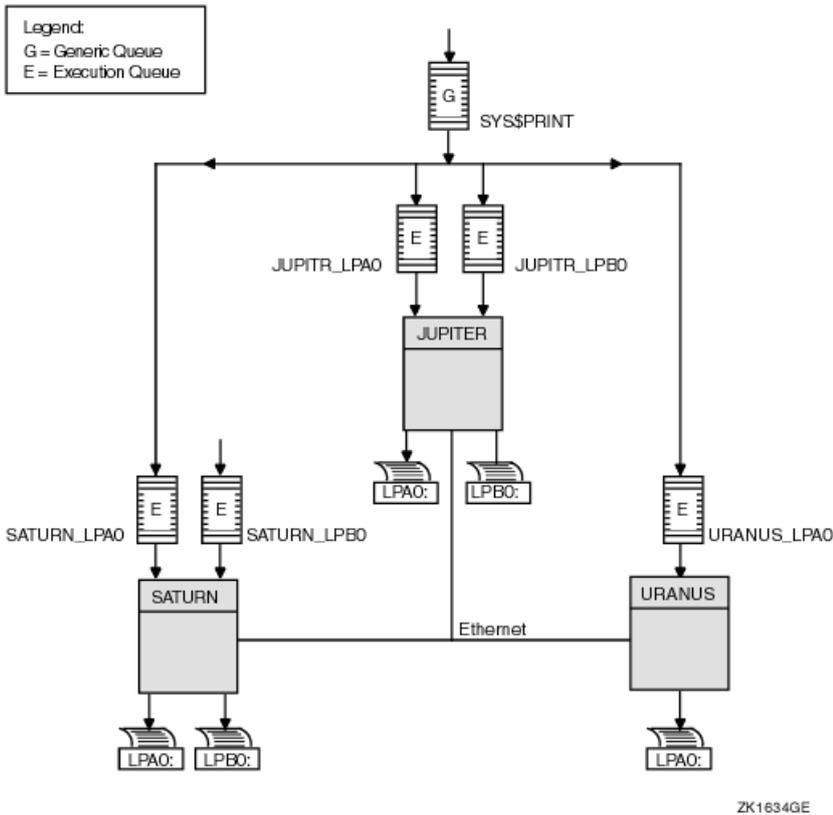
7.8. Setting Up Clusterwide Generic Print Queues

The clusterwide queue database enables you to establish generic queues that function throughout the cluster. Jobs queued to clusterwide generic queues are placed in any assigned print queue that is available, regardless of its location in the cluster. However, the file queued for printing must be accessible to the computer to which the printer is connected.

7.8.1. Sample Configuration

Figure 7.3 illustrates a clusterwide generic print queue in which the queues for all LPA0 printers in the cluster are assigned to a clusterwide generic queue named SYS\$PRINT.

A clusterwide generic print queue needs to be initialized and started only once. The most efficient way to start your queues is to create a common command procedure that is executed by each OpenVMS Cluster computer (see Section 7.12.3).

Figure 7.3. Clusterwide Generic Print Queue Configuration

7.8.2. Command Example

The following command initializes and starts the clusterwide generic queue SYSS\$PRINT:

```
$ INITIALIZE/QUEUE/GENERIC=( JUPITR_LPA0 , SATURN_LPA0 , URANUS_LPA0 ) /START SYSS$PRINT
```

Jobs queued to SYSS\$PRINT are placed in whichever assigned print queue is available. Thus, in this example, a print job from JUPITER that is queued to SYSS\$PRINT can be queued to JUPITR_LPA0, SATURN_LPA0, or URANUS_LPA0.

7.9. Setting Up Execution Batch Queues

Generally, you set up execution batch queues on each OpenVMS Cluster computer using the same procedures you use for a standalone computer. For more detailed information about how to do this, see the *VSI OpenVMS System Manager's Manual*.

7.9.1. Before You Begin

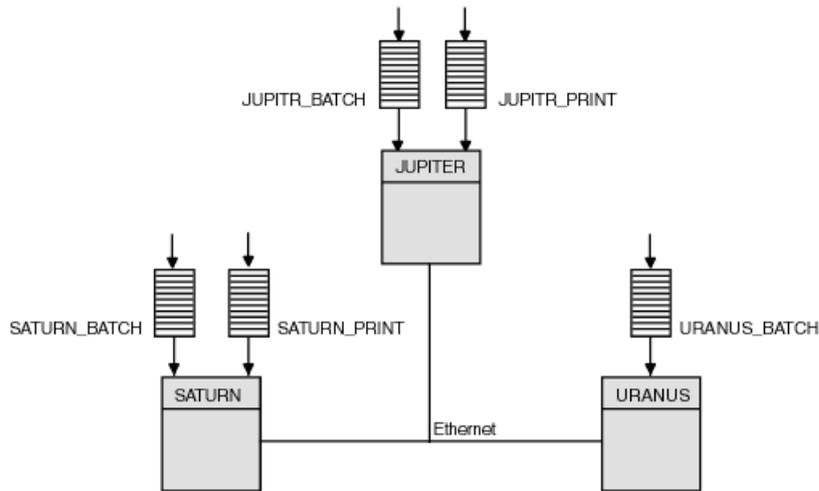
Before you establish batch queues, you should decide which type of queue configuration best suits your cluster. As system manager, you are responsible for setting up batch queues to maintain efficient batch job processing on the cluster. For example, you should do the following:

- Determine what type of processing will be performed on each computer.
- Set up local batch queues that conform to these processing needs.

- Decide whether to set up any clusterwide generic queues that will distribute batch job processing across the cluster.
- Decide whether to use autostart queues for startup simplicity.

Once you determine the strategy that best suits your needs, you can create a command procedure to set up your queues. Figure 7.4 shows a batch queue configuration for a cluster consisting of computers JUPITER, SATURN, and URANUS.

Figure 7.4. Sample Batch Queue Configuration



ZK1635GE

7.9.2. Batch Command Format

You create a batch queue with a unique name by specifying the DCL command INITIALIZE/QUEUE/BATCH in the following format

```
$ INITIALIZE/QUEUE/BATCH/ON=node : [ /START ] [ /NAME_OF_MANAGER=
name-of-manager ] queue-name
```

Qualifier	Description
/ON	Specifies the computer on which the batch queue runs.
/START	Starts the queue.
/NAME_OF_MANAGER	Specifies the name of the queue manager if you are running multiple queue managers.

7.9.3. Autostart Command Format

You can initialize and start an autostart batch queue by specifying the DCL command INITIALIZE/QUEUE/BATCH. Use the following command format:

```
INITIALIZE/QUEUE/BATCH/AUTOSTART_ON=node::queue-name
```

When you use the /AUTOSTART_ON qualifier, you must initially activate the queue for autostart, either by specifying the /START qualifier with the INITIALIZE/QUEUE command or by entering a START/QUEUE command. However, the queue cannot begin processing jobs until the ENABLE AUTOSTART/QUEUES command is entered for a node on which the queue can run.

Rule: Generic queues cannot be autostart queues. Note that you cannot specify both /ON and /AUTOSTART_ON.

7.9.4. Examples

The following commands make the local batch queue assignments for JUPITR, SATURN, and URANUS shown in Figure 7.4:

```
$ INITIALIZE/QUEUE/BATCH/ON=JUPITR::/START/NAME_OF_MANAGER=BATCH_QUEUE
  JUPITR_BATCH
$ INITIALIZE/QUEUE/BATCH/ON=SATURN::/START/NAME_OF_MANAGER=BATCH_QUEUE
  SATURN_BATCH
$ INITIALIZE/QUEUE/BATCH/ON=URANUS::/START/NAME_OF_MANAGER=BATCH_QUEUE
  URANUS_BATCH
```

Because batch jobs on each OpenVMS Cluster computer are queued to SYSS\$BATCH by default, you should consider defining a logical name to establish this queue as a clusterwide generic batch queue that distributes batch job processing throughout the cluster (see Example 7.2). Note, however, that you should do this only if you have a common-environment cluster.

7.10. Setting Up Clusterwide Generic Batch Queues

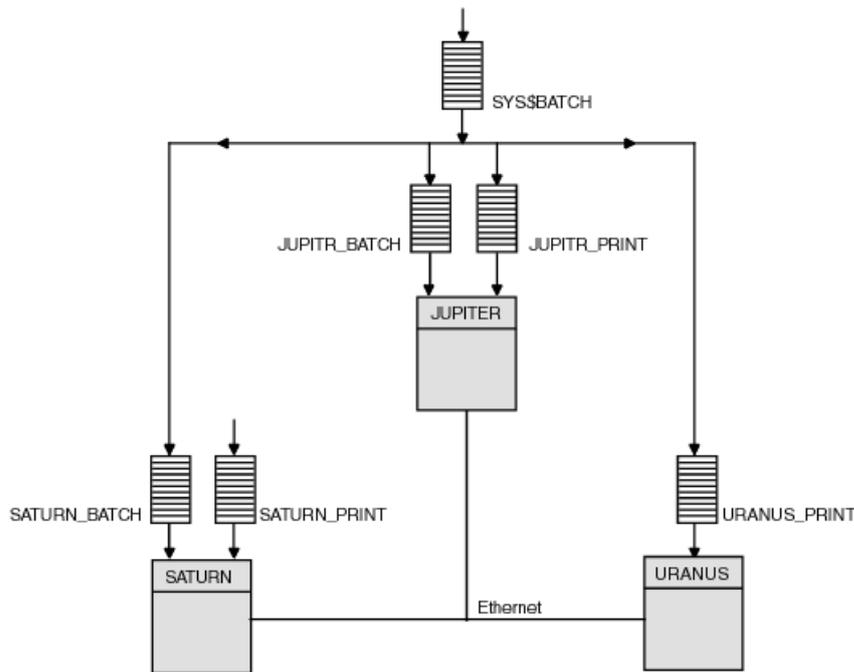
In an OpenVMS Cluster system, you can distribute batch processing among computers to balance the use of processing resources. You can achieve this workload distribution by assigning local batch queues to one or more clusterwide generic batch queues. These generic batch queues control batch processing across the cluster by placing batch jobs in assigned batch queues that are available. You can create a clusterwide generic batch queue as shown in Example 7.2.

A clusterwide generic batch queue needs to be initialized and started only once. The most efficient way to perform these operations is to create a common command procedure that is executed by each OpenVMS Cluster computer (see Example 7.2).

7.10.1. Sample Configuration

In Figure 7.5, batch queues from each OpenVMS Cluster computer are assigned to a clusterwide generic batch queue named SYSS\$BATCH. Users can submit a job to a specific queue (for example, JUPITR_BATCH or SATURN_BATCH), or, if they have no special preference, they can submit it by default to the clusterwide generic queue SYSS\$BATCH. The generic queue in turn places the job in an available assigned queue in the cluster.

If more than one assigned queue is available, the operating system selects the queue that minimizes the ratio (executing jobs/job limit) for all assigned queues.

Figure 7.5. Clusterwide Generic Batch Queue Configuration

ZK1636GE

7.11. Starting Local Batch Queues

Normally, you use local batch execution queues during startup to run batch jobs to start layered products. For this reason, these queues must be started before the `ENABLE AUTOSTART` command is executed, as shown in the command procedure in Example 7.1.

7.11.1. Startup Command Procedure

Start the local batch execution queue in each node's startup command procedure `SYSTARTUP_VMS.COM`. If you use a common startup command procedure, add commands similar to the following to your procedure:

```
$ SUBMIT/PRIORITY=255/NOIDENT/NOLOG/QUEUE=node_BATCH LAYERED_PRODUCT.COM
$ START/QUEUE node_BATCH
$ DEFINE/SYSTEM/EXECUTIVE SYS$BATCH node_BATCH
```

Submitting the startup command procedure `LAYERED_PRODUCT.COM` as a high-priority batch job before the queue starts ensures that the job is executed immediately, regardless of the job limit on the queue. If the queue is started before the command procedure was submitted, the queue might reach its job limit by scheduling user batch jobs, and the startup job would have to wait.

7.12. Using a Common Command Procedure

Once you have created queues, you must start them to begin processing batch and print jobs. In addition, you must make sure the queues are started each time the system reboots, by enabling autostart for autostart queues or by entering `START/QUEUE` commands for nonautostart queues. To do so, create a command procedure containing the necessary commands.

7.12.1. Command Procedure

You can create a common command procedure named, for example, QSTARTUP.COM, and store it on a shared disk. With this method, each node can share the same copy of the common QSTARTUP.COM procedure. Each node invokes the common QSTARTUP.COM procedure from the common version of SYSTARTUP. You can also include the commands to start queues in the common SYSTARTUP file instead of in a separate QSTARTUP.COM file.

7.12.2. Examples

Example 7.1 shows commands used to create OpenVMS Cluster queues.

Example 7.1. Sample Commands for Creating OpenVMS Cluster Queues

```

$
❶
$ DEFINE/FORM LN_FORM 10 /WIDTH=80 /STOCK=DEFAULT /TRUNCATE
$ DEFINE/CHARACTERISTIC 2ND_FLOOR 2
.
.
.
❷
$ INITIALIZE/QUEUE/AUTOSTART_ON=( JUPITR::LPA0: ) /START JUPITR_PRINT
$ INITIALIZE/QUEUE/AUTOSTART_ON=( SATURN::LPA0: ) /START SATURN_PRINT
$ INITIALIZE/QUEUE/AUTOSTART_ON=( URANUS::LPA0: ) /START URANUS_PRINT
.
.
.
❸
$ INITIALIZE/QUEUE/BATCH/START/ON=JUPITR:: JUPITR_BATCH
$ INITIALIZE/QUEUE/BATCH/START/ON=SATURN:: SATURN_BATCH
$ INITIALIZE/QUEUE/BATCH/START/ON=URANUS:: URANUS_BATCH
.
.
.
❹
$ INITIALIZE/QUEUE/START -
_ $ /AUTOSTART_ON=( JUPITR::LTA1: , SATURN::LTA1 , URANUS::LTA1 ) -
_ $ /PROCESSOR=LATSYM /FORM_MOUNTED=LN_FORM -
_ $ /RETAIN=ERROR /DEFAULT=( NOBURST , FLAG=ONE , NOTRAILER ) -
_ $ /RECORD_BLOCKING LN03$PRINT
$
$ INITIALIZE/QUEUE/START -
_ $ /AUTOSTART_ON=( JUPITR::LTA2: , SATURN::LTA2 , URANUS::LTA2 ) -
_ $ /PROCESSOR=LATSYM /RETAIN=ERROR -
_ $ /DEFAULT=( NOBURST , FLAG=ONE , NOTRAILER ) /RECORD_BLOCKING -
_ $ /CHARACTERISTIC=2ND_FLOOR LA210$PRINT
$
❺
$ ENABLE AUTOSTART/QUEUES/ON=SATURN
$ ENABLE AUTOSTART/QUEUES/ON=JUPITR
$ ENABLE AUTOSTART/QUEUES/ON=URANUS❻
$ INITIALIZE/QUEUE/START SYS$PRINT -
_ $ /GENERIC=( JUPITR_PRINT , SATURN_PRINT , URANUS_PRINT )

```

```

$
❶
$ INITIALIZE/QUEUE/BATCH/START SYSS$BATCH -
_ $ /GENERIC=( JUPITR_BATCH , SATURN_BATCH , URANUS_BATCH )
$

```

Following are descriptions of each command or group of commands in Example 7.1.

- ❶ Define all printer forms and characteristics.
- ❷ Initialize local print queues. In the example, these queues are autostart queues and are started automatically when the node executes the ENABLE AUTOSTART/QUEUES command. Although the /START qualifier is specified to activate the autostart queues, they do not begin processing jobs until autostart is enabled.

To enable autostart each time the system reboots, add the ENABLE AUTOSTART/QUEUES command to your queue startup command procedure, as shown in Example 7.2.

- ❸ Initialize and start local batch queues on all nodes, including satellite nodes. In this example, the local batch queues are not autostart queues.
- ❹ Initialize queues for remote LAT printers. In the example, these queues are autostart queues and are set up to run on one of three nodes. The queues are started on the first of those three nodes to execute the ENABLE AUTOSTART command.

You must establish the logical devices LTA1 and LTA2 in the LAT startup command procedure LAT\$\$SYSTARTUP.COM on each node on which the autostart queue can run. For more information, see the description of editing LAT\$\$SYSTARTUP.COM in the *VSI OpenVMS System Manager's Manual*.

Although the /START qualifier is specified to activate these autostart queues, they will not begin processing jobs until autostart is enabled.

- ❺ Enable autostart to start the autostart queues automatically. In the example, autostart is enabled on node SATURN first, so the queue manager starts the autostart queues that are set up to run on one of several nodes.
- ❻ Initialize and start the generic output queue SYSS\$PRINT. This is a nonautostart queue (generic queues cannot be autostart queues). However, generic queues are not stopped automatically when a system is shut down, so you do not need to restart the queue each time a node reboots.
- ❼ Initialize and start the generic batch queue SYSS\$BATCH. Because this is a generic queue, it is not stopped when the node shuts down. Therefore, you do not need to restart the queue each time a node reboots.

7.12.3. Example

Example 7.2 illustrates the use of a common QSTARTUP command procedure on a shared disk.

Example 7.2. Common Procedure to Start OpenVMS Cluster Queues

```

$!
$! QSTARTUP.COM -- Common procedure to set up cluster queues
$!
$!
❶
$ NODE = F$GETSYI("NODENAME")
$!
$! Determine the node-specific subroutine

```

```

$!
$ IF (NODE .NES. "JUPITR") .AND. (NODE .NES. "SATURN") .AND. (NODE .NES. -
"URANUS")
$   THEN
$     GOSUB SATELLITE_STARTUP
$   ELSE
❷
$!
$!   Configure remote LAT devices.
$!
$   SET TERMINAL LTA1: /PERM /DEVICE=LN03 /WIDTH=255 /PAGE=60 -
/LOWERCASE /NOBROAD
$   SET TERMINAL LTA2: /PERM /DEVICE=LA210 /WIDTH=255 /PAGE=66 -
/NOBROAD
$   SET DEVICE LTA1: /SPOOLED=(LN03$PRINT,SYS$SYSDEVICE:)
$   SET DEVICE LTA2: /SPOOLED=(LA210$PRINT,SYS$SYSDEVICE:)
❸
$   START/QUEUE/BATCH 'NODE'_BATCH
$   GOSUB 'NODE'_STARTUP
$   ENDIF
$ GOTO ENDING
$!
$! Node-specific subroutines start here
$!
❹
$ SATELLITE_STARTUP:
$!
$! Start a batch queue for satellites.
$!
$ START/QUEUE/BATCH 'NODE'_BATCH
$ RETURN
$!
❺
$JUPITR_STARTUP:
$!
$! Node-specific startup for JUPITR::
$! Setup local devices and start nonautostart queues here
$!
$ SET PRINTER/PAGE=66 LPA0:
$ RETURN
$!
$SATURN_STARTUP:
$!
$! Node-specific startup for SATURN::
$! Setup local devices and start nonautostart queues here
$!
.
.
.
$ RETURN
$!
$URANUS_STARTUP:
$!
$! Node-specific startup for URANUS::
$! Setup local devices and start nonautostart queues here
$!
.

```

```

.
.
$ RETURN
$!
$ENDING:
⑥
$! Enable autostart to start all autostart queues
$!
$ ENABLE AUTOSTART/QUEUES
$ EXIT

```

Following are descriptions of each phase of the common QSTARTUP.COM command procedure in Example 7.2.

- ① Determine the name of the node executing the procedure.
- ② On all large nodes, set up remote devices connected by the LAT. The queues for these devices are autostart queues and are started automatically when the ENABLE AUTOSTART/QUEUES command is executed at the end of this procedure.

In the example, these autostart queues were set up to run on one of three nodes. The queues start when the first of those nodes executes the ENABLEAUTOSTART/QUEUES command. The queue remains running as long as one of those nodes is running and has autostart enabled.

- ③ On large nodes, start the local batch queue. In the example, the local batch queues are nonautostart queues and must be started explicitly with START/QUEUE commands.
- ④ On satellite nodes, start the local batch queue.
- ⑤ Each node executes its own subroutine. On node JUPITR, set up the line printer device LPA0:. The queue for this device is an autostart queue and is started automatically when the ENABLE AUTOSTART/QUEUES command is executed.
- ⑥ Enable autostart to start all autostart queues.

7.13. Disabling Autostart During Shutdown

By default, the shutdown procedure disables autostart at the beginning of the shutdown sequence. Autostart is disabled to allow autostart queues with failover lists to fail over to another node. Autostart also prevents any autostart queue running on another node in the cluster to fail over to the node being shutdown.

7.13.1. Options

You can change the time at which autostart is disabled in the shutdown sequence in one of two ways:

Option	Description
1	<p>Define the logical name SHUTDOWN\$DISABLE_AUTOSTART as follows:</p> <pre>\$ DEFINE/SYSTEM/EXECUTIVE SHUTDOWN\$DISABLE_AUTOSTART number-of-minutes</pre> <p>Set the value of <code>number-of-minutes</code> to the number of minutes before shutdown when autostart is to be disabled. You can add this logical name definition to SYLOGICALS.COM. The value of <code>number-of-minutes</code> is the default value for</p>

Option	Description
	the node. If this number is greater than the number of minutes specified for the entire shutdown sequence, autostart is disabled at the beginning of the sequence.
2	Specify the <code>DISABLE_AUTOSTART number-of-minutes</code> option during the shutdown procedure. (The value you specify for <code>number-of-minutes</code> overrides the value specified for the <code>SHUTDOWN\$DISABLE_AUTOSTART</code> logical name).

Reference: See the *VSI OpenVMS System Manager's Manual* for more information about changing the time at which autostart is disabled during the shutdown sequence.

Chapter 8. Configuring an OpenVMS Cluster System

This chapter provides an overview of the cluster configuration command procedures and describes the preconfiguration tasks required before running either command procedure. Then it describes each major function of the command procedures and the postconfiguration tasks, including running AUTOGEN.COM.

8.1. Overview of the Cluster Configuration Procedures

Two similar command procedures are provided for configuring and reconfiguring an OpenVMS Cluster system: CLUSTER_CONFIG_LAN.COM and CLUSTER_CONFIG.COM. The choice depends on whether you use the LANCP utility or DECnet for satellite booting in your cluster. CLUSTER_CONFIG_LAN.COM provides satellite booting services with the LANCP utility; CLUSTER_CONFIG.COM provides satellite booting services with DECnet.

Also, to configure an Integrity server system use CLUSTER_CONFIG_LAN.COM and to configure an Alpha system, use either CLUSTER_CONFIG_LAN.COM or CLUSTER_CONFIG.COM. You can use only CLUSTER_CONFIG_LAN.COM for configuring Cluster over IP.

In a satellite environment, you may want to determine which command procedure is used for configuring a cluster. To determine whether CLUSTER_CONFIG or CLUSTER_CONFIG_LAN is used in cluster configuration, see the SYS\$SYSTEM:MODPARAMS.DAT file. While configuring a cluster, the command procedure name is added as a comment in the MODPARAMS.DAT file.

During the ADD operation, a comment similar to the following is added to MODPARAMS.DAT for CLUSTER_CONFIG:

```
! CLUSTER_CONFIG creating for ADD operation on 4-APR-2009 14:21:00.89
```

For CLUSTER_CONFIG_LAN:

```
! CLUSTER_CONFIG_LAN creating for ADD operation on 5-APR-2009 14:21:00.89
```

Similar comments are added for the 'CHANGE' operation. For multiple entries in MODPARAMS.DAT, the last entry reflects the latest procedure name that is used to configure the cluster. See Section 4.5 for the factors to consider when choosing a satellite booting service.

These configuration procedures automate most of the tasks required to configure an OpenVMS Cluster system. When you invoke CLUSTER_CONFIG_LAN.COM or CLUSTER_CONFIG.COM, the following configuration options are displayed:

- Add a computer to the cluster
- Remove a computer from the cluster
- Change a computer's characteristics
- Create a duplicate system disk
- Make a directory structure for a new root on a system disk
- Delete a root from a system disk

By selecting the appropriate option, you can configure the cluster easily and reliably without invoking any OpenVMS utilities directly. Table 8.1 summarizes the functions that the configuration procedures perform for each configuration option.

The phrase **cluster configuration command procedure**, when used in this chapter, refers to both CLUSTER_CONFIG_LAN.COM and CLUSTER_CONFIG.COM. The questions of the two configuration procedures are identical except where they pertain to LANCP and DECnet.

Note: For help on any question in these command procedures, type a question mark (?) at the question.

Table 8.1. Summary of Cluster Configuration Functions

Option	Functions Performed
ADD	<p>Enables a node as a cluster member:</p> <ul style="list-style-type: none"> • Establishes the new computer's root directory on a cluster common system disk and generates the computer's system parameter files, (IA64VMSSYS.PAR for Integrity server systems or ALPHAVMSSYS.PAR for Alpha systems), and MODPARAMS.DAT in its SYSS\$SPECIFIC:[SYSEXE] directory. • Generates the new computer's page and swap files (PAGEFILE.SYS and SWAPFILE.SYS). • Sets up a cluster quorum disk (optional). • Sets disk allocation class values, or port allocation class values (Alpha only), or both, with the ALLOCLASS parameter for the new computer, if the computer is being added as a disk server. If the computer is being added as a tape server, sets a tape allocation class value with the TAPE_ALLOCLASS parameter. <p>Note: ALLOCLASS must be set to a value greater than zero if you are configuring an Alpha computer on a shared SCSI bus and you are not using a portal location class.</p> <ul style="list-style-type: none"> • Generates an initial (temporary) startup procedure for the new computer. This initial procedure: <ul style="list-style-type: none"> • Runs NETCONFIG.COM to configure the network. • Runs AUTOGEN to set appropriate system parameter values for the computer. • Reboots the computer with normal startup procedures. • If the new computer is a satellite node, the configuration procedure updates: <ul style="list-style-type: none"> • Network databases for the computer on which the configuration procedure is executed to add the new computer. • SYSS\$MANAGER:NETNODE_UPDATE.COM command procedure on the local computer (as described in Section 10.4.2).

Option	Functions Performed
REMOVE	Disables a node as a cluster member: <ul style="list-style-type: none"> • Deletes another computer's root directory and its contents from the local computer's system disk. If the computer being removed is a satellite, the cluster configuration command procedure updates SYS \$MANAGER:NETNODE_UPDATE.COM on the local computer. • Updates the permanent and volatile remote node network databases on the local computer. • Removes the quorum disk.
CHANGE	Displays the CHANGE menu and prompts for appropriate information to: <ul style="list-style-type: none"> • Enable or disable the local computer as a disk server • Enable or disable the local computer as a boot server • Enable or disable IP for cluster communications on the local computer • Enable or disable the Ethernet or FDDI LAN for cluster communications on the local computer • Enable or disable a quorum disk on the local computer • Change a satellite's Ethernet or FDDI hardware address • Enable or disable the local computer as a tape server • Change the local computer's ALLOCLASS or TAPE_ALLOCLASS value • Change the local computer's shared SCSI port allocation class value • Enable or disable MEMORY CHANNEL for node-to-node cluster communications on the local computer
CREATE	Duplicates the local computer's system disk and removes all system roots from the new disk.
MAKE	Creates a directory structure for a new root on a system disk.
DELETE	Deletes a root from a system disk.

8.1.1. Before Configuring the System

Before invoking either the CLUSTER_CONFIG_LAN.COM or the CLUSTER_CONFIG.COM procedure to configure an OpenVMS Cluster system, perform the tasks described in Table 8.2.

Table 8.2. Preconfiguration Tasks

Task	Procedure
Determine whether the computer uses DECdtm.	When you add a computer to or remove a computer from a cluster that uses DECdtm services, there are a number of tasks you need to do in order to ensure the integrity of your data.

Task	Procedure
	<p>Reference: See the chapter about DECdtm services in the <i>VSI OpenVMS System Manager's Manual</i> for step-by-step instructions on setting up DECdtm in an OpenVMS Cluster system.</p> <p>If you are not sure whether your cluster uses DECdtm services, enter this command sequence:</p> <pre>\$ SET PROCESS /PRIVILEGES=SYSPRV \$ RUN SYS\$SYSTEM:LMCP LMCP> SHOW LOG</pre> <p>If your cluster does not use DECdtm services, the SHOW LOG command will display a “file not found” error message. If your cluster uses DECdtm services, it displays a list of the files that DECdtm uses to store information about transactions.</p>
<p>Ensure the network software providing the satellite booting service is up and running and all computers are connected to the LAN.</p>	<p>For nodes that will use the LANCP utility for satellite booting, run the LANCP utility and enter the LANCP command LIST DEVICE/MOPDLL to display a list of LAN devices on the system:</p> <pre>\$ RUN SYS\$SYSTEM:LANCP LANCP> LIST DEVICE/MOPDLL</pre> <p>For nodes running DECnet for OpenVMS, enter the DCL command SHOW NETWORK to determine whether the network is up and running:</p> <pre>\$ SHOW NETWORK Product: DECNET Node: CHBOSE Address(es): 25.169 Product: TCP/IP Node: chbose.ind.hp.com Address(es): 18.156.235.23</pre> <p>This example shows that the node CHBOSE is running DECnet for OpenVMS and node chbose.ind.hp.com is running TCP/IP. If DECnet has not been started, the message “SHOW-I-NONET, Network Unavailable” is displayed.</p> <p>For nodes running DECnet-Plus, refer to <i>VSI OpenVMS DECnet Network Management Utilities</i> for information about determining whether the DECnet-Plus network is up and running.</p>
<p>Select MOP and disk servers.</p>	<p>Every OpenVMS Cluster configured with satellite nodes must include at least one Maintenance Operations Protocol (MOP) and disk server. When possible, select multiple computers as MOP and disk servers. Multiple servers give better availability, and they distribute the work load across more LAN adapters.</p> <p>Follow these guidelines when selecting MOP and disk servers:</p> <ul style="list-style-type: none"> • Ensure that MOP servers have direct access to the system disk. • Ensure that disk servers have direct access to the storage that they are serving.

Task	Procedure
	<ul style="list-style-type: none"> Choose the most powerful computers in the cluster. Low-powered computers can become overloaded when serving many busy satellites or when many satellites boot simultaneously. Note, however, that two or more moderately powered servers may provide better performance than a single high-powered server. If you have several computers of roughly comparable power, it is reasonable to use them all as boot servers. This arrangement gives optimal load balancing. In addition, if one computer fails or is shut down, others remain available to serve satellites. After compute power, the most important factor in selecting a server is the speed of its LAN adapter. Servers should be equipped with the highest-bandwidth LAN adapters in the cluster. If you are interconnecting the cluster using IP, note the local LAN adapter on which the IP address will be configured and used for SCS.
Make sure you are logged in to a privileged account.	<p>Log in to a privileged account.</p> <p>Rules: If you are adding a satellite, you must be logged into the system manager's account on a boot server. Note that the process privileges SYSPRV, OPER, CMKRNL, BYPASS, and NETMBX are required, because the procedure performs privileged system operations.</p>
Coordinate cluster common files.	If your configuration has two or more system disks, follow the instructions in Chapter 5 to coordinate the cluster common files.
Optionally, disable broadcast messages to your terminal.	While adding and removing computers, many such messages are generated. To disable the messages, you can enter the DCL command <code>REPLY/DISABLE=(NETWORK, CLUSTER)</code> . See also Section 10.5 for more information about controlling OPCOM messages.
Predetermine answers to the questions asked by the cluster configuration procedure.	Table 8.3 describes the data requested by the cluster configuration command procedures.

8.1.2. Data Requested by the Cluster Configuration Procedures

The following table describes the questions asked by the cluster configuration command procedures and describes how you might answer them. The table is supplied here so that you can determine answers to the questions before you invoke the procedure.

Because many of the questions are configuration specific, Table 8.3 lists the questions according to configuration type, and not in the order they are asked.

Table 8.3. Data Requested by CLUSTER_CONFIG_LAN.COM and CLUSTER_CONFIG.COM

Information Required	How to Specify or Obtain
For all configurations	

Information Required	How to Specify or Obtain
Device name of cluster system disk on which root directories will be created	Press Return to accept the default device name which is the translation of the SYSSYSDEVICE: logical name, or specify a logical name that points to the common system disk.
Computer's root directory name on cluster system disk	<p>Press Return to accept the procedure-supplied default, or specify a name in the form SYS x:</p> <ul style="list-style-type: none"> • For both Integrity servers and Alpha systems with direct access to the system disk, the valid range of hexadecimal values is much larger. It includes both the VAX range of 1 through 9 or A through D, and also the range 10 through FFFF. Note that SYSE and SYSF are reserved for system use. • For satellites, x must be in the range of 10 through FFFF.
Workstation windowing system	System manager specifies. Workstation software must be installed before workstation satellites are added. If it is not, the procedure indicates that fact.
Location and sizes of page and swap files	<p>This information is requested only when you add a computer to the cluster. Press Return to accept the default size and location. (The default sizes displayed in brackets by the procedure are minimum values. The default location is the device name of the cluster system disk).</p> <p>If your configuration includes satellite nodes, you may realize a performance improvement by locating satellite page and swap files on a satellite's local disk, if such a disk is available. The potential for performance improvement depends on the configuration of your OpenVMS Cluster system disk and network.</p> <p>To set up page and swap files on a satellite's local disk, the cluster configuration procedure creates a command procedure called SATELLITE_PAGE.COM in the satellite's [SYS n.SYSEX] directory on the boot server's system disk. The SATELLITE_PAGE.COM procedure performs the following functions:</p> <ul style="list-style-type: none"> • Mounts the satellite's local disk with a volume label that is unique in the cluster in the format <i>node-name</i>_SCSSYSTEMID. <p>Reference: Refer to Section 8.6.5 for information about altering the volume label.</p> <ul style="list-style-type: none"> • Installs the page and swap files on the satellite's local disk. <p>Note: For page and swap disks that are shadowed, you must edit the MOUNT and INIT commands in SATELLITE_PAGE.COM to the appropriate syntax for mounting any specialized "local" disks (that is, host-based shadowing disks (DS xxx), or host-based RAID disks (DP xxx), or DECram disks (MDA xxx)) on the newly added node. CLUSTER_CONFIG(_LAN).COM does not create the MOUNT and INIT commands required for SHADOW, RAID, or DECram disks.</p>

Information Required	How to Specify or Obtain
	<p>Note: To relocate the satellite's page and swap files (for example, from the satellite's local disk to the boot server's system disk, or the reverse) or to change file sizes:</p> <ol style="list-style-type: none"> 1. Create new PAGE and SWAP files on a shared device, as shown: <pre data-bbox="592 439 1414 495">\$ MCR SYSGEN CREATE device:[dir] PAGEFILE.SYS/SIZE= block-count</pre> <p>Note: If page and swap files will be created for a shadow set, you must edit SATELLITE_PAGE accordingly.</p> 2. Rename the SYSS\$SPECIFIC:[SYSEXE]PAGEFILE.SYS and SWAPFILE.SYS files to PAGEFILE.TMP and SWAPFILE.TMP. 3. Reboot, and then delete the .TMP files. 4. Modify the SYSS\$MANAGER:SYPAGSWPFILES.COM procedure to load the files.
Value for local computer's allocation class (ALLOCLASS or TAPE_ALLOCLASS) parameter.	The ALLOCLASS parameter can be used for a node allocation class or, on Alpha computers, a port allocation class. Refer to Section 6.2.1 for complete information about specifying allocation classes.
Physical device name of quorum disk	System manager specifies.
For systems running DECnet for OpenVMS	
Computer's DECnet node address for Phase IV	<p>For the DECnet node address, you obtain this information as follows:</p> <ul style="list-style-type: none"> • If you are adding a computer, the network manager supplies the address. • If you are removing a computer, use the SHOW NETWORK command (as shown in Table 8.2).
Computer's DECnet node name	Network manager supplies. The name must be from 1 to 6 alphanumeric characters and <i>cannot</i> include dollar signs (\$) or underscores (_).
For systems running DECnet-Plus	
Computer's DECnet node address for Phase IV (if you need Phase IV compatibility)	<p>For the DECnet node address, you obtain this information as follows:</p> <ul style="list-style-type: none"> • If you are adding a computer, the network manager supplies the address. • If you are removing a computer, use the SHOW NETWORK command (as shown in Table 8.2).
Node's DECnet full name	<p>Determine the full name with the help of your network manager. Enter a string comprised of:</p> <ul style="list-style-type: none"> • The namespace name, ending with a colon (:). This is optional. • The root directory, designated by a period (.).

Information Required	How to Specify or Obtain
	<ul style="list-style-type: none"> • Zero or more hierarchical directories, designated by a character string followed by a period (.). • The simple name, a character string that, combined with the directory names, uniquely identifies the node. For example: .SALES.NETWORKS.MYNODE MEGA:.INDIANA.JONES COLUMBUS:.FLATWORLD
SCS node name for this node	Enter the OpenVMS Cluster node name, which is a string of 6 or fewer alphanumeric characters.
DECnet synonym	Press Return to define a DECnet synonym, which is a short name for the node's full name. Otherwise, enter N.
Synonym name for this node	<p>Enter a string of 6 or fewer alphanumeric characters. By default, it is the first 6 characters of the last simple name in the full name. For example:</p> <p>Synonym: BLACKH</p> <p>Note: The node synonym does not need to be the same as the OpenVMS Cluster node name.</p>
MOP service client name for this node	Enter the name for the node's MOP service client when the node is configured as a boot server. By default, it is the OpenVMS Cluster node name (for example, the SCS node name). This name does not need to be the same as the OpenVMS Cluster node name.
For systems running TCP/IP or the LANCP Utility for satellite booting, or both	
Computer's SCS node name (SCSNODE) and SCS system ID (SCSSYSTEMID)	These prompts are described in Section 4.2.3. If a system is running TCP/IP, the procedure does not ask for a TCP/IP host name because a cluster node name (SCSNODE) does not have to match a TCP/IP host name. The TCP/IP host name might be longer than six characters, whereas the SCSNODE name must be no more than six characters. Note that if the system is running both DECnet and IP, then the procedure uses the DECnet defaults.
For LAN configurations	
Cluster group number and password	This information is requested only when the CHANGE option is chosen. See Section 2.5 for information about assigning cluster group numbers and passwords.
Satellite's LAN hardware address	<p>Address has the form xx-xx-xx-xx-xx-xx. You must include the hyphens when you specify a hardware address. For getting the hardware address, execute the following command at the satellite's console:</p> <p>On Integrity servers:</p> <pre>Shell> lanaddress</pre> <p>On Alpha systems:</p> <pre>>>> SHOW NETWORK</pre>

Information Required	How to Specify or Obtain
	These commands display the hardware address of the LAN devices that can be used for satellite booting. Note that you can also use the SHOW CONFIG command at LANCP.
For IP configurations	
UDP port number	UDP port number is the port number used for cluster communication. UDP port number must be same on all members of the cluster. Also, ensure that there is no other cluster in your environment using the same UDP port number or this port number is used by any other application.
IP multicast address	<p>Enter the IP multicast address for cluster, if IP multicasting is enabled. By default, the IP multicast address is selected from the administratively scoped IP multicast address range of 239.242.x.y. The last two octets x and y are generated based on the cluster group number. For example, if the cluster group number is 1985, the multicast is calculated as follows:</p> $X = 1985 / 256 \quad Y = 1985 - (256 * X)$ <p>The system administrator can override the default multicast address with a unique address for their environment.</p>
IP unicast address	If the node that is configured uses IP unicast to discover a remote node, you need the IP unicast address of the existing members or any new member in the cluster.
IP address	It is the IP address of the local system from where the cluster is configured.
Gateway and Network mask address	In the configuration option, select option 4 to add the TCP/IP gateway and network mask address to the cluster over IP database.
IP interface address	<p>In the configuration option for the selected address, select option 4 to add to the cluster over IP database. The interface information along with the default route is entered in the TCPIP\$CLUSTER.DAT as shown in the following example:</p> <pre>interface=IE0,EIA0,10.0.1.2,255.255.255.0 default_route=10.0.1.1</pre>
IP interface address for satellite booting	<p>To select the IP interface to be used for satellite booting.</p> <p>For Alpha systems:</p> <ul style="list-style-type: none"> Execute the following command at the satellite's console: <pre>>>> SHOW DEVICE</pre> <p>From the output, the LAN interface will be EIA0 on which IP address will be configured and used for cluster configuration.</p> <p>For Integrity server systems:</p>

Information Required	How to Specify or Obtain
	<ul style="list-style-type: none"> The IP interface name will either start from 'EI' or 'EW'. If it is the first interface, it will be EIA0 or EWA0. Note the mac address of the interface that you want to use from the Shell prompt. <p>To get the interface information on Integrity servers, execute the following command on EFI Shell:</p> <pre>Shell> lanaddress</pre> <p>Assume the interface which is active is EIA0. Configure the satellite with EIA0, if it does not boot with EIA0, try with EWA0 subsequently.</p>

8.1.3. Invoking the Procedure

Once you have made the necessary preparations, you can invoke the cluster configuration procedure to configure your OpenVMS Cluster system. Log in to the system manager account and make sure your default is SYSS\$MANAGER. Then, invoke the procedure at the DCL command prompt as follows:

```
$ @CLUSTER_CONFIG_LAN
```

or

```
$ @CLUSTER_CONFIG
```

Caution: Do not invoke multiple sessions simultaneously. You can run only one cluster configuration session at a time.

Once invoked, both procedures display the following information and menu. (The only difference between CLUSTER_CONFIG_LAN.COM and CLUSTER_CONFIG.COM at this point is the command procedure name that is displayed.) Depending on the menu option you select, the procedure interactively requests configuration information from you. (Predetermine your answers as described in Table 8.3).

```
Cluster/IPCI Configuration Procedure
CLUSTER_CONFIG_LAN Version V2.84
Executing on an IA64 System
```

```
DECnet Phase IV is installed on this node.
```

```
IA64 satellites will use TCP/IP BOOTP and TFTP services for downline
loading.
```

```
TCP/IP is installed and running on this node.
```

```
Enter a "?" for help at any prompt. If you are familiar with
the execution of this procedure, you may want to mute extra notes
and explanations by invoking it with "@CLUSTER_CONFIG_LAN BRIEF".
```

```
CALVIN is an IA64 system and currently a member of a cluster
so the following functions can be performed:
```

MAIN Menu

1. ADD an IA64 node to the cluster.
2. REMOVE a node from the cluster.
3. CHANGE a cluster member's characteristics.
4. CREATE a duplicate system disk for CALVIN.

5. MAKE a directory structure for a new root on a system disk.
6. DELETE a root from a system disk.
7. EXIT from this procedure.

Enter choice [7]:

```
.
.
.
.
.
```

This chapter contains a number of sample sessions showing how to run the cluster configuration procedures. Although the CLUSTER_CONFIG_LAN.COM and the CLUSTER_CONFIG.COM procedure function the same for Integrity server systems and Alpha systems, the questions and format may appear slightly different according to the type of computer system.

8.2. Adding Computers

In most cases, you invoke either CLUSTER_CONFIG_LAN.COM or CLUSTER_CONFIG.COM on an active OpenVMS Cluster computer and select the ADD function to enable a computer as an OpenVMS Cluster member. However, in some circumstances, you may need to perform extra steps to add computers. Use the information in Table 8.4 to determine your actions.

Table 8.4. Preparing to Add Computers to an OpenVMS Cluster

IF...	THEN...
You are adding your first satellite node to the OpenVMS Cluster.	Follow these steps: <ol style="list-style-type: none"> 1. Log in to the computer that will be enabled as the cluster boot server. 2. Invoke the cluster configuration procedure, and execute the CHANGE function described in Section 8.4 to enable the local computer as a boot server. 3. After the CHANGE function completes, execute the ADD function to add satellites to the cluster.
The cluster uses DECdtm services.	You must create a transaction log for the computer when you have configured it into your cluster. For step-by-step instructions on how to do this, see the chapter on DECdtm services in the <i>VSI OpenVMS System Manager's Manual</i> .
You add a CI connected computer that boots from a cluster common system disk.	You must create a new default bootstrap command procedure for the computer before booting it into the cluster. For instructions, refer to your computer-specific installation and operations guide.
You are adding computers to a cluster with more than one common system disk.	You must use a different device name for each system disk on which computers are added. For this reason, the cluster configuration procedure supplies as a default device name the logical volume name (for example, DISK\$MARS_SYS1) of SYSSYSDEVICE: on the local system. Using different device names ensures that each computer added has a unique root directory specification, even if the system disks contain roots with the same name—for example, DISK\$MARS_SYS1:[SYS10] and DISK\$MARS_SYS2:[SYS10].

IF...	THEN...
You add a voting member to the cluster.	You must, after the ADD function completes, reconfigure the cluster according to the instructions in Section 8.6.

Caution: If either the local or the new computer fails before the ADD function completes, you must, after normal conditions are restored, perform the REMOVE option to erase any invalid data and then restart the ADD option. Section 8.3 describes the REMOVE option.

8.2.1. Controlling Conversational Bootstrap Operations

When you add a satellite to the cluster using either cluster configuration command procedure, the procedure asks whether you want to allow conversational bootstrap operations for the satellite (default is No).

If you select the default, the NISCS_CONV_BOOT system parameter in the satellite's system parameter file remains set to 0 to disable such operations. The parameter file (IA64VMSSYS.PAR for Integrity servers or ALPHAVMSSYS.PAR for Alpha systems) resides in the satellite's root directory on a boot server's system disk (*device:[SYS.x.SYSEXE]*). You can enable conversational bootstrap operations for a given satellite at any time by setting this parameter to 1.

Example:

To enable such operations for an OpenVMS Alpha satellite booted from root 10 on device \$1\$DGA11, you would proceed as follows:

Step	Action
1	Log in as system manager on the boot server.
2	On Integrity servers or Alpha systems, invoke the System Generation utility (SYSGEN) and enter the following commands: <pre>\$ RUN SYS\$SYSTEM:SYSGEN SYSGEN> USE \$1\$DGA11:[SYS10.SYSEXE]ALPHAVMSSYS.PAR SYSGEN> SET NISCS_CONV_BOOT 1 SYSGEN> WRITE \$1\$DGA11:[SYS10.SYSEXE]ALPHAVMSSYS.PAR SYSGEN> EXIT \$</pre>
3	Modify the satellite's MODPARAMS.DAT file so that NISCS_CONV_BOOT is set to 1.

8.2.2. Common AUTOGEN Parameter Files

When adding a node or a satellite to an OpenVMS Cluster, the cluster configuration command procedure adds one of the following lines in the MODPARAMS.DAT file:

WHEN the node being added is a...	THEN...
Satellite node	The following line is added to the MODPARAMS.DAT file: <pre>AGEN\$INCLUDE_PARAMS SYS\$MANAGER:AGEN \$NEW_SATELLITE_DEFAULTS.DAT</pre>
Nonsatellite node	The following line is added to the MODPARAMS.DAT file:

WHEN the node being added is a...	THEN...
	AGEN\$INCLUDE_PARAMS SYS\$MANAGER:AGEN \$NEW_NODE_DEFAULTS.DAT

The AGEN\$NEW_SATELLITE_DEFAULTS.DAT and AGEN\$NEW_NODE_DEFAULTS.DAT files hold AUTOGEN parameter settings that are common to all satellite nodes or nonsatellite nodes in the cluster. Use of these files simplifies system management, because you can maintain common system parameters in either one or both of these files. When adding or changing the common parameters, this eliminates the need to make modifications in the MODPARAMS.DAT files located on every node in the cluster.

Initially, these files contain no parameter settings. You edit the AGEN \$NEW_SATELLITE_DEFAULTS.DAT and AGEN\$NEW_NODE_DEFAULTS.DAT files, as appropriate, to add, modify, or edit system parameters. For example, you might edit the AGEN \$NEW_SATELLITE_DEFAULTS.DAT file to set the MIN_GBLPAGECNT parameter to 5000. AUTOGEN makes the MIN_GBLPAGECNT parameter and all other parameter settings in the AGEN \$NEW_SATELLITE_DEFAULTS.DAT file common to all satellite nodes in the cluster.

AUTOGEN uses the parameter settings in the AGEN\$NEW_SATELLITE_DEFAULTS.DAT or AGEN\$NEW_NODE_DEFAULTS.DAT files the first time it is run, and with every subsequent execution.

8.2.3. Examples

Examples 8.1, 8.2, and 8.3 describes the use of CLUSTER_CONFIG_LAN.COM on BHAGAT to add, respectively, a boot server running DECnet for OpenVMS, a boot server running DECnet-Plus, and a satellite node.

This section also illustrates the use of CLUSTER_CONFIG_LAN.COM to create and configure a two node cluster using IPCI, to add a new node to an IPCI cluster, to add a new node to an IPCI cluster with a Shared System Disk, and to add an Integrity server satellite node to an IPCI cluster.

Example 8.1. Sample Interactive CLUSTER_CONFIG_LAN.COM Session to Add a Computer as a Boot Server

```

$ @CLUSTER_CONFIG_LAN.COM
      Cluster/IPCI Configuration Procedure
      CLUSTER_CONFIG_LAN Version V2.84
      Executing on an IA64 System

      DECnet-Plus is installed on this node.
      IA64 satellites will use TCP/IP BOOTP and TFTP services for downline
loading.
      TCP/IP is installed and running on this node.

      Enter a "?" for help at any prompt.  If you are familiar with
      the execution of this procedure, you may want to mute extra notes
      and explanations by invoking it with "@CLUSTER_CONFIG_LAN BRIEF".

      BHAGAT is an IA64 system and currently a member of a cluster
      so the following functions can be performed:

```

MAIN Menu

1. ADD an IA64 node to the cluster.

2. REMOVE a node from the cluster.
3. CHANGE a cluster member's characteristics.
4. CREATE a duplicate system disk for BHAGAT.
5. MAKE a directory structure for a new root on a system disk.
6. DELETE a root from a system disk.
7. EXIT from this procedure.

Enter choice [7]: 1

This ADD function will add a new IA64 node to the cluster.

WARNING: If the node being added is a voting member, EXPECTED_VOTES for every cluster member must be adjusted. For complete instructions check the section on configuring a cluster in the "OpenVMS Cluster Systems" manual.

CAUTION: If this cluster is running with multiple system disks and common system files will be used, please, do not proceed unless appropriate logical names are defined for cluster common files in SYLOGICALS.COM. For instructions, refer to the "OpenVMS Cluster Systems" manual.

If this cluster will run IPCI, then TCP/IP installed on the system should be version 5.7 and above or else IPCI configuration will be aborted.

Do you want to continue [Y]?[Return]

Is the node to be a clustered node with a shared SCSI/FIBRE-CHANNEL bus (Y/N)? Y

Will the node be a satellite [Y]? N

What is the node's SCS node name? MOON

What is the node's SCSSYSTEMID number? 24.123

NOTE: 24.123 equates to an SCSSYSTEMID of 24699

Will MOON be a boot server [Y]? [Return]

TCP/IP BOOTP and TFTP services must be enabled on IA64 boot nodes.

Use SYS\$MANAGER:TCPIP\$CONFIG.COM on MOON to enable BOOTP and TFTP service after MOON has booted into the cluster.

This procedure will now ask you for the device name of MOON's system root.

The default device name (DISK\$BHAGAT_831H1:) is the logical volume name of SYS\$SYSDEVICE:.

What is the device name for MOON's system root [default DISK \$BHAGAT_831H1:]?

What is the name of MOON's system root [SYS1]? [Return]

Creating directory tree SYS1 ...

System root SYS1 created

ENABLE IP for cluster communications (Y/N)? N

CAUTION: If you do not define port allocation classes later in this procedure for shared SCSI buses, all nodes sharing a SCSI bus must have the same non-zero ALLOCLASS value. If multiple nodes connect to a shared SCSI bus without the same allocation class for the bus, system booting will halt due to the error or IO AUTOCONFIGURE after boot will keep the bus offline.

WARNING: If BHAGAT is sharing the same SCSI bus with MOON, then BHAGAT's ALLOCLASS parameter or port allocation class for the shared bus must be changed from 0 to the same non-zero value that will be entered for MOON. Use the CHANGE option of CLUSTER_CONFIG_LAN.COM to change BHAGAT's ALLOCLASS parameter before MOON is booted.

Enter a value for MOON's ALLOCLASS parameter [1]: [Return]
 Does this cluster contain a quorum disk [N]? [Return]
 Size of pagefile for MOON [RETURN for AUTOGEN sizing]? [Return]

A temporary pagefile will be created until resizing by AUTOGEN. The default size below is arbitrary and may or may not be appropriate.

Size of temporary pagefile [10000]? [Return]
 Size of swap file for MOON [RETURN for AUTOGEN sizing]? [Return]

A temporary swap file will be created until resizing by AUTOGEN. The default size below is arbitrary and may or may not be appropriate.

Size of temporary swap file [8000]? [Return]

Each shared SCSI bus must have a positive allocation class value. A shared bus uses a PK adapter. A private bus may use: PK, DR, DV, DQ.

When adding a node with SCSI-based cluster communications, the shared SCSI port allocation classes may be established in SYS\$DEVICES.DAT. Otherwise, the system's disk allocation class will apply.

A private SCSI bus need not have an entry in SYS\$DEVICES.DAT. If it has an entry, its entry may assign any legitimate port allocation class value:

n where n = a positive integer, 1 to 32767 inclusive
 0 no port allocation class and disk allocation class does not apply
 -1 system's disk allocation class applies (system parameter ALLOCLASS)

When modifying port allocation classes, SYS\$DEVICES.DAT must be updated for all affected nodes, and then all affected nodes must be rebooted. The following dialog will update SYS\$DEVICES.DAT on MOON.

Enter [RETURN] to continue:

There are currently no entries in SYS\$DEVICES.DAT for MOON. After the next boot, any SCSI controller on MOON will use MOON's disk allocation class.

Assign port allocation class to which adapter [RETURN for none]: [Return]
 Will a disk local only to MOON (and not accessible at this time to BHAGAT)
 be used for paging and swapping (Y/N)? N

If you specify a device other than DISK\$BHAGAT_831H1: for MOON's
 page and swap files, this procedure will create PAGEFILE_MOON.SYS
 and SWAPFILE_MOON.SYS in the <SYSEXE> directory on the device you
 specify.

What is the device name for the page and swap files [DISK\$BHAGAT_831H1:]?
 %SYSGEN-I-CREATED, BHAGAT\$DKA100:<SYS1.SYSEXE>PAGEFILE.SYS;1 created
 %SYSGEN-I-CREATED, BHAGAT\$DKA100:<SYS1.SYSEXE>SWAPFILE.SYS;1 created
 The configuration procedure has completed successfully.

MOON has been configured to join the cluster.
 The first time MOON boots, AUTOGEN.COM will run automatically.

Example 8.2. Sample Interactive CLUSTER_CONFIG_LAN.COM Session to Add a Computer Running DECnet-Plus

```
$ @CLUSTER_CONFIG.COM
      Cluster/IPCI Configuration Procedure
      CLUSTER_CONFIG_LAN Version V2.84
      Executing on an Alpha System

DECnet-Plus is installed on this node.
Alpha satellites will use LANCP, not DECnet, for MOP downline loading.

      Enter a "?" for help at any prompt.  If you are familiar with
      the execution of this procedure, you may want to mute extra notes
      and explanations by invoking it with "@CLUSTER_CONFIG_LAN BRIEF".

BISMIL is an Alpha system and currently a member of a cluster
so the following functions can be performed:
```

MAIN Menu

1. ADD an Alpha node to the cluster.
2. REMOVE a node from the cluster.
3. CHANGE a cluster member's characteristics.
4. CREATE a duplicate system disk for BISMIL.
5. MAKE a directory structure for a new root on a system disk.
6. DELETE a root from a system disk.
7. EXIT from this procedure.

Enter choice [7]: 1

This ADD function will add a new Alpha node to the cluster.

WARNING: If the node being added is a voting member, EXPECTED_VOTES for
 every cluster member must be adjusted. For complete
 instructions
 check the section on configuring a cluster in the "OpenVMS
 Cluster
 Systems" manual.

CAUTION: If this cluster is running with multiple system disks and common system files will be used, please, do not proceed unless appropriate logical names are defined for cluster common files in SYLOGICALS.COM. For instructions, refer to the "OpenVMS Cluster Systems" manual.

If this cluster will run IPCI, then TCP/IP installed on the system should be version 5.7 and above or else IPCI configuration will be aborted.

Do you want to continue [Y]? [Return]

Is the node to be a clustered node with a shared SCSI/FIBRE-CHANNEL bus (Y/N)? Y

Will the node be a satellite [Y]? N

What is the node's SCS node name? MOON

DECnet is running on this node. Even though you are configuring a LAN-based cluster, the DECnet database will provide some information and may be updated.

What is the node's DECnet full name? local:.MOON

Do you want to define a DECnet synonym [Y]? N

What is the MOP service client name for this node [MOON]? VENUS

What is the node's SCSSYSTEMID number? 24.123

NOTE: 24.123 equates to an SCSSYSTEMID of 24699

Will MOON run DECnet [Y]? [Return]

Note:

This procedure will not update any network databases with information about MOON. You must do that yourself.

Will MOON be a boot server [Y]? [Return]

This procedure will now ask you for the device name of MOON's system root.

The default device name (DISK\$ALPHA732:) is the logical volume name of SYS\$SYSDEVICE:.

What is the device name for MOON's system root [default DISK\$ALPHA732:]?

What is the name of MOON's system root [SYS1]? [Return]

Creating directory tree SYS1 ...

System root SYS1 created

ENABLE IP for cluster communications (Y/N)? N

CAUTION: If you do not define port allocation classes later in this procedure for shared SCSI buses, all nodes sharing a SCSI bus must have the same non-zero ALLOCLASS value. If multiple nodes connect to a shared SCSI bus without the same allocation class for the bus, system booting will halt due to the error or IO AUTOCONFIGURE after boot will keep the bus offline.

WARNING: If BISMIL is sharing the same SCSI bus with MOON, then BISMIL's ALLOCLASS parameter or port allocation class for the shared bus

must be changed from 0 to the same non-zero value that will be entered for MOON. Use the CHANGE option of CLUSTER_CONFIG_LAN.COM to change BISMIL's ALLOCLASS parameter before MOON is booted.

Enter a value for MOON's ALLOCLASS parameter [1]: [Return]
Does this cluster contain a quorum disk [N]? [Return]
Size of pagefile for MOON [RETURN for AUTOGEN sizing]? [Return]

A temporary pagefile will be created until resizing by AUTOGEN. The default size below is arbitrary and may or may not be appropriate.

Size of temporary pagefile [10000]? [Return]
Size of swap file for MOON [RETURN for AUTOGEN sizing]? [Return]

A temporary swap file will be created until resizing by AUTOGEN. The default size below is arbitrary and may or may not be appropriate.

Size of temporary swap file [8000]? [Return]
Each shared SCSI bus must have a positive allocation class value. A shared bus uses a PK adapter. A private bus may use: PK, DR, DV, DQ.

When adding a node with SCSI-based cluster communications, the shared SCSI port allocation classes may be established in SYS\$DEVICES.DAT. Otherwise, the system's disk allocation class will apply.

A private SCSI bus need not have an entry in SYS\$DEVICES.DAT. If it has an entry, its entry may assign any legitimate port allocation class value:

n where n = a positive integer, 1 to 32767 inclusive
0 no port allocation class and disk allocation class does not apply
-1 system's disk allocation class applies (system parameter ALLOCLASS)

When modifying port allocation classes, SYS\$DEVICES.DAT must be updated for all affected nodes, and then all affected nodes must be rebooted. The following dialog will update SYS\$DEVICES.DAT on MOON.

Enter [RETURN] to continue: [Return]

There are currently no entries in SYS\$DEVICES.DAT for MOON. After the next boot, any SCSI controller on MOON will use MOON's disk allocation class.

Assign port allocation class to which adapter [RETURN for none]:
Will a local (non-HSx) disk on MOON and not on a hierarchical storage controller be used for paging and swapping (Y/N)? N

If you specify a device other than DISK\$ALPHA732: for MOON's page and swap files, this procedure will create PAGEFILE_MOON.SYS and SWAPFILE_MOON.SYS in the <SYSEXE> directory on the device you specify.

What is the device name for the page and swap files [DISK\$ALPHA732:]?
 %SYSGEN-I-CREATED, BISMIL\$DKB100:<SYS1.SYSEXE>PAGEFILE.SYS;1 created
 %SYSGEN-I-CREATED, BISMIL\$DKB100:<SYS1.SYSEXE>SWAPFILE.SYS;1 created
 The configuration procedure has completed successfully.

MOON has been configured to join the cluster.

Before booting MOON, you must create a new default bootstrap command procedure for MOON. For instructions, see your processor-specific installation and operations guide.

The first time MOON boots, NET\$CONFIGURE.COM and AUTOGEN.COM will run automatically.

The following parameters have been set for MOON:

```
VOTES = 1
QDSKVOTES = 1
```

After MOON has booted into the cluster, you must increment the value for EXPECTED_VOTES in every cluster member's MODPARAMS.DAT. You must then reconfigure the cluster, using the procedure described in the "OpenVMS Cluster Systems" manual.

Example 8.3. Sample Interactive CLUSTER_CONFIG_LAN.COM Session to Add a Satellite with Local Page and Swap Files

```
$ @CLUSTER_CONFIG_LAN.COM
      Cluster/IPCI Configuration Procedure
      CLUSTER_CONFIG_LAN Version V2.84
      Executing on an IA64 System

DECnet-Plus is installed on this node.
IA64 satellites will use TCP/IP BOOTP and TFTP services for downline
loading.
TCP/IP is installed and running on this node.

      Enter a "?" for help at any prompt.  If you are familiar with
      the execution of this procedure, you may want to mute extra notes
      and explanations by invoking it with "@CLUSTER_CONFIG_LAN BRIEF".

BHAGAT is an IA64 system and currently a member of a cluster
so the following functions can be performed:

MAIN Menu

1. ADD an IA64 node to the cluster.
2. REMOVE a node from the cluster.
3. CHANGE a cluster member's characteristics.
4. CREATE a duplicate system disk for BHAGAT.
5. MAKE a directory structure for a new root on a system disk.
6. DELETE a root from a system disk.
7. EXIT from this procedure.

Enter choice [7]: 1

      This ADD function will add a new IA64 node to the cluster.
```

WARNING: If the node being added is a voting member, EXPECTED_VOTES for every cluster member must be adjusted. For complete instructions check the section on configuring a cluster in the "OpenVMS Cluster Systems" manual.

CAUTION: If this cluster is running with multiple system disks and common system files will be used, please, do not proceed unless appropriate logical names are defined for cluster common files in SYLOGICALS.COM. For instructions, refer to the "OpenVMS Cluster Systems" manual.

If this cluster will run IPCI, then TCP/IP installed on the system should be version 5.7 and above or else IPCI configuration will be aborted.

Do you want to continue [Y]? [Return]

Is the node to be a clustered node with a shared SCSI/FIBRE-CHANNEL bus (Y/N)? N

Will the node be a satellite [Y]? [Return]

What is the node's SCS node name? GOMTHI

DECnet is running on this node. Even though you are configuring a LAN-based cluster, the DECnet database will provide some information and may be updated.

What is the node's DECnet full name? local:.GOMTHI

Do you want to define a DECnet synonym [Y]? N

What is the node's SCSSYSTEMID number? 25.171

NOTE: 25.171 equates to an SCSSYSTEMID of 25771

WARNING:

The DECnet databases on BHAGAT will not be updated with information on GOMTHI. You must see to it that network databases on this and all other cluster members are updated. For help, refer to the "OpenVMS Cluster Systems" manual.

Does GOMTHI need to be registered in the DECnet namespace [N]?[Return]

What is the Cluster Alias full name? [Return]

Will GOMTHI run DECnet [Y]? [Return]

This procedure will now ask you for the device name of GOMTHI's system root.

The default device name (DISK\$BHAGAT_SYS:) is the logical volume name of SYS\$SYSDEVICE:.

What is the device name for GOMTHI's system root [default DISK \$BHAGAT_SYS:]?

What is the name of GOMTHI's system root [SYS10]? [Return]

What is GOMTHI's LAN adapter hardware address? 00-30-6E-4C-BB-1A

What is GOMTHI's TCP/IP address? 16.181.160.129

Would you like GOMTHI added as a TCP/IP host shortcut for 16.181.160.129
[Y]? [Return]

What is GOMTHI's TCP/IP gateway or gateways (leave blank if none)?
16.181.160.1

What is GOMTHI's TCP/IP network mask [255.255.255.0]? 255.255.252.0

NOTE: Make sure to set the VMS_FLAGS console variable
to 0,200000 on node GOMTHI so it will use
the memory-disk method to boot as a satellite.
The command to update this variable from the
console EFI shell of GOMTHI is:
set vms_flags "0,200000"

Allow conversational bootstraps on GOMTHI [N]? [Return]

The following workstation windowing options are available:

1. No workstation software
2. DECwindows Workstation Software

Enter choice [1]: [Return]

Creating directory tree SYS10 ...
System root SYS10 created

ENABLE IP for cluster communications (Y/N)? N

Will GOMTHI be a disk server [N]? Y

Enter a value for GOMTHI's ALLOCLASS parameter [0]: [Return]

Updating BOOTP database with satellite information for GOMTHI...

Size of pagefile for GOMTHI [RETURN for AUTOGEN sizing]? [Return]

A temporary pagefile will be created until resizing by AUTOGEN. The
default size below is arbitrary and may or may not be appropriate.

Size of temporary pagefile [10000]? [Return]

Size of swap file for GOMTHI [RETURN for AUTOGEN sizing]? [Return]

A temporary swap file will be created until resizing by AUTOGEN. The
default size below is arbitrary and may or may not be appropriate.

Size of temporary swap file [8000]? [Return]

NOTE: IA64 satellite node GOMTHI requires DOSD if capturing the
system state in a dumpfile is desired after a system crash.

Will a local disk on GOMTHI be used for paging and swapping (Y/N)? Y

This procedure will now wait until GOMTHI is a member of
the cluster. Once GOMTHI joins the cluster, this procedure
will ask you which local disk it can use for paging and swapping.

Please boot GOMTHI now. Make sure the default boot device is
set to the appropriate clustered-disk access path: LAN device for
satellite nodes; or shared-bus (CI/DSSI/SCSI/FC) disk device. See
the hardware user manual or the console help command for instructions
to do this.

Waiting for GOMTHI to boot...

Waiting for GOMTHI to boot...

Waiting for GOMTHI to boot...

Waiting for GOMTHI to boot...
 Waiting for GOMTHI to boot...

Node GOMTHI is now a cluster member. This procedure will pause for up to 4 minutes, while attempting to detect local disks on GOMTHI, to use for paging and swapping.

The local disks on GOMTHI are:

Device	Device	Error	Volume	Free
Trans Mnt	Status	Count	Label	Blocks
Name				
Count	Cnt			
GOMTHI\$DQA0:	Online	0		
GOMTHI\$DKA0:	Online	0		
GOMTHI\$DKA100:	Online	0		
GOMTHI\$DKB200:	Online	0		

If the paging and swapping disk you plan to use is not displayed, it may not yet be configured. Please wait a few moments and hit a carriage return for an updated display.

Which disk can be used for paging and swapping? GOMTHI\$DKA100:
 May this procedure INITIALIZE GOMTHI\$DKA100 [Y]? N

In order to ensure that this disk has a unique volume name this procedure wishes to change its name from [GOMTHI_831H1] to [GOMTHI_25771]. If the satellite being added may also be booted standalone and refers to this disk by name you may retain the old volume name if there are no other disks with the same name in this cluster.

May the volume name of this disk be changed to GOMTHI_25771 [Y]? N
 %DELETE-W-SEARCHFAIL, error searching for SYS\$COMMON:
 [SYSMGR]CLU2020042F.TMP1;*
 -RMS-E-FNF, file not found
 Mounting GOMTHI\$DKA100...

What is the file specification for the pagefile on
 GOMTHI\$DKA100: [<SYSEXE>PAGEFILE.SYS]? [Return]
 %CREATE-I-CREATED, GOMTHI\$DKA100:<SYSEXE> created
 %SYSGEN-I-CREATED, GOMTHI\$DKA100:<SYSEXE>PAGEFILE.SYS;1 created

What is the file specification for the swapfile on
 GOMTHI\$DKA100: [<SYSEXE>SWAPFILE.SYS]? [Return]
 %SYSGEN-I-CREATED, GOMTHI\$DKA100:<SYSEXE>SWAPFILE.SYS;1 created

SATELLITE_PAGE.COM and INSTALL_PAGE.COM will now be created for local page/swap disk/file installation.

```

***** ! SHADOWED PAGE or SWAP DISK WARNING ! *****
****  Edit these procedures to include any          ****
****  local configuration commands necessary for ****
****  shadowed disk mounting, prior to reboot. ****
*****

```

AUTOGEN will now reconfigure and reboot GOMTHI automatically.

These operations will complete in a few minutes, and a completion message will be displayed at your terminal.

```
Waiting for GOMTHI to reboot...
Waiting for GOMTHI to reboot...
```

The configuration procedure has completed successfully.

8.2.3.1. Creating and Configuring a Two-Node Cluster Using IP

Cluster over IP can be used to create and configure a two-node cluster. Node ORCHID is a standalone node at SITE A and node TULIP at SITE B, is already a member (the only member) of the cluster. In this scenario, Cluster over IP is not configured in TULIP. SITE A and SITE B can be in the same or different LAN, building or any other geographical location. It is required to have IP connectivity between SITE A and SITE B and should be within the supported inter site distance.

Step 1. Configuring Node TULIP to Enable Cluster over IP

To configure the node TULIP (OpenVMS Alpha node) for enabling the Cluster over IP feature, execute the CLUSTER_CONFIG_LAN.COM procedure on node TULIP and select the appropriate option as illustrated:

Example 8.4. Configuring Node TULIP to Enable Cluster over IP

```
TULIP$@SYS$MANAGER:CLUSTER_CONFIG_LAN
          Cluster/IPCI Configuration Procedure
          CLUSTER_CONFIG_LAN Version V2.84
          Executing on an Alpha System
```

DECnet-Plus is installed on this node.
Alpha satellites will use LANCP, not DECnet, for MOP downline loading.

Enter a "?" for help at any prompt. If you are familiar with the execution of this procedure, you may want to mute extra notes and explanations by invoking it with "@CLUSTER_CONFIG_LAN BRIEF".

TULIP is an Alpha system and currently a member of a cluster so the following functions can be performed:

MAIN Menu

1. ADD an Alpha node to the cluster.
2. REMOVE a node from the cluster.
3. CHANGE a cluster member's characteristics.
4. CREATE a duplicate system disk for TULIP.
5. MAKE a directory structure for a new root on a system disk.
6. DELETE a root from a system disk.
7. EXIT from this procedure.

Enter choice [7]: 3 **1**

CHANGE Menu

1. Enable TULIP as a boot server.
2. Disable TULIP as a boot server.
3. Enable a quorum disk for TULIP

4. Disable a quorum disk for TULIP.
5. Enable TULIP as a disk server.
6. Disable TULIP as a disk server.
7. Change TULIP's ALLOCLASS value.
8. Enable TULIP as a tape server.
9. Disable TULIP as a tape server.
10. Change TULIP's TAPE_ALLOCLASS value.
11. Change an Alpha satellite node's LAN adapter hardware address.
12. Enable Cluster Communication using IP on TULIP.
13. Disable Cluster Communication using IP on TULIP.
14. Enable the LAN for cluster communications on TULIP.
15. Disable the LAN for cluster communications on TULIP.
16. Enable Memory Channel for cluster communications on TULIP.
17. Disable Memory Channel for cluster communications on TULIP.
18. Change TULIP's shared SCSI port allocation class value.
19. Return to MAIN menu.

Enter choice [19]: 12 ②

ENABLE IP for cluster communications (Y/N)? Y ③

UDP port number to be used for Cluster Communication over IP[49152]?Y ④

Enable IP multicast for cluster communication(Y/N)[Y]? Y ⑤

What is the IP multicast address[239.242.7.193]? [Return] ⑥

What is the TTL (time to live) value for IP multicast packets [32]?

[Return] ⑦

Do you want to enter unicast address(es)(Y/N)[Y]? [Return] ⑧

What is the unicast address [Press Enter to end the list]? 10.0.1.2 ⑨

What is the unicast address[Press Enter to end the list]? [Return] ⑩

```
*****
Cluster Communications over IP has been enabled. Now
CLUSTER_CONFIG_LAN will run the SYS$MANAGER:TCPIP$CONFIG
procedure. Please select the IP interfaces to be used for
Cluster Communications over IP (IPCI). This can be done
selecting "Core Environment" option from the main menu
followed by the "Interfaces" option. You may also use
this opportunity to configure other aspects.
*****
```

Press Enter to continue.

Checking TCP/IP Services for OpenVMS configuration database files.

TCP/IP Services for OpenVMS Configuration Menu

Configuration options:

- 1 - Core environment
- 2 - Client components
- 3 - Server components
- 4 - Optional components
- 5 - Shutdown TCP/IP Services for OpenVMS
- 6 - Startup TCP/IP Services for OpenVMS
- 7 - Run tests
- A - Configure options 1 - 4
- [E] - Exit configuration procedure

Enter configuration option: 1 ⑪

TCP/IP Services for OpenVMS Core Environment Configuration Menu

Configuration options:

- 1 - Domain
- 2 - Interfaces
- 3 - Routing
- 4 - BIND Resolver
- 5 - Time Zone
- A - Configure options 1 - 5
- [E] - Exit menu

Enter configuration option: 2 **12**

TCP/IP Services for OpenVMS Interface & Address Configuration Menu

Hostname Details: Configured=TULIP, Active=TULIP

Configuration options:

- 0 - Set The Target Node (Current Node: TULIP)
- 1 - IE0 Menu (EIA0: TwistedPair 100mbps)
- 2 - 10.0.2.2/23 TULIP
Configured,Active
- 3 - IE1 Menu (EIB0: TwistedPair 100mbps)
- 4 - 10.0.2.224/23 *noname*
Configured,Active
- I - Information about your configuration
- [E] - Exit menu

Enter configuration option: 2 **13**

TCP/IP Services for OpenVMS Address Configuration Menu (Node: TULIP)

IE0 10.0.2.2/23 TULIP Configured, Active IE0

Configuration options

- 1 - Change address
- 2 - Set "TULIP" as the default hostname
- 3 - Delete from configuration database
- 4 - Add to IPCI database
- 5 - Deactivate from live system
- 6 - Add standby aliases to configuration database (for
failSAFE IP)
- [E] - Exit menu

Enter configuration option: 4 **14**

Updated Interface in IPCI configuration file: SYS\$SYSROOT:[SYSEXE]TCPIP
\$CLUSTER.DAT;

Updated Default Route in IPCI configuration file: SYS\$SYSROOT:[SYSEXE]TCPIP
\$CLUSTER.DAT;

Added address IE1:10.0.2.2 to IPCI database

TCP/IP Services for OpenVMS Interface & Address Configuration Menu

Hostname Details: Configured=TULIP, Active=TULIP

Configuration options:

```
0 - Set The Target Node (Current Node:tulip)
1 - IE0 Menu (EIA0: TwistedPair 100mbps)
2 - 10.0.2.2/23 TULIP
Configured,IPCI,Active
3 - IE1 Menu (EIB0: TwistedPair 100mbps)
4 - 10.0.2.224/23 *noname*
Configured,Active
I - Information about your configuration
[E]- Exit menu
```

Enter configuration option: E ⑮

TCP/IP Services for OpenVMS Core Environment Configuration Menu

Configuration options:

```
1 - Domain
2 - Interfaces
3 - Routing
4 - BIND Resolver
5 - Time Zone
A - Configure options 1 - 5
[E] - Exit menu
```

Enter configuration option: E

TCP/IP Services for OpenVMS Configuration Menu

Configuration options:

```
2 - Client components
3 - Server components
4 - Optional components
5 - Shutdown TCP/IP Services for OpenVMS
6 - Startup TCP/IP Services for OpenVMS
7 - Run tests
A - Configure options 1 - 4
[E] - Exit configuration procedure
```

Enter configuration option: E

The configuration procedure has completed successfully.

Tulip has been enabled for IP communications ⑯
Please run AUTOGEN to reboot TULIP:

TULIP\$ @SYS\$UPDATE:AUTOGEN GETDATA REBOOT ⑰

① TULIP is a single-member cluster without Cluster over IP enabled. The cluster member characteristics can be changed to enable Cluster over IP for this node by selecting option 3.

- ② Select option 12 to enable cluster over IP. By selecting this option, the SYSGEN parameter, NISCS_USE_UDP is set to 1, which enables the PEDRIVER to use IP for cluster communication. This requires reboot of the node. If LAN is not already selected as the cluster interconnect, this option sets NISCS_LOAD_PEA0 to 1 to load PEDRIVER during the next reboot.
- ③ Enable IP for cluster communication.
- ④ You can enable IP multicast for cluster communication if your environment allows IP multicast traffic between cluster nodes. Check with your network administrator, if IP multicasting is enabled in your environment.
- ⑤ You can enable IP multicast for cluster communication if your environment allows IP multicast traffic between cluster nodes. Check with your network administrator, if IP multicasting is enabled in your environment.
- ⑥ Enter the IP multicast address for the cluster, if IP multicasting is enabled. By default, the IP multicast address is selected from the administratively scoped IP multicast address ranging from 239.242.x.y. The last two octets x and y are generated based on the cluster group number. In the above example cluster group number is 1985 and can be calculated as follows:

$$X = 1985 / 256$$

$$Y = 1985 - (256 * X)$$

The system administrator can override the default multicast address by a unique address for their environment.

- ⑦ TTL is the time to live for IP multicast packets. It specifies the number of hops allowed for IP multicast packets.
- ⑧ Enter "Yes" to enter the IP unicast addresses for the remote nodes of the cluster, which are not reachable using IP multicast address.
- ⑨ In this example, 10.0.1.2 is the IP unicast address for the node ORCHID. Although, the IP multicast is selected, ORCHID's IP address is entered because the IP multicast connectivity between SITE A and SITE B is presumed to be not existing in this example. Note: Enter the list of IP address of the cluster. All the information entered in [4],[6],[7] and [9] are entered into the SYSSYSTEM:PE\$IP_CONFIG.DAT file. The PE\$IP_CONFIG.DAT file is generated as shown in the following example. Also, to allow the remote node to join the cluster, the Unicast list in the PE\$IP_CONFIG.DAT on the local node must contain the IP address of the remote node. In this example, TULIP must have ORCHID's IP address and ORCHID must have TULIP's IP address.

```
! CLUSTER_CONFIG_LAN creating for CHANGE operation on 10-JUL-2008
14:14:06.57
```

```
multicast_address=239.242.7.193
```

```
ttl=32
```

```
udp_port=49152
```

```
unicast=10.0.1.2
```

- ⑩ Press **Return** after entering the Unicast list.

- ❶ CLUSTER_CONFIG_LAN.COM invokes TCPIP\$CONFIG.COM to configure the IP interfaces used for cluster communication. Select the **core environment** option. Assuming that TCP/IP is already configured, the node can be pinged from outside the subnet.
- ❷ Select the interface option from the core environment menu.
- ❸ Select the appropriate interface for cluster communication. In this example, option 2 is selected.
- ❹ In the configuration option for the selected address, select option 4 to add to IPCI database. The interface information along with the default route is entered in the TCPIP\$CLUSTER.DAT as shown:


```
interface=IE0,EIA0,10.0.2.2,255.255.254.0

default_route=10.0.2.1
```
- ❺ Exit from the TCP/IP configuration procedure, which returns to CLUSTER_CONFIG_LAN.COM.
- ❻ Proceed with cluster configuration.
- ❼ After rebooting the system, run AUTOGEN. PEDRIVER in ORCHID will start using IP in addition to LAN for cluster communication and must be able to join TULIP.

Step 2. Configuring Node ORCHID as a Cluster Member

To configure ORCHID with Cluster over IP enabled, execute CLUSTER_CONFIG_LAN.COM on node ORCHID and select the appropriate option as shown in the following example:

Example 8.5. Configuring Node ORCHID to Enable Cluster over IP

```
ORCHID$ @SYS$MANAGER:CLUSTER_CONFIG_LAN.COM
          Cluster/IPCI Configuration Procedure
          CLUSTER_CONFIG_LAN Version V2.84
          Executing on an IA64 System
DECnet-Plus is installed on this node.
IA64 satellites will use TCP/IP BOOTP and TFTP services for downline
loading.
TCP/IP is installed and running on this node.
```

Enter a "?" for help at any prompt. If you are familiar with the execution of this procedure, you may want to mute extra notes and explanations by invoking it with "@CLUSTER_CONFIG_LAN BRIEF".

This IA64 node is not currently a cluster member.

MAIN Menu

1. ADD ORCHID to existing cluster, or form a new cluster.
2. MAKE a directory structure for a new root on a system disk.
3. DELETE a root from a system disk.
4. EXIT from this procedure.

Enter choice [4]: 1 ❶

Is the node to be a clustered node with a shared SCSI/FIBRE-CHANNEL bus (Y/N)? N

What is the node's SCS node name? ORCHID

IA64 node, using LAN/IP for cluster communications. PEDRIVER will be loaded.

No other cluster interconnects are supported for IA64 nodes.

Enter this cluster's group number: 1985

Enter this cluster's password:

Re-enter this cluster's password for verification:

ENABLE IP for cluster communications (Y/N)? Y ②

UDP port number to be used for Cluster Communication over IP[49152]?

[Return] ③

Enable IP multicast for cluster communication(Y/N)[Y]? Y [Return] ④

What is IP the multicast address[239.242.7.193]? [Return] ⑤

What is the TTL (time to live) value for IP multicast packets [32]?

[Return] ⑥

Do you want to enter unicast address(es)(Y/N)[Y]?[Return] ⑦

What is the unicast address[Press [RETURN] to end the list]? 10.0.2.2 ⑧

What is the unicast address[Press [RETURN] to end the list]?[Return] ⑨

```
*****
                Cluster Communications over IP has been enabled.  Now
                CLUSTER_CONFIG_LAN will run the SYS$MANAGER:TCPIP$CONFIG
procedure.
```

```
                Pleas select the IP interfaces to be used for Cluster
Communications
                over IP (IPCI).  This can be done selecting "Core
Environment" option
                from the main menu followed by the "Interfaces" option.
You may
                also use this opportunity to configure other aspects.
```

```
*****
```

Press Return to continue ...

TCP/IP Network Configuration Procedure

This procedure helps you define the parameters required to run TCP/IP Services for OpenVMS on this system.

%TCPIP-I-IPCI, TCP/IP Configuration is limited to IPCI.

-TCPIP-I-IPCI, Rerun TCPIP\$CONFIG after joining the cluster.

TCP/IP Services for OpenVMS Interface & Address Configuration Menu

Hostname Details: Configured=Not Configured, Active=nodeg

Configuration options:

- 0 - Set The Target Node (Current Node: ORCHID)
- 1 - IE0 Menu (EIA0: TwistedPair 100mbps)
- 2 - IE1 Menu (EIB0: TwistedPair 100mbps)
- [E] - Exit menu

Enter configuration option: 1 ⑩

* IPCI Address Configuration *

Only IPCI addresses can be configured in the current environment. After configuring your IPCI address(es) it will be necessary to run TCPIP\$CONFIG once your node has joined the cluster.

IPv4 Address may be entered with CIDR bits suffix.
E.g. For a 16-bit netmask enter 10.0.1.1/16

Enter IPv4 Address []:10.0.1.2/24 ①
Default netmask calculated from class of IP address: 255.0.0.0

IPv4 Netmask may be entered in dotted decimal notation,
(e.g. 255.255.0.0), or as number of CIDR bits (e.g. 16)

Enter Netmask or CIDR bits [255.0.0.0]: 255.255.254.0 ②

Requested configuration:

```
Node      : ORCHID
Interface: IE0
IPCI      : Yes
Address   : 10.0.1.2/23
Netmask   : 255.255.254.0 (CIDR bits: 23)
```

* Is this correct [YES]:

Updated Interface in IPCI configuration file: SYS\$SYSROOT:[SYSEXE]TCPIP
\$CLUSTER.DAT; ③

TCP/IP Services for OpenVMS Interface & Address Configuration Menu

Hostname Details: Configured=Not Configured, Active=ORCHID

Configuration options:

```
0 - Set The Target Node (Current Node: ORCHID)
1 - IE0 Menu (EIA0: TwistedPair 100mbps)
2 - 10.0.1.2 /23 *noname*          IPCI
3 - IE1 Menu (EIB0: TwistedPair 100mbps)
[E] - Exit menu
```

Enter configuration option: E ④

Enter your Default Gateway address []: 10.0.1.1 ⑤

* The default gateway will be: 10.0.1.1. Correct [NO]: YES

Updated Default Route in IPCI configuration file: SYS\$SYSROOT:[SYSEXE]TCPIP
\$CLUSTER.DAT;

TCPIP-I-IPCIDONE, Finished configuring IPCI address information ⑥

Will ORCHID be a boot server [Y]? [Return] ⑦

TCP/IP BOOTP and TFTP services must be enabled on IA64 boot nodes.

Use SYS\$MANAGER:TCPIP\$CONFIG.COM on ORCHID to enable BOOTP and
TFTP
services after ORCHID has booted into the cluster.

Enter a value for ORCHID's ALLOCLASS parameter [7]:
Does this cluster contain a quorum disk [N]? [Return]

The EXPECTED_VOTES system parameter of members of a cluster indicates the total number of votes present when all cluster members are booted, and is used to determine the minimum number of votes (QUORUM) needed for cluster operation.

EXPECTED_VOTES value for this cluster: 1

Warning: Setting EXPECTED_VOTES to 1 allows this node to boot without being able to see any other nodes in the cluster. If there is another instance of the cluster in existence that is unreachable via SCS but shares common drives (such as a Fibrechannel fabric) this may result in severe disk corruption.

Do you wish to re-enter the value of EXPECTED_VOTES [Y]? N

The use of a quorum disk is recommended for small clusters to maintain cluster quorum if cluster availability with only a single cluster node is a requirement.

For complete instructions, check the section on configuring a cluster in the "OpenVMS Cluster Systems" manual.

WARNING: ORCHID will be a voting cluster member. EXPECTED_VOTES for this and every other cluster member should be adjusted at a convenient time before a reboot. For complete instructions, check the section on configuring a cluster in the "OpenVMS Cluster Systems" manual.

Execute AUTOGEN to compute the SYSGEN parameters for your configuration and reboot ORCHID with the new parameters. This is necessary before ORCHID can become a cluster member.

Do you want to run AUTOGEN now [Y]? N

Please run AUTOGEN to reboot ORCHID:

```
ORCHID$ @SYS$UPDATE:AUTOGEN GETDATA REBOOT ③
```

- ① Node ORCHID is currently a standalone, Integrity server node and is made as a member of a cluster. Only LAN or IP is used for cluster communication and no other interconnect is supported.
- ② Select IP for cluster communication in addition to LAN by entering "YES". The SYSGEN parameter, NISCS_USE_UDP is set to 1 and PEDRIVER uses IP in addition to LAN for cluster communication when the node is rebooted.
- ③ The UDP port number to be used for cluster communication. The UDP port number must be same on all members of the cluster. Also, ensure that there is no other cluster in your

environment using the same UDP port number and this port number must not be used by any other application.

- ④ You can enable IP multicast for cluster communication if your environment allows IP multicast traffic between cluster nodes. Check with your network administrator to see if IP multicasting is enabled in your environment.
- ⑤ Enter the IP multicast address for cluster, if IP multicasting is enabled. By default, the IP multicast address is selected from the administratively scoped IP multicast address range of 239.242.x.y. The last two octets x and y are generated based on the cluster group number. In the above example cluster group number is 1985 and is calculates as follows:

```
X= 1985/256
Y= 1985 - (256 *x)
```

The system administrator can override the default multicast address by a unique address for their environment.

- ⑥ TTL is the time to live for IP multicast packets. It specifies the number of hops allowed for IP multicast packets.
- ⑦ Enter "Yes" to enter the IP unicast address for remote nodes of the cluster which are not reachable using IP multicast address.
- ⑧ In this example, 10.0.2.2 is the IP unicast address of node TULIP. Although, IP multicast is selected, TULIP's IP address is entered because the IP multicast connectivity between SITE A and SITE B is presumed to be non-existent in this example. NOTE: Enter the list of IP unicast address of the cluster. All the information entered in [2], [3], [5], [6], and [7] are entered in the SYS\$SYSTEM:PE\$IP_CONFIG.DAT file. The PE\$IP_CONFIG.DAT file is generated as shown in following example. Also, the Unicast list in PE\$IP_CONFIG.DAT in the local node should contain the remote node IP address for the local node to allow the remote node to join the cluster.

In this example, ORCHID must have TULIP's IP address and TULIP must have ORCHID's IP address. SYSTEM:PE\$IP_CONFIG.DAT in node ORCHID:

```
! CLUSTER_CONFIG_LAN creating for CHANGE operation on 10-JUL-
2008 14:14:06.57

multicast_address=239.242.7.193

ttl=32

udp_port=49152

unicast=10.0.2.2
```

- ⑨ Press **Return** after entering the unicast list.
- ⑩ CLUSTER_CONFIG_LAN.COM invokes TCPIP\$CONFIG.COM to configure the IP interfaces used for cluster communication. Currently, ORCHID is a standalone node, when TCPIP \$CONFIG is invoked by the CLUSTER_CONFIG_LAN procedure, TCP/IP configuration is limited to IPCI. The interface, IE0 is selected for enabling cluster communications. Note: TCPIP \$CONFIG must be invoked after joining the cluster for other TCP/IP configuration, such as, FTP, TELNET.

- ❶ IPv4 address for the IE0 interface is 10.0.1.2
- ❷ Network mask for the IE0 interface is 255.255.254.0
- ❸ The IE0 interface information along with network mask is entered in the TCPIP \$CLUSTER.DAT file.
- ❹ Exit the interface menu after selecting the interface for cluster communication.
- ❺ The default gateway address for the interface IE0 is entered. Only one default gateway address is allowed for Cluster over IP communication.
- ❻ After the interface and default gateway are selected, TCPIP\$CONFIG updates the TCPIP \$CLUSTER.DAT with the default route or gateway information. This also completes the TCPIP \$CONFIG required for cluster communications using IP. The interface information along with the default route is entered in the TCPIP\$CLUSTER.DAT file as shown:


```
interface=IE0,EIA0,10.0.1.2,255.255.254.0

default_route=10.0.1.1
```
- ❼ Proceed with cluster configuration.
- ❽ After rebooting the system, run AUTOGEN. PEDRIVER in ORCHID will start using IP in addition to LAN for cluster communication and must be able to join TULIP.

8.2.3.2. Adding a new Node to a Cluster over IP

This section describes how to add a new node, JASMIN to an existing two-node cluster. Nodes, ORCHID and TULIP are currently members of a two-node cluster, which are at SITE A and SITE B. For more information about configuring a node with IP as interconnect, see Section 8.2.3.1. Node JASMIN is currently a standalone node at SITE C with IP connectivity to both SITE A and SITE B.

Step 1. Ensuring IP Connectivity

Ensure that the IP connectivity between the node JASMIN and the nodes ORCHID and TULIP is working fine. Use the TCP/IP PING utility to test the IP connectivity between JASMIN and other nodes, ORCHID and TULIP.

If PING fails, set up TCP/IP configuration properly so that node JASMIN can ping both ORCHID and TULIP.

Step 2. Executing the CLUSTER_CONFIG_LAN.COM

Execute CLUSTER_CONFIG_LAN.COM on node JASMIN. Because, the node JASMIN is a standalone node, complete the procedure described in Section 8.2.3.1. Complete the sequence of steps provided in the following example while entering the unicast list.

```
Do you want to enter unicast address(es) (Y/N) [Y]? [Return]
What is the unicast address [Press [RETURN] to end the list]? 10.0.3.2
What is the unicast address [Press [RETURN] to end the list]? 10.0.2.2
What is the unicast address [Press [RETURN] to end the list]? 10.0.1.2 ❶
What is the unicast address [Press [RETURN] to end the list]? [Return]
SYS$SYSTEM:PE$IP_CONFIG.DAT file generated in node JASMIN shown below
! CLUSTER_CONFIG_LAN creating for CHANGE operation on 10-JUL-2008
 14:14:06.57
multicast_address=239.242.7.193
ttl=32
```

```
udp_port=49152
unicast=10.0.3.2
Unicast=10.0.2.2
Unicast=10.0.1.2
```

- ❶ Enter the IP address of JASMIN, ORCHID and TULIP while configuring the node JASMIN.

Note

The unicast list must be consistent in all nodes of the cluster. Hence, while entering the unicast list in JASMIN, enter the IP addresses of all the three nodes of the cluster (that is, JASMIN, ORCHID and TULIP). You can also enter the local nodes IP addresses along with the Unicast list as it facilitates system management.

Step 3. Completing the Configuration Procedure Continue to run the CLUSTER_CONFIG_LAN.COM to complete the cluster configuration procedure, see Section 8.2.3.1 for more details.

Step 4. Updating the PE\$IP_CONFIG.DAT File To ensure that the nodes join the cluster, it is required to have PE\$IP_CONFIG.DAT consistent through all the members of the cluster. Copy the SYS\$SYSTEM:PE\$IP_CONFIG.DAT file that is created on node JASMIN to the other nodes, ORCHID, and TULIP.

Step 5. Refreshing the Unicast List

On both ORCHID and TULIP, to update the new unicast list in the PE\$IP_CONFIG.DAT file, enter the following command for PEDRIVER:

```
$MC SCACP RELOAD
```

You can also use SYSMAN and run the command cluster wide.

Note

The following rule is applicable when IP unicast address is used for node discovery. A node is allowed to join the cluster only if its IP address is present in the existing members of the SYS\$SYSTEM:PE\$IP_CONFIG.DAT file.

Step 6. Running AUTOGEN and Rebooting the Node

After the first boot of JASMIN, AUTOGEN.COM runs automatically to join the existing cluster consisting of nodes ORCHID and LOTUS.

```
JASMIN$ @SYS$UPDATE:AUTOGEN GETDATA REBOOT
```

8.2.3.3. Adding a new Node to a Cluster over IP with a Shared System Disk

This section describes how to add a new node JASMIN that has a shared system disk of TULIP. ORCHID and TULIP are currently members of two-node cluster which are at SITE A and SITE B.

Step 1. Obtaining the Interface Information

Node JASMIN is an OpenVMS Alpha node and is directly connected to the system disk of one of the node TULIP. In this configuration, Node JASMIN is connected in network, but not yet booted.

To configure a Cluster over IP, the interface information of JASMIN is required. This information can be obtained from the ">>>" prompt on JASMIN by executing the following command:

```
P00>>>SHOW DEVICE
dga5245.1003.0.3.0      $1$DGA5245      COMPAQ HSV110 (C)COMPAQ  3028
dga5245.1004.0.3.0      $1$DGA5245      COMPAQ HSV110 (C)COMPAQ  3028
dga5890.1001.0.3.0      $1$DGA5890      COMPAQ HSV110 (C)COMPAQ  3028
dga5890.1002.0.3.0      $1$DGA5890      COMPAQ HSV110 (C)COMPAQ  3028
dka0.0.0.2004.0         DKA0            COMPAQ BD03685A24      HPB7
dka100.1.0.2004.0       DKA100          COMPAQ BD01864552      3B08
dka200.2.0.2004.0       DKA200          COMPAQ BD00911934      3B00
dqa0.0.0.15.0          DQA0            HL-DT-ST CD-ROM GCR-8480  2.11
dva0.0.0.1000.0         DVA0
eia0.0.0.2005.0         EIA0            00-06-2B-03-2D-7D
pga0.0.0.3.0           PGA0            WWN 1000-0000-c92a-78e9
pka0.7.0.2004.0         PKA0            SCSI Bus ID 7
pkb0.6.0.2.0           PKB0            SCSI Bus ID 6 5.57
P00>>>
```

From the output, the interface will be EIA0 on which the IP address will be configured and can be used for cluster formation.

To obtain the interface information on Integrity server system, execute the following commands on the EFI Shell:

```
shell>fs0:
fs0:\> cd efi
fs0:\EFI> cd vms
fs0:\EFI\VMS> vms_show device
VMS: EIA0            00-30-6E-F3-EC-6E
EFI: Acpi(HWP0002,0)/Pci(3|0)/Mac(00306EF3EC6E)
VMS: DKA100         HP 36.4GST336754LC      HPC2
EFI: Acpi(HWP0002,100)/Pci(1|0)/Scsi(Pun1,Lun0)
VMS: DKA0           HP 36.4GMAS3367NC      HPC3      X8_3_XBJL
EFI: fs0: Acpi(HWP0002,100)/Pci(1|0)/Scsi(Pun0,Lun0)
VMS: EWA0           00-30-6E-F3-3C-28
EFI: Acpi(HWP0002,100)/Pci(2|0)/Mac(00306EF33C28)
fs0:\EFI\VMS>
```

From the output, the interface will be EIA0. Here fs0: is the partition of the shared system disk.

Step 2. Executing CLUSTER_CONFIG_LAN.COM

Execute the following command procedure on node TULIP:

```
TULIP$ @SYS$SYSROOT:[SYSMGR]CLUSTER_CONFIG_LAN.COM;1
```

```
Cluster/IPCI Configuration Procedure
CLUSTER_CONFIG_LAN Version V2.84
Executing on an Alpha System
```

DECnet Phase IV is installed on this node.

Alpha satellites will use LANCP, not DECnet, for MOP downline loading.

```
Enter a "?" for help at any prompt. If you are familiar with
the execution of this procedure, you may want to mute extra notes
and explanations by invoking it with "@CLUSTER_CONFIG_LAN BRIEF".
```

```
TULIP is an Alpha system and currently a member of a cluster
```

so the following functions can be performed:

MAIN Menu

1. ADD an Alpha node to the cluster.
2. REMOVE a node from the cluster.
3. CHANGE a cluster member's characteristics.
4. CREATE a duplicate system disk for Tulip.
5. MAKE a directory structure for a new root on a system disk.
6. DELETE a root from a system disk.
7. EXIT from this procedure.

Enter choice [7]: 1

This ADD function will add a new Alpha node to the cluster.

WARNING: If the node being added is a voting member, EXPECTED_VOTES for every cluster member must be adjusted. For complete instructions check the section on configuring a cluster in the "OpenVMS Cluster Systems" manual.

CAUTION: If this cluster is running with multiple system disks and common system files will be used, please, do not proceed unless appropriate logical names are defined for cluster common files in SYLOGICALS.COM. For instructions, refer to the "OpenVMS Cluster Systems" manual.

Do you want to continue [Y]?Y

Is the node to be a clustered node with a shared SCSI/FIBRE-CHANNEL bus (Y/N)? Y

Will the node be a satellite [Y]? N

What is the node's SCS node name? JASMIN

What is the node's SCSSYSTEMID number? 14487

Will JASMIN be a boot server [Y]?Y

This procedure will now ask you for the device name of JASMIN's system root.

The default device name (DISK\$TULIPSYS:) is the logical volume name of SYS\$SYSDEVICE:.

What is the device name for JASMIN's system root

[default DISK\$TULIPSYS:]?

What is the name of JASMIN's system root [SYS3]?SYS3

Creating directory tree SYS3 ...

System root SYS3 created

ENABLE IP for cluster communications (Y/N)? Y

UDP port number to be used for Cluster Communication over IP[49152]?

[Return]

Enable IP multicast for cluster communication(Y/N)[Y]?Y

What is the IP multicast address[224.0.0.3]? [Return]

What is the TTL (time to live) value for IP multicast packets [1] ?

[Return]

Do you want to enter unicast address(es)(Y/N)[Y]?Y

What is the unicast address[Press [RETURN] to end the list]? 10.0.1.2

What is the unicast address[Press [RETURN] to end the list]? 10.0.2.2

What is the unicast address[Press [RETURN] to end the list]? 10.0.2.3

What is the unicast address[Press [RETURN] to end the list]? [Return]

```
*****
Cluster Communications over IP has been enabled. Now
CLUSTER_CONFIG_LAN will run the SYS$MANAGER:TCPIP$CONFIG
procedure. Please select the IP interfaces to be used for
Cluster Communications over IP (IPCI). This can be done
selecting "Core Environment" option from the main menu
followed by the "Interfaces" option. You may also use
this opportunity to configure other aspects.
*****
```

Press Return to continue ...

Checking TCP/IP Services for OpenVMS configuration database files.

TCP/IP Services for OpenVMS Configuration Menu

Configuration options:

- 1 - Core environment
- 2 - Client components
- 3 - Server components
- 4 - Optional components
- 5 - Shutdown TCP/IP Services for OpenVMS
- 6 - Startup TCP/IP Services for OpenVMS
- 7 - Run tests
- A - Configure options 1 - 4
- [E] - Exit configuration procedure

Enter configuration option: 1

TCP/IP Services for OpenVMS Core Environment Configuration Menu

Configuration options:

- 1 - Domain
- 2 - Interfaces
- 3 - Routing
- 4 - BIND Resolver
- 5 - Time Zone
- A - Configure options 1 - 5
- [E] - Exit menu

Enter configuration option: 2

TCP/IP Services for OpenVMS Interface & Address Configuration Menu

Hostname Details: Configured=TULIP, Active=TULIP

Configuration options:

- 0 - Set The Target Node (Current Node: TULIP)
- 1 - WE0 Menu (EWA0: TwistedPair 1000mbps)
- 2 - 10.0.2.2/8 Tulip Configured,IPCI
- 3 - WE1 Menu (EWB0: TwistedPair 1000mbps)
- 4 - WE2 Menu (EWC0: TwistedPair 1000mbps)
- 5 - WE3 Menu (EWD0: TwistedPair 1000mbps)

- 6 - WE4 Menu (EWE0: TwistedPair 1000mbps)
- 7 - WE5 Menu (EWF0: TwistedPair 1000mbps)
- 8 - WE6 Menu (EWG0: Multimode 10000mbps)
- 9 - WE7 Menu (EWH0: TwistedPair 1000mbps)
- 10 - IE0 Menu (EIA0: TwistedPair 100mbps)
- 11 - IE1 Menu (EIB0: TwistedPair 100mbps)

Enter configuration option or press ENTER key to continue: 0 ❶

Enter name of node to manage [TULIP]: JASMIN

JASMIN is not currently a cluster member.

* Continue configuring JASMIN [NO]: Y ❷

Enter system device for JASMIN [\$10\$DGA165:]: ❸

Enter system root for JASMIN []: SYS3 ❹

TCP/IP Services for OpenVMS Interface & Address Configuration Menu

Hostname Details: Configured=Not Configured

Configuration options:

- 0 - Set The Target Node (Current Node: JASMIN - \$10\$DGA165:[sys3.]
- A - Add an Interface
- [E] - Exit menu

Enter configuration option: A

Enter controller name (e.g. EIA or EWC, etc): [ENTER when done] EIA ❺

Controller Name : EIA
TCP/IP Interface Name : IE0

* Is this correct [NO]: Y

Interface Menu: IE0

TCP/IP Services for OpenVMS Interface IE0 Configuration Menu (Node: JASMIN)

Configuration options:

- 1 - Add a primary address on IE0
- 2 - Add an alias address on IE0
- 3 - Enable DHCP client to manage address on IE0
- [E] - Exit menu

Enter configuration option: 1 ❻

* Is this address used by Clusters over IP (IPCI) [NO]: Y ❼

IPv4 Address may be entered with CIDR bits suffix.

E.g. For a 16-bit netmask enter 10.0.1.1/16

Enter IPv4 Address []: 10.0.2.3

Default netmask calculated from class of IP address: 255.0.0.0

IPv4 Netmask may be entered in dotted decimal notation,
(e.g. 255.255.0.0), or as number of CIDR bits (e.g. 16)

Enter Netmask or CIDR bits [255.0.0.0]:

Enter hostname []: JASMIN

Requested configuration:

```
Node      : JASMIN
Interface: IE0
IPCI      : Yes
Address   : 10.0.2.3/8
Netmask   : 255.0.0.0 (CIDR bits: 8)
Hostname  : JASMIN
```

* Is this correct [YES]:Y

Added hostname JASMIN (10.0.2.3) to host database

NOTE:

The system hostname is not configured.
It will now be set to jasmin (10.0.2.3).
This can be changed later via the Interface Configuration Menu.

Updated system hostname in configuration database

Added address IE0:10.0.2.3 to configuration database

Updated Interface in IPCI configuration file: \$10\$DGA165:[SYS3.SYSEXE]TCPIP
\$CLUSTER.DAT;

Updated Default Route in IPCI configuration file: \$10\$DGA165:
[SYS3.SYSEXE]TCPIP\$CLUSTER.DAT;

TCP/IP Services for OpenVMS Interface & Address Configuration Menu

Hostname Details: Configured=JASMIN

Configuration options:

- 0 - Set The Target Node (Current Node: JASMIN -
\$10\$DGA165:[sys3.]
 - 1 - IE0 Menu (EIA0:)
 - 2 - 10.0.2.3/8 JASMIN
- Configured,IPCI
- I - Information about your configuration
 - A - Add an Interface
 - [E] - Exit menu

Enter configuration option:

TCP/IP Services for OpenVMS Core Environment Configuration Menu

Configuration options:

- 1 - Domain
- 2 - Interfaces
- 3 - Routing
- 4 - BIND Resolver
- 5 - Time Zone
- A - Configure options 1 - 5
- [E] - Exit menu

Enter configuration option:

TCP/IP Services for OpenVMS Configuration Menu

Configuration options:

- 1 - Core environment
- 2 - Client components
- 3 - Server components
- 4 - Optional components
- 5 - Shutdown TCP/IP Services for OpenVMS
- 6 - Startup TCP/IP Services for OpenVMS
- 7 - Run tests
- A - Configure options 1 - 4
- [E] - Exit configuration procedure

Enter configuration option:

SYS\$SYSTEM:PE\$IP_CONFIG.DAT file generated in node JASMIN's root shown below

```
! CLUSTER_CONFIG_LAN creating for CHANGE operation on 15-JUL-2008
15:23:56.05
multicast_address=224.0.0.3
ttl=1
udp_port=49152
unicast=10.0.2.3
unicast=10.0.2.2
unicast=10.0.1.2
```

SYS\$SYSTEM:TCPIP\$CLUSTER.DAT file generated in node JASMIN's root shown below

```
interface=IE0,EIA0,10.0.2.3,255.0.0.0
default_route=16.116.40.1
```

- ❶ In the TCP/IP configuration, select option 0 to set the target node to JASMIN, which is the new node added to the cluster.
- ❷ Proceed with the configuration procedure to configure node JASMIN.
- ❸ Enter the system device for JASMIN, which is \$10\$DGA165.
- ❹ Enter JASMIN's root, which is SYS3.
- ❺ Enter the controller information on which IP will be configured for cluster traffic. This is the controller information that has been obtained from the console of the machine JASMIN as explained in the beginning of the configuration.
- ❻ Select the option to add the primary address for IE0 (IP interface name of controller EIA).
- ❼ Enable the use of IE0 for Cluster over IP and proceed with the rest of the configuration.

Step 3. Completing the Configuration Procedure

Continue to run the CLUSTER_CONFIG_LAN.COM to complete the cluster configuration procedure. For more information, see Section 8.2.3.1.

Step 4. Updating the PE\$IP_CONFIG.DAT file

To ensure that the nodes join the cluster, PE\$IP_CONFIG.DAT must be consistent through all the members of the cluster. Copy the SYSS\$SYSTEM:PE\$IP_CONFIG.DAT file that is created on node JASMIN to the other nodes, ORCHID and TULIP.

Step 5. Refreshing the Unicast list

On both ORCHID and TULIP, to update the new Unicast list in the PE\$IP_CONFIG.DAT file, enter the following command for PEDRIVER:

```
$ MC SCACP RELOAD
```

You can also use SYSMAN and run the command cluster wide.

Note

The following rule is applicable when IP unicast address is used for node discovery. A node is allowed to join the cluster only if its IP address is present in the existing members SYSS\$SYSTEM:PE\$IP_CONFIG.DAT file.

Step 6. Running AUTOGEN and Rebooting the Node

After the first boot of JASMIN, AUTOGEN.COM runs automatically. JASMIN will now be able to join the existing cluster consisting of nodes ORCHID and LOTUS.

```
JASMIN$ @SYS$UPDATE:AUTOGEN GETDATA REBOOT
```

8.2.3.4. Adding an Integrity server Satellite node to a Cluster over IP

This section describes how to add a satellite node to an existing two-node cluster. JASMIN is an Integrity server satellite node and is added to a cluster that has two nodes, ORCHID and TULIP. TULIP is the boot server for the satellite node.

Note

For both Alpha and Integrity server satellite nodes, the satellite node and its boot server must exist in the same LAN segment.

Step 1. Selecting the Interface for Satellite Booting

To select the interface to be used for satellite booting, assume that the satellite node does not have any disk running OpenVMS connected to it. Note: If you are adding Alpha systems as satellite nodes, you can get the information from ">>>>" prompt by executing the following command:

```
P00>>>SHOW DEVICE
dga5245.1003.0.3.0      $1$DGA5245   COMPAQ HSV110 (C)COMPAQ  3028
dga5245.1004.0.3.0      $1$DGA5245   COMPAQ HSV110 (C)COMPAQ  3028
dga5890.1001.0.3.0      $1$DGA5890   COMPAQ HSV110 (C)COMPAQ  3028
dga5890.1002.0.3.0      $1$DGA5890   COMPAQ HSV110 (C)COMPAQ  3028
dka0.0.0.2004.0         DKA0          COMPAQ BD03685A24  HPB7
dka100.1.0.2004.0       DKA100        COMPAQ BD01864552  3B08
dka200.2.0.2004.0       DKA200        COMPAQ BD00911934  3B00
dqa0.0.0.15.0           DQA0          HL-DT-ST CD-ROM GCR-8480  2.11
dva0.0.0.1000.0         DVA0
eia0.0.0.2005.0         EIA0          00-06-2B-03-2D-7D
```

```

pga0.0.0.3.0          PGA0          WWN 1000-0000-c92a-78e9
pka0.7.0.2004.0      PKA0          SCSI Bus ID 7
pkb0.6.0.2.0        PKB0          SCSI Bus ID 6 5.57
P00>>>

```

From the output, the LAN interface will be EIA0 on which IP address is configured and used for cluster configuration.

Note

The Alpha console uses the MOP protocol for network load of satellite systems. Because the MOP protocol is non-routable, the satellite boot server or servers and all satellites booting from them must reside in the same LAN. In addition, the boot server must have at least one LAN device enabled for cluster communications to permit the Alpha satellite nodes to access the system disk.

On Integrity server systems, the interface name either starts with 'EI' or 'EW'. If it is the first interface, it will be EIA0 or EWA0. Note the mac address of the interface that you want to use from the Shell prompt. To obtain the interface information on Integrity servers, execute the following command on EFI Shell:

```
Shell> lanaddress
```

```
LAN Address Information
```

LAN Address	Path
Mac(00306E4A133F)	Acpi(HWP0002,0)/Pci(3 0)/Mac(00306E4A133F))
*Mac(00306E4A02F9)	Acpi(HWP0002,100)/Pci(2 0)/Mac(00306E4A02F9))

```
Shell>
```

Assuming that the interface which is active is EIA0, configure the satellite with EIA0, if it does not boot with EIA0 then try with EWA0 subsequently.

Step 2. Executing CLUSTER_CONFIG_LAN.COM

Execute CLUSTER_CONFIG_LAN on the boot server node TULIP and select the appropriate option as described in the following example:

```

TULIP$ @SYS$SYSROOT:[SYSMGR]CLUSTER_CONFIG_LAN.COM
                Cluster/IPCI Configuration Procedure
                CLUSTER_CONFIG_LAN Version V2.84
                Executing on an IA64 System

    DECnet-Plus is installed on this node.
    IA64 satellites will use TCP/IP BOOTP and TFTP services for downline
loading

    TCP/IP is installed and running on this node.

    Enter a "?" for help at any prompt.  If you are familiar with
    the execution of this procedure, you may want to mute extra notes
    and explanations by invoking it with "@CLUSTER_CONFIG_LAN BRIEF".

    TULIP is an IA64 system and currently a member of a cluster
    so the following functions can be performed:

```

```
MAIN Menu
```

1. ADD an IA64 node to the cluster.
2. REMOVE a node from the cluster.
3. CHANGE a cluster member's characteristics.
4. CREATE a duplicate system disk for TULIP.
5. MAKE a directory structure for a new root on a system disk.
6. DELETE a root from a system disk.
7. EXIT from this procedure.

Enter choice [7]:

This ADD function will add a new IA64 node to the cluster.

WARNING: If the node being added is a voting member, EXPECTED_VOTES for every cluster member must be adjusted. For complete instructions check the section on configuring a cluster in the "OpenVMS Cluster Systems" manual.

CAUTION: If this cluster is running with multiple system disks and common system files will be used, please, do not proceed unless appropriate logical names are defined for cluster common files in SYLOGICALS.COM. For instructions, refer to the "OpenVMS Cluster Systems" manual.

Do you want to continue [Y]? Y

Is the node to be a clustered node with a shared SCSI/FIBRE-CHANNEL bus (Y/N)? N

Will the node be a satellite [Y]? [Return]

What is the node's SCS node name? JASMIN

What is the node's SCSSYSTEMID number? 25482

WARNING:

DECnet is not running.

No DECnet databases will be updated with information on JASMIN.

Does JASMIN need to be registered in the DECnet namespace [N]?[Return]

What is the Cluster Alias fullname?

This procedure will now ask you for the device name of JASMIN's system root.

The default device name (DISK\$TULIPSYS:) is the logical volume name of SYS\$SYSDEVICE:.

What is the device name for JASMIN's system root [default DISK\$TULIPSYS:]? [Return]

What is the name of JASMIN's system root [SYS14]? [Return]

What is JASMIN's LAN adapter hardware address? 00-30-6E-4A-02-F9 ❶

What is JASMIN's TCP/IP address [10.0.2.3]? [Return] ❷

What is JASMIN's TCP/IP gateway or gateways (leave blank if none)?

10.0.2.1 ❸

What is JASMIN's TCP/IP network mask [255.255.255.0]? 255.255.254.0 ❹

NOTE: Make sure to set the VMS_FLAGS console variable to 0,200000 on node JASMIN so it will use

the memory-disk method to boot as a satellite.
 The command to update this variable from the
 console EFI shell of JASMIN is:
 set vms_flags "0,200000"

Allow conversational bootstraps on JASMIN [N]? [Return]

The following workstation windowing options are available:

1. No workstation software
2. DECwindows Workstation Software

Enter choice [1]:

Creating directory tree SYS14 ...
 System root SYS14 created

ENABLE IP for cluster communications (Y/N)? Y ⑤

UDP port number to be used for Cluster Communication over IP[49152]?
 [Return] ⑥

Enable IP multicast for cluster communication(Y/N)[Y]? Y ⑦

What is IP the multicast address[224.0.0.3]? [Return]

What is the TTL (time to live) value for IP multicast packets [1] ? 32]?
 [Return] ⑧

Do you want to enter unicast address(es)(Y/N)[Y]? Y ⑨

What is the unicast address[Press [RETURN] to end the list]? 10.0.2.3

What is the unicast address[Press [RETURN] to end the list]? 10.0.2.2

What is the unicast address[Press [RETURN] to end the list]? 10.0.1.2

What is the unicast address[Press [RETURN] to end the list]? [Return]

```
*****
Cluster Communications over IP has been enabled. Now
CLUSTER_CONFIG_LAN will run the SYS$MANAGER:TCPIP$CONFIG
procedure. Please select the IP interfaces to be used for
Cluster Communications over IP (IPCI). This can be done
selecting "Core Environment" option from the main menu
followed by the "Interfaces" option. You may also use
this opportunity to configure other aspects.
*****
```

Press Return to continue ...

Checking TCP/IP Services for OpenVMS configuration database files.

Configuration options:

- 1 - Core environment
- 2 - Client components
- 3 - Server components
- 4 - Optional components
- 5 - Shutdown TCP/IP Services for OpenVMS
- 6 - Startup TCP/IP Services for OpenVMS
- 7 - Run tests
- A - Configure options 1 - 4
- [E] - Exit configuration procedure

Enter configuration option: 1

TCP/IP Services for OpenVMS Core Environment Configuration Menu

Configuration options:

- 1 - Domain
- 2 - Interfaces
- 3 - Routing
- 4 - BIND Resolver
- 5 - Time Zone
- A - Configure options 1 - 5
- [E] - Exit menu

Enter configuration option: 2

TCP/IP Services for OpenVMS Interface & Address Configuration Menu

Hostname Details: Configured=[], Active=[]

Configuration options:

- 0 - Set The Target Node (Current Node: TULIP)
- 1 - IE0 Menu (EIA0: TwistedPair 100mbps)
- 2 - 15.146.235.222/23 *noname* Configured
- 3 - 15.146.235.254/23 [] IPCI
- 4 - IE1 Menu (EIB0: TwistedPair 100mbps)
- 5 - 15.146.235.222/23 *noname* Configured,Active
- I - Information about your configuration
- [E] - Exit menu

Enter configuration option: 0 ⑩

Enter name of node to manage [TULIP]: JASMIN

JASMIN is not currently a cluster member.

* Continue configuring JASMIN [NO]: YES ⑪

Enter system device for JASMIN [DSA2:]: ⑫

Enter system root for JASMIN []: SYS14 ⑬

TCP/IP Services for OpenVMS Interface & Address Configuration Menu

Hostname Details: Configured=JASMIN

Configuration options:

- 0 - Set The Target Node (Current Node: JASMIN - DSA2:[SYS14])
- A - Add an Interface
- [E] - Exit menu

Enter configuration option: a

Enter controller name (e.g. EIA or EWC, etc): [ENTER when done] EIA ⑭

Controller Name : EIA

TCP/IP Interface Name : IE0

* Is this correct [NO]: y

Interface Menu:IE0

TCP/IP Services for OpenVMS Interface IE0 Configuration Menu (Node: JASMIN)

Configuration options:

- 1 - Add a primary address on IE0
- 2 - Add an alias address on IE0
- 3 - Enable DHCP client to manage address on IE0
- [E] - Exit menu

Enter configuration option: 1 **1**

* Is this address used by Clusters over IP (IPCI) [NO]: Y **1**
IPv4 Address may be entered with CIDR bits suffix.
E.g. For a 16-bit netmask enter 10.0.1.1/16

Enter IPv4 Address []: 10.0.2.3

Default netmask calculated from class of IP address: 255.0.0.0

IPv4 Netmask may be entered in dotted decimal notation,
(e.g. 255.255.0.0), or as number of CIDR bits (e.g. 16)

Enter Netmask or CIDR bits [255.0.0.0]:

Enter hostname []: JASMIN

Requested configuration:

Node : JASMIN
Interface: IE0
IPCI : Yes
Address : 10.0.2.3/8
Netmask : 255.0.0.0 (CIDR bits: 8)
Hostname : jasmin

* Is this correct [YES]:

Added hostname jasmin (10.0.2.3) to host database

NOTE:

The system hostname is not configured.
It will now be set to jasmin (10.0.2.3).
This can be changed later via the Interface Configuration Menu.

Updated system hostname in configuration database

Added address IE1:10.0.2.3 to configuration database

Updated Interface in IPCI configuration file: DSA2:[SYS14.SYSEXE]TCPIP
\$CLUSTER.DAT;

Updated Default Route in IPCI configuration file: DSA2:[SYS14.SYSEXE]TCPIP
\$CLUSTER.DAT;

TCP/IP Services for OpenVMS Interface & Address Configuration Menu

Hostname Details: Configured=JASMIN

Configuration options:

- 0 - Set The Target Node (Current Node: JASMIN - DSA2:[SYS14.])
- 1 - IE1 Menu (EIB0:)
- 2 - 10.0.2.3/8 JASMIN Configured,IPCI
- I - Information about your configuration

- A - Add an Interface
- [E] - Exit menu

Enter configuration option:

TCP/IP Services for OpenVMS Core Environment Configuration Menu

Configuration options:

- 1 - Domain
- 2 - Interfaces
- 3 - Routing
- 4 - BIND Resolver
- 5 - Time Zone
- A - Configure options 1 - 5
- [E] - Exit menu

Enter configuration option: E

TCP/IP Services for OpenVMS Configuration Menu

Configuration options:

- 1 - Core environment
- 2 - Client components
- 3 - Server components
- 4 - Optional components
- 5 - Shutdown TCP/IP Services for OpenVMS
- 6 - Startup TCP/IP Services for OpenVMS
- 7 - Run tests
- A - Configure options 1 - 4
- [E] - Exit configuration procedure

Enter configuration option: E

SYS\$SYSTEM:PE\$IP_CONFIG.DAT file generated in node JASMIN's root shown below

```
! CLUSTER_CONFIG_LAN creating for CHANGE operation on 15-JUL-2008
15:23:56.05
multicast_address=224.0.0.3
ttl=32
udp_port=49152
unicast=10.0.2.3
unicast=10.0.2.2
unicast=10.0.1.2
```

SYS\$SYSTEM:TCPIP\$CLUSTER.DAT file generated in node JASMIN's root shown below

```
interface=IE0,EIA0,10.0.2.3,255.0.0.0
default_route=16.116.40.1
```

Note

Assuming that the interface which is active is EIA, configure the satellite with EIA, if it does not boot with EIA then try with EIB subsequently. If the wrong interface name is given, satellite node fails with the messages during booting.

- ❶ Enter the LAN adapter's hardware address.
- ❷ Enter the TCP/IP address.
- ❸ Enter the TCP/IP gateway.
- ❹ Enter the TCP/IP network mask address.
- ❺ Enable IP for cluster communication.
- ❻ The UDP port number to be used for cluster communication. The UDP port number must be same on all members of the cluster. Also, ensure that there is no other cluster in your environment using the same UDP port number and this port number must not be used by any other application.
- ❼ Enter the IP multicast address for cluster, if IP multicasting is enabled. By default, the IP multicast address is selected from the administratively scoped IP multicast address range of 239.242.x.y. The last two octets x and y are generated based on the cluster group number. In the above example, the cluster group number is 1985 and can be calculated as follows:

$$X = 1985 / 256$$
$$Y = 1985 - (256 * X)$$

The system administrator can override the default multicast address with a unique address for their environment. The multicast address is modified based on the cluster group number or it can be added to .DAT file.
- ❽ TTL is the time-to-live for IP multicast packets. It specifies the number of hops allowed for IP multicast packets.
- ❾ Enter "yes" to enter the IP Unicast address of remote nodes of the cluster, which are not reachable using IP multicast address.
- ❿ In the TCP/IP configuration, select option 0 to set the target node to JASMIN, which is the satellite node, and will be added to the cluster.
- ⓫ Proceed with configuration steps to configure node JASMIN.
- ⓬ Enter the system device for JASMIN, which DSA2.
- ⓭ Enter JASMIN's root, which SYS14.
- ⓮ Enter the controller information on which IP will be configured for cluster communication. The controller information is obtained from the console of JASMIN as explained in the beginning of the configuration.
- ⓯ Select an option to add a primary address for IE0 (IP interface name of controller EIA).

- ⑩ Enable the use of IE0 for Cluster over IP and proceed with the rest of the configuration.

Step 3. Executing the CLUSTER_CONFIG_LAN.COM Procedure

Continue to run the CLUSTER_CONFIG_LAN.COM to complete the cluster configuration procedure.

```
Adjusting protection on DSA2:[SYS14.][SYSEXEXE]PE$IP_CONFIG.DAT;1
Will JASMIN be a disk server [N]? Y
Enter a value for JASMIN's ALLOCLASS parameter [0]: 15
Updating BOOTP database with satellite information for JASMIN..
Size of pagefile for JASMIN [RETURN for AUTOGEN sizing]?
```

A temporary pagefile will be created until resizing by AUTOGEN. The default size below is arbitrary and may or may not be appropriate.

```
Size of temporary pagefile [10000]? [Return]
Size of swap file for JASMIN [RETURN for AUTOGEN sizing]? [Return]
```

A temporary swap file will be created until resizing by AUTOGEN. The default size below is arbitrary and may or may not be appropriate.

```
Size of temporary swap file [8000]? [Return]
```

NOTE: IA64 satellite node JASMIN requires DOSD if capturing the system state in a dumpfile is desired after a system crash.

```
Will a local disk on JASMIN be used for paging and swapping (Y/N)? N
```

If you specify a device other than DISK\$TULIPSYS: for JASMIN's page and swap files, this procedure will create PAGEFILE_JASMIN.SYS and SWAPFILE_JASMIN.SYS in the <SYSEXEXE> directory on the device you specify.

```
What is the device name for the page and swap files [DISK$TULIPSYS:]?
[Return]
%SYSGEN-I-CREATED, DSA2:<SYS14.SYSEXEXE>PAGEFILE.SYS;1 created
%SYSGEN-I-CREATED, DSA2:<SYS14.SYSEXEXE>SWAPFILE.SYS;1 created
The configuration procedure has completed successfully.
```

The node JASMIN is configured to join the cluster. After the first boot of JASMIN, AUTOGEN.COM will run automatically.

Step 4. Updating the PE\$IP_CONFIG.DAT File

To ensure that the nodes join the cluster, PE\$IP_CONFIG.DAT must be consistent through all the members of the cluster. Copy the SYSS\$SYSTEM:PE\$IP_CONFIG.DAT file that is created on node JASMIN's root to the other nodes, ORCHID and TULIP.

Step 5. Refreshing the Unicast list

On both ORCHID and TULIP, to update the new unicast list in the PE\$IP_CONFIG.DAT file, enter the following command for PEDRIVER:

```
$MC SCACP RELOAD
```

You can also use SYSMAN and run the command cluster wide.

Note

The following rule is applicable when IP unicast address is used for node discovery. A node is allowed to join the cluster only if its IP address is present in the existing members of the SYS\$SYSTEM:PE\$IP_CONFIG.DAT file.

Step 6. Running AUTOGEN and Rebooting the Node

After the first boot of JASMIN, AUTOGEN.COM runs automatically. JASMIN will now be able to join the existing cluster consisting of nodes ORCHID and TULIP.

```
JASMIN$ @SYS$UPDATE:AUTOGEN GETDATA REBOOT
```

8.2.3.5. Adding an Integrity server Node to a Cluster over IP with Logical LAN Failover set

This section describes how to add a node, ORCHID to an existing two-node cluster, JASMIN and TULIP. The Logical LAN failover set is created and configured on ORCHID. ORCHID can survive failure if a local LAN card fails and it will switchover to other interface configured in the logical LAN failover set.

Step 1. Configuring the Logical LAN Failover set

Execute the following commands to create a logical LAN failover set.

```
$ MC LANCP
LANCP>DEFINE DEVICE LLB/ENABLE/FAILOVER=(EIA0, EIB0)
```

Reboot the system and during reboot, the following console message is displayed:

```
%LLB0, Logical LAN event at 2-SEP-2008 14:52:50.06
%LLB0, Logical LAN failset device created
```

Step 2: Executing CLUSTER_CONFIG_LAN

Execute CLUSTER_CONFIG_LAN.COM on node ORCHID and select the appropriate option as shown:

```
ORCHID$@SYS$MANAGER:CLUSTER_CONFIG_LAN
Cluster/IPCI Configuration Procedure
CLUSTER_CONFIG_LAN Version V2.84
Executing on an IA64 System
```

```
DECnet-Plus is installed on this node.
```

```
IA64 satellites will use TCP/IP BOOTP and TFTP services for downline
loading.
```

```
TCP/IP is installed and running on this node.
```

```
Enter a "?" for help at any prompt. If you are familiar with
the execution of this procedure, you may want to mute extra notes
and explanations by invoking it with "@CLUSTER_CONFIG_LAN BRIEF".
```

```
This IA64 node is not currently a cluster member.
```

MAIN Menu

1. ADD ORCHID to existing cluster, or form a new cluster.
2. MAKE a directory structure for a new root on a system disk.

3. DELETE a root from a system disk.
4. EXIT from this procedure.

```
Enter choice [4]: 1
Is the node to be a clustered node with a shared SCSI/FIBRE-CHANNEL bus (Y/N)? n
What is the node's SCS node name? orchid
    IA64 node, using LAN/IP for cluster communications. PEDRIVER will be
    loaded.
    No other cluster interconnects are supported for IA64 nodes.
Enter this cluster's group number: 1985
Enter this cluster's password:
Re-enter this cluster's password for verification:
```

```
ENABLE IP for cluster communications (Y/N)? Y
UDP port number to be used for Cluster Communication over IP[49152]?
[Return]
Enable IP multicast for cluster communication(Y/N)[Y]? [Return]
What is IP the multicast address[239.242.7.193]? 239.242.7.193
What is the TTL (time to live) value for IP multicast packets [32]?
[Return]
Do you want to enter unicast address(es)(Y/N)[Y]? [Return]
What is the unicast address[Press [RETURN] to end the list]? 10.0.1.2
What is the unicast address[Press [RETURN] to end the list]? 10.0.2.3
What is the unicast address[Press [RETURN] to end the list]? 10.0.2.2
What is the unicast address[Press [RETURN] to end the list]? [Return]
```

```
*****
Cluster Communications over IP has been enabled. Now
CLUSTER_CONFIG_LAN will run the SYS$MANAGER:TCPIP$CONFIG
procedure. Please select the IP interfaces to be used for
Cluster Communications over IP (IPCI). This can be done
selecting "Core Environment" option from the main menu
followed by the "Interfaces" option. You may also use
this opportunity to configure other aspects.
```

```
*****
Press Return to continue ...
```

TCP/IP Network Configuration Procedure

This procedure helps you define the parameters required to run TCP/IP Services for OpenVMS on this system.

```
%TCPIP-I-IPCI, TCP/IP Configuration is limited to IPCI.
-TCPPIP-I-IPCI, Rerun TCPIP$CONFIG after joining the cluster.
```

TCP/IP Services for OpenVMS Interface & Address Configuration Menu

Hostname Details: Configured=Not Configured, Active=nodeg

Configuration options:

- 0 - Set The Target Node (Current Node: ORCHID)
- 1 - LEO Menu (LLA0: TwistedPair 100mbps)
- 2 - IE1 Menu (EIB0: TwistedPair 100mbps)
- [E] - Exit menu

Enter configuration option: 1

* IPCI Address Configuration *

Only IPCI addresses can be configured in the current environment.
After configuring your IPCI address(es) it will be necessary to
run TCPIP\$CONFIG once your node has joined the cluster.

IPv4 Address may be entered with CIDR bits suffix.
E.g. For a 16-bit netmask enter 10.0.1.1/16

Enter IPv4 Address []:10.0.1.2
Default netmask calculated from class of IP address: 255.0.0.0

IPv4 Netmask may be entered in dotted decimal notation,
(e.g. 255.255.0.0), or as number of CIDR bits (e.g. 16)

Enter Netmask or CIDR bits [255.0.0.0]: 255.255.255.0

Requested configuration:

Node : ORCHID
Interface: IE0
IPCI : Yes
Address : 10.0.1.2/24
Netmask : 255.255.254.0 (CIDR bits: 23)

* Is this correct [YES]:

Updated Interface in IPCI configuration file: SYS\$SYSROOT:[SYSEXE]TCPIP
\$CLUSTER.
DAT;

TCP/IP Services for OpenVMS Interface & Address Configuration Menu

Hostname Details: Configured=Not Configured, Active=nodeg

Configuration options:

0 - Set The Target Node (Current Node: ORCHID)
1 - LE0 Menu (LLA0: TwistedPair 100mbps)
2 - 10.0.1.2/24 ORCHID IPCI
3 - IE1 Menu (EIB0: TwistedPair 100mbps)
[E] - Exit menu

Enter configuration option: E

Enter your Default Gateway address []: 10.0.1.1

* The default gateway will be: 10.0.1.1 Correct [NO]: YES

Updated Default Route in IPCI configuration file: SYS\$SYSROOT:[SYSEXE]TCPIP
\$CLUSTER.DAT;

TCPIP-I-IPCIDONE, Finished configuring IPCI address information.

SYS\$SYSTEM:PE\$IP_CONFIG.DAT file generated in node ORCHID's root shown
below

```
! CLUSTER_CONFIG_LAN creating for CHANGE operation on 15-JUL-2008
 15:23:56.05
multicast_address=239.242.7.193
ttl=32
udp_port=49152
unicast=10.0.2.3
unicast=10.0.2.2
unicast=10.0.1.2
```

SYSS\$SYSTEM:TCPIP\$CLUSTER.DAT file generated in node ORCHID's root shown below

```
interface=LE1,LLB0,10.0.1.2,255.0.0.0
default_route=10.0.1.1
```

Step 3. Completing the Configuration Procedure

Continue to run the CLUSTER_CONFIG_LAN.COM to complete the cluster configuration procedure. For more information, see Section 8.2.3.1.

Step 4. Updating the PE\$IP_CONFIG.DAT file

To ensure that the nodes join the cluster, PE\$IP_CONFIG.DAT must be consistent through all the members of the cluster. Copy the SYSS\$SYSTEM:PE\$IP_CONFIG.DAT file that is created on node JASMIN to the other nodes ORCHID and TULIP.

Step 5. Refreshing the Unicast list

On both ORCHID and TULIP, to update the new unicast list in the PE\$IP_CONFIG.DAT file, enter the following command for PEDRIVER:

```
$MC SCACP RELOAD
```

You can also use SYSMAN and run the command cluster wide.

Step 6. Running AUTOGEN and Rebooting the Node

After the first boot of ORCHID, AUTOGEN.COM will run automatically. ORCHID will now be able to join the existing cluster consisting of nodes JASMIN and LOTUS.

```
ORCHID$ @SYS$UPDATE:AUTOGEN GETDATA REBOOT
```

8.2.4. Adding a Quorum Disk

To enable a quorum disk on a node or nodes, use the cluster configuration procedure as described in Table 8.5.

Table 8.5. Preparing to Add a Quorum Disk Watcher

IF...	THEN...
Other cluster nodes are already enabled as quorum disk watchers.	Perform the following steps: <ol style="list-style-type: none"> 1. Log in to the computer that is to be enabled as the quorum disk watcher and run CLUSTER_CONFIG_LAN.COM or CLUSTER_CONFIG.COM.

IF...	THEN...
	<ol style="list-style-type: none"> 2. Execute the CHANGE function and select menu item 7 to enable a quorum disk. (See Section 8.4). 3. Update the current system parameters and reboot the node. (See Section 8.6.1).
The cluster does not contain any quorum disk watchers.	<p>Perform the following steps:</p> <ol style="list-style-type: none"> 1. Perform the preceding steps 1 and 2 for each node to be enabled as a quorum disk watcher. 2. Reconfigure the cluster according to the instructions in Section 8.6.

8.3. Removing Computers

To disable a computer as an OpenVMS Cluster member:

1. Determine whether removing a member will cause you to lose quorum. Use the SHOW CLUSTER command to display the CL_QUORUM and CL_VOTES values.

IF removing members...	THEN...
Will cause you to lose quorum	<p>Perform the steps in the following list:</p> <p>Caution: Do not perform these steps until you are ready to reboot the entire OpenVMS Cluster system. Because you are reducing quorum for the cluster, the votes cast by the node being removed could cause a cluster partition to be formed.</p> <ul style="list-style-type: none"> • Reset the EXPECTED_VOTES parameter in the AUTOGEN parameter files and current system parameter files (see Section 8.6.1). • Shut down the cluster (see Section 8.6.2), and reboot without the node that is being removed. <p>Note: Be sure that you do not specify an automatic reboot on that node.</p>
Will not cause you to lose quorum	<p>Proceed as follows:</p> <ul style="list-style-type: none"> • Perform an orderly shutdown on the node being removed by invoking the SYSSYSTEM:SHUTDOWN.COM command procedure (described in Section 8.6.3). • If the node was a voting member, use the DCL command SET CLUSTER/EXPECTED_VOTES to reduce the value of quorum.

Reference: Refer also to Section 10.11 for information about adjusting expected votes.

2. Invoke CLUSTER_CONFIG_LAN.COM or CLUSTER_CONFIG.COM on an active OpenVMS Cluster computer and select the REMOVE option.
3. Use the information in Table 8.6 to determine whether additional actions are required.

Table 8.6. Preparing to Remove Computers from an OpenVMS Cluster

IF...	THEN...
You are removing a voting member.	You must, after the REMOVE function completes, reconfigure the cluster according to the instructions in Section 8.6.
The page and swap files for the computer being removed do not reside on the same disk as the computer's root directory tree.	The REMOVE function does not delete these files. It displays a message warning that the files will not be deleted, as in Example 8.6. If you want to delete the files, you must do so after the REMOVE function completes.
You are removing a computer from a cluster that uses DECdtm services.	Make sure that you have followed the step-by-step instructions in the chapter on DECdtm services in the <i>VSI OpenVMS System Manager's Manual</i> . These instructions describe how to remove a computer safely from the cluster, thereby preserving the integrity of your data.

Note: When the REMOVE function deletes the computer's entire root directory tree, it generates OpenVMS RMS informational messages while deleting the directory files. You can ignore these messages.

8.3.1. Example

Example 8.6 illustrates the use of CLUSTER_CONFIG_LAN.COM on BHAGAT to remove satellite GOMTHI from the cluster.

Example 8.6. Sample Interactive CLUSTER_CONFIG_LAN.COM Session to Remove a Satellite with Local Page and Swap Files

```

$ @CLUSTER_CONFIG_LAN.COM
      Cluster/IPCI Configuration Procedure
      CLUSTER_CONFIG_LAN Version V2.84
      Executing on an IA64 System

DECnet-Plus is installed on this node.
IA64 satellites will use TCP/IP BOOTP and TFTP services for downline
loading.
TCP/IP is installed and running on this node.

      Enter a "?" for help at any prompt.  If you are familiar with
      the execution of this procedure, you may want to mute extra notes
      and explanations by invoking it with "@CLUSTER_CONFIG_LAN BRIEF".

BHAGAT is an IA64 system and currently a member of a cluster
so the following functions can be performed:

MAIN Menu

1. ADD an IA64 node to the cluster.
2. REMOVE a node from the cluster.
3. CHANGE a cluster member's characteristics.
4. CREATE a duplicate system disk for BHAGAT.
5. MAKE a directory structure for a new root on a system disk.

```

6. DELETE a root from a system disk.
7. EXIT from this procedure.

Enter choice [7]: 2

The REMOVE command disables a node as a cluster member.

- o It deletes the node's root directory tree.
- o If the node has entries in SYS\$DEVICES.DAT, any port allocation class for shared SCSI bus access on the node must be re-assigned.

If the node being removed is a voting member, EXPECTED_VOTES in each remaining cluster member's MODPARAMS.DAT must be adjusted. The cluster must then be rebooted.

For instructions, see the "OpenVMS Cluster Systems" manual.

CAUTION: The REMOVE command does not remove the node name from any network databases. Also, if a satellite has been set up for booting with multiple hardware addresses, the satellite's aliases are not cleared from the LANACP boot database.

```
What is the node's SCS node name? GOMTHI
Verifying BOOTP satellite node database...
Verifying that $1$DKA0:[SYS10] is GOMTHI's root...
Are you sure you want to remove node GOMTHI (Y/N)? Y
```

```
WARNING: GOMTHI's page and swap files will not be deleted.
They do not reside on $1$DKA0:.
```

```
Deleting directory tree $1$DKA0:<SYS10...>
```

```
%DELETE-I-FILDEL, $1$DKA0:<SYS10.SYS$I18N.LOCALES>SYSTEM.DIR;1 deleted (16
blocks)
.
.
.
.
%DELETE-I-FILDEL, $1$DKA0:<SYS10>VPM$SERVER.DIR;1 deleted (16 blocks)
%DELETE-I-TOTAL, 21 files deleted (336 blocks)
%DELETE-I-FILDEL, $1$DKA0:<0,0>SYS10.DIR;1 deleted (16 blocks)
```

```
System root $1$DKA0:<SYS10> deleted.
Updating BOOTP database...
Removing rights identifier for GOMTHI...
The configuration procedure has completed successfully.
```

8.3.2. Removing a Quorum Disk

To disable a quorum disk on a node or nodes, use the cluster configuration command procedure as described in Table 8.7.

Table 8.7. Preparing to Remove a Quorum Disk Watcher

IF...	THEN...
Other cluster nodes will still be enabled as quorum disk watchers.	Perform the following steps: <ol style="list-style-type: none"> 1. Log in to the computer that is to be disabled as the quorum disk watcher and run CLUSTER_CONFIG_LAN.COM or CLUSTER_CONFIG.COM. 2. Execute the CHANGE function and select menu item 7 to disable a quorum disk (see Section 8.4). 3. Reboot the node (see Section 8.6.7).
All quorum disk watchers will be disabled.	Perform the following steps: <ol style="list-style-type: none"> 1. Perform the preceding steps 1 and 2 for all computers with the quorum disk enabled. 2. Reconfigure the cluster according to the instructions in Section 8.6.

8.4. Changing Computer Characteristics

As your processing needs change, you may want to add satellites to an existing OpenVMS Cluster, or you may want to change an OpenVMS Cluster that is based on one interconnect (such as the CI or DSSI interconnect, or HSC subsystem) to include several interconnects.

Table 8.8 describes the operations you can accomplish when you select the CHANGE option from the main menu of the cluster configuration command procedure.

Note: All operations except changing a satellite's LAN (Ethernet or FDDI) hardware address must be executed on the computer whose characteristics you want to change.

Table 8.8. CHANGE Options of the Cluster Configuration Procedure

Option	Operation Performed
Enable the local computer as a disk server	Loads the MSCP server by setting, in MODPARAMS.DAT, the value of the MSCP_LOAD parameter to 1 and the MSCP_SERVE_ALL parameter to 1 or 2.
Disable the local computer as a disk server	Sets MSCP_LOAD to 0.
Enable the local computer as a boot server	If you are setting up an OpenVMS Cluster that includes satellites, you must perform this operation once before you attempt to add satellites to the cluster. You thereby enable MOP service for the LAN adapter circuit that the computer uses to service operating system load requests from satellites. When you enable the computer as a boot server, it automatically becomes a disk server (if it is not one already) because it must serve its system disk to satellites.
Disable the local computer as a boot server	Disables DECnet MOP service for the computer's adapter circuit.

Option	Operation Performed
Enable IP for cluster communications on the local computer	Loads the port driver PEDRIVER by setting the value of the NISCS_LOAD_PEA0 parameter to 1 in MODPARAMS.DAT. Also, sets the value of NISCS_USE_UDP to 1 in MODPARAMS.DAT, which enables loading of the IP configuration files. Creates the cluster security database file, SYSSYSTEM:[SYSEXE]CLUSTER_AUTHORIZE.DAT, on the local computer's system disk.
Disable IP for cluster communications on the local computer	Sets NISCS_USE_UDP to 0.
Enable the LAN for cluster communications on the local computer	Loads the port driver PEDRIVER by setting the value of the NISCS_LOAD_PEA0 parameter to 1 in MODPARAMS.DAT. Creates the cluster security database file, SYSSYSTEM:[SYSEXE]CLUSTER_AUTHORIZE.DAT, on the local computer's system disk. Caution: The VAXCLUSTER system parameter must be set to 2 if the NISCS_LOAD_PEA0 parameter is set to 1. This ensures coordinated access to shared resources in the cluster and prevents accidental data corruption.
Disable the LAN for cluster communications on the local computer	Sets NISCS_LOAD_PEA0 to 0.
Enable a quorum disk on the local computer	In MODPARAMS.DAT, sets the DISK_QUORUM system parameter to a device name; sets the value of QDSKVOTES to 1 (default value).
Disable a quorum disk on the local computer	In MODPARAMS.DAT, sets a blank value for the DISK_QUORUM system parameter; sets the value of QDSKVOTES to 1.
Change a satellite's LAN hardware address	Changes a satellite's hardware address if its LAN device needs replacement. Both the permanent and volatile network databases and NETNODE_UPDATE.COM are updated on the local computer. Rule: <i>You must perform this operation on each computer enabled as a boot server for the satellite.</i>
Enable the local computer as a tape server	Loads the TMSCP server by setting, in MODPARAMS.DAT, the value of the TMSCP_LOAD parameter to 1 and the TMSCP_SERVE_ALL parameter to 1 or 2.
Disable the local computer as a tape server	Sets TMSCP_LOAD to zero.
Change the local computer's node allocation class value	Sets a value for the computer's ALLOCLASS parameter in MODPARAMS.DAT.
Change the local computer's tape allocation class value	Sets a value from 1 to 255 for the computer's TAPE_ALLOCLASS parameter in MODPARAMS.DAT. The default value is zero. You must specify a nonzero tape allocation class parameter if this node is locally connected to a dual-ported tape, or if it will be serving any multiple-host tapes (for example, TF nn or HSC connected tapes) to other cluster members. Satellites usually have TAPE_ALLOCLASS set to zero.

Option	Operation Performed
Change the local computer's port allocation class value	Sets a value for the computer's ALLOCLASS parameter in MODPARAMS.DAT for all devices attached to it.
Enable MEMORY CHANNEL	Sets MC_SERVICES_P2 to 1 to load the PMDRIVER (PMA0) cluster driver. This system parameter enables MEMORY CHANNEL on the local computer for node-to-node cluster communications.
Disable MEMORY CHANNEL	Sets MC_SERVICES_P2 to 0 so that the PMDRIVER (PMA0) cluster driver is not loaded. The setting of 0 disables MEMORYCHANNEL on the local computer as the node-to-node cluster communications interconnect.

8.4.1. Preparation

You usually need to perform a number of steps before using the cluster configuration command procedure to change the configuration of your existing cluster.

Table 8.9 suggests several typical configuration changes and describes the procedures required to make them.

Table 8.9. Tasks Involved in Changing OpenVMS Cluster Configurations

Task	Procedure
Add satellite nodes	<p>Perform these operations on the computer that will be enabled as a cluster boot server:</p> <ol style="list-style-type: none"> 1. Execute the CHANGE function to enable the first installed computer as a boot server (see Example 8.9). 2. Execute the ADD function to add the satellite (as described in Section 8.2). 3. Reconfigure the cluster according to the postconfiguration instructions in Section 8.6.
Change an existing CI or DSSI cluster to include satellite nodes	<p>To enable cluster communications over the LAN (Ethernet or FDDI) on all computers, and to enable one or more computers as boot servers, proceed as follows:</p> <ol style="list-style-type: none"> 1. Log in as system manager on each computer, invoke either CLUSTER_CONFIG_LAN.COM or CLUSTER_CONFIG.COM, and execute the CHANGE function to enable LAN communications. <p>Rule: <i>You must perform this operation on all computers.</i></p> <p>Note: You must establish a cluster group number and password on all system disks in the OpenVMS Cluster before you can successfully add a satellite node using the CHANGE function of the cluster configuration procedure.</p> <ol style="list-style-type: none"> 2. Execute the CHANGE function to enable one or more computers as boot servers. 3. Reconfigure the cluster according to the postconfiguration instructions in Section 8.6.

Task	Procedure
Change an existing LAN-based cluster to include CI and DSSI interconnects	<p>Before performing the operations described here, be sure that the computers and HSC subsystems or RF disks you intend to include in your new configuration are correctly installed and checked for proper operation.</p> <p>The method you use to include CI and DSSI interconnects with an existing LAN-based cluster configuration depends on whether your current boot server is capable of being configured as a CI or DSSI computer.</p> <p>Note: The following procedures assume that the system disk containing satellite roots will reside on an HSC disk (for CI configurations) or an RF disk (for DSSI configurations).</p> <ul style="list-style-type: none"> • If the boot server can be configured as a CI or DSSI computer, proceed as follows: <ol style="list-style-type: none"> 1. Log in as system manager on the boot server and perform an image backup operation to back up the current system disk to a disk on an HSC subsystem or RF storage device. (For more information about backup operations, refer to the <i>VSI OpenVMS System Management Utilities Reference Manual</i>). 2. Modify the computer's default bootstrap command procedure to boot the computer from the HSC or RF disk, according to the instructions in the appropriate system-specific installation and operations guide. 3. Shut down the cluster. Shut down the satellites first, and then shut down the boot server. 4. Boot the boot server from the newly created system disk on the HSC or RF storage subsystem. 5. Reboot the satellites. • If your current boot server cannot be configured as a CI or a DSSI computer, proceed as follows: <ol style="list-style-type: none"> 1. Shut down the old local area cluster. Shut down the satellites first, and then shut down the boot server. 2. Install the OpenVMS operating system on the new CI computer's HSCsystem disk or on the new DSSI computer's RF disk, as

Task	Procedure
	<p>appropriate. When the installation procedure asks whether you want to enable the LAN for cluster communications, answer YES.</p> <ol style="list-style-type: none"> 3. When the installation completes, log in as system manager, and configure and start the DECnet for OpenVMS network as described in Chapter 4. 4. Execute the CHANGE function of either CLUSTER_CONFIG_LAN.COM or CLUSTER_CONFIG.COM to enable the computer as a boot server. 5. Log in as system manager on the newly added computer and execute the ADD function of either CLUSTER_CONFIG_LAN.COM or CLUSTER_CONFIG.COM to add the former LAN cluster members (including the former boot server) as satellites. <ul style="list-style-type: none"> • Reconfigure the cluster according to the postconfiguration instructions in Section 8.6.
Convert a standalone computer to an OpenVMS Cluster computer	<p>Execute either CLUSTER_CONFIG_LAN.COM or CLUSTER_CONFIG.COM on a standalone computer to perform either of the following operations:</p> <ul style="list-style-type: none"> • Add the standalone computer with its own system disk to an existing cluster. • Set up the standalone computer to form a new cluster if the computer was not set up as a cluster computer during installation of the operating system. • Reconfigure the cluster according to the postconfiguration instructions in Section 8.6. <p>See Example 8.13, which illustrates the use of CLUSTER_CONFIG.COM on standalone computer PLUTO to convert PLUTO to a cluster boot server.</p> <p>If your cluster uses DECdtm services, you must create a transaction log for the computer when you have configured it into your cluster. For step-by-step instructions on how to do this, see the chapter on DECdtm services in the <i>VSI OpenVMS System Manager's Manual</i>.</p>
Enable or disable disk-serving or tape-serving functions	<p>After invoking either CLUSTER_CONFIG_LAN.COM or CLUSTER_CONFIG.COM to enable or disable the disk or tape serving functions, run AUTOGEN with the REBOOT option to reboot the local computer (see Section 8.6.1).</p>

Note: When the cluster configuration command procedure sets or changes values in MODPARAMS.DAT, the new values are always appended at the end of the file so that they override earlier values. You may want to edit the file occasionally and delete lines that specify earlier values.

8.4.2. Examples

Examples 8.7 through 8.13 illustrate the use of CLUSTER_CONFIG_LAN.COM to perform the following operations:

- Enable a computer as a disk server (Example 8.7).
- Change a computer's ALLOCLASS value (Example 8.8).
- Enable a computer as a boot server (Example 8.9).
- Specify a new hardware address for a satellite node that boots from a common system disk (Example 8.10).
- Enable a computer as a tape server (Example 8.11).
- Change a computer's TAPE_ALLOCLASS value (Example 8.12).
- Convert a standalone computer to a cluster boot server (Example 8.13).

Example 8.7. Sample Interactive CLUSTER_CONFIG_LAN.COM Session to Enable the Local Computer as a Disk Server

```
$ @CLUSTER_CONFIG_LAN.COM
      Cluster/IPCI Configuration Procedure
      CLUSTER_CONFIG_LAN Version V2.84
      Executing on an IA64 System

DECnet-Plus is installed on this node.
IA64 satellites will use TCP/IP BOOTP and TFTP services for downline
loading.
TCP/IP is installed and running on this node.

      Enter a "?" for help at any prompt.  If you are familiar with
      the execution of this procedure, you may want to mute extra notes
      and explanations by invoking it with "@CLUSTER_CONFIG_LAN BRIEF".

BHAGAT is an IA64 system and currently a member of a cluster
so the following functions can be performed:

MAIN Menu

1. ADD an IA64 node to the cluster.
2. REMOVE a node from the cluster.
3. CHANGE a cluster member's characteristics.
4. CREATE a duplicate system disk for BHAGAT.
5. MAKE a directory structure for a new root on a system disk.
6. DELETE a root from a system disk.
7. EXIT from this procedure.

Enter choice [7]: 3

CHANGE Menu

1. Enable BHAGAT as a boot server.
2. Disable BHAGAT as a boot server.
3. Enable a quorum disk for BHAGAT.
4. Disable a quorum disk for BHAGAT.
```

5. Enable BHAGAT as a disk server.
6. Disable BHAGAT as a disk server.
7. Change BHAGAT's ALLOCLASS value.
8. Enable BHAGAT as a tape server.
9. Disable BHAGAT as a tape server.
10. Change BHAGAT's TAPE_ALLOCLASS value.
11. Change an IA64 satellite node's LAN adapter hardware address.
12. Enable Cluster Communication using IP on BHAGAT.
13. Disable Cluster Communication using IP on BHAGAT.
14. Change BHAGAT's shared SCSI port allocation class value.
15. Reset an IA64 satellite node's boot environment file protections.
16. Return to MAIN menu.

Enter choice [16]: 5

Enter a value for BHAGAT's ALLOCLASS parameter [1]:

The configuration procedure has completed successfully.

BHAGAT has been enabled as a disk server. In MODPARAMS.DAT:

```
MSCP_LOAD has been set to 1
MSCP_SERVE_ALL has been set to 2
```

Please run AUTOGEN to reboot BHAGAT:

```
$ @SYS$UPDATE:AUTOGEN GETDATA REBOOT
```

If you have changed BHAGAT's ALLOCLASS value, you must reconfigure the cluster, using the procedure described in the OpenVMS Cluster Systems manual.

Example 8.8. Sample Interactive CLUSTER_CONFIG_LAN.COM Session to Change the Local Computer's ALLOCLASS Value

```
$ @CLUSTER_CONFIG_LAN.COM
```

```
Cluster/IPCI Configuration Procedure
CLUSTER_CONFIG_LAN Version V2.84
Executing on an IA64 System
```

DECnet-Plus is installed on this node.

IA64 satellites will use TCP/IP BOOTP and TFTP services for downline loading.

TCP/IP is installed and running on this node.

Enter a "?" for help at any prompt. If you are familiar with the execution of this procedure, you may want to mute extra notes and explanations by invoking it with "@CLUSTER_CONFIG_LAN BRIEF".

BHAGAT is an IA64 system and currently a member of a cluster so the following functions can be performed:

MAIN Menu

1. ADD an IA64 node to the cluster.
2. REMOVE a node from the cluster.
3. CHANGE a cluster member's characteristics.
4. CREATE a duplicate system disk for BHAGAT.
5. MAKE a directory structure for a new root on a system disk.

6. DELETE a root from a system disk.
7. EXIT from this procedure.

Enter choice [7]: 3

CHANGE Menu

1. Enable BHAGAT as a boot server.
2. Disable BHAGAT as a boot server.
3. Enable a quorum disk for BHAGAT.
4. Disable a quorum disk for BHAGAT.
5. Enable BHAGAT as a disk server.
6. Disable BHAGAT as a disk server.
7. Change BHAGAT's ALLOCLASS value.
8. Enable BHAGAT as a tape server.
9. Disable BHAGAT as a tape server.
10. Change BHAGAT's TAPE_ALLOCLASS value.
11. Change an IA64 satellite node's LAN adapter hardware address.
12. Enable Cluster Communication using IP on BHAGAT.
13. Disable Cluster Communication using IP on BHAGAT.
14. Change BHAGAT's shared SCSI port allocation class value.
15. Reset an IA64 satellite node's boot environment file protections.
16. Return to MAIN menu.

Enter choice [16]: 7

Enter a value for BHAGAT's ALLOCLASS parameter [1]: 2

The configuration procedure has completed successfully.

Since you have changed BHAGAT's ALLOCLASS value, you must reconfigure the cluster, using the procedure described in the "OpenVMS Cluster Systems" manual. This includes running AUTOGEN for BHAGAT as shown below, before rebooting the cluster:

```
$ @SYS$UPDATE:AUTOGEN GETDATA REBOOT
```

If you have changed BHAGAT's ALLOCLASS value, you must reconfigure the cluster, using the procedure described in the OpenVMS Cluster Systems manual.

Example 8.9. Sample Interactive CLUSTER_CONFIG_LAN.COM Session to Enable the Local Computer as a Boot Server

```
$ @CLUSTER_CONFIG_LAN.COM
```

```
Cluster/IPCI Configuration Procedure
CLUSTER_CONFIG_LAN Version V2.84
Executing on an IA64 System
```

DECnet-Plus is installed on this node.

IA64 satellites will use TCP/IP BOOTP and TFTP services for downline loading.

TCP/IP is installed and running on this node.

Enter a "?" for help at any prompt. If you are familiar with the execution of this procedure, you may want to mute extra notes and explanations by invoking it with "@CLUSTER_CONFIG_LAN BRIEF".

BHAGAT is an IA64 system and currently a member of a cluster

so the following functions can be performed:

MAIN Menu

1. ADD an IA64 node to the cluster.
2. REMOVE a node from the cluster.
3. CHANGE a cluster member's characteristics.
4. CREATE a duplicate system disk for BHAGAT.
5. MAKE a directory structure for a new root on a system disk.
6. DELETE a root from a system disk.
7. EXIT from this procedure.

Enter choice [7]: 3

CHANGE Menu

1. Enable BHAGAT as a boot server.
2. Disable BHAGAT as a boot server.
3. Enable a quorum disk for BHAGAT.
4. Disable a quorum disk for BHAGAT.
5. Enable BHAGAT as a disk server.
6. Disable BHAGAT as a disk server.
7. Change BHAGAT's ALLOCLASS value.
8. Enable BHAGAT as a tape server.
9. Disable BHAGAT as a tape server.
10. Change BHAGAT's TAPE_ALLOCLASS value.
11. Change an IA64 satellite node's LAN adapter hardware address.
12. Enable Cluster Communication using IP on BHAGAT.
13. Disable Cluster Communication using IP on BHAGAT.
14. Change BHAGAT's shared SCSI port allocation class value.
15. Reset an IA64 satellite node's boot environment file protections.
16. Return to MAIN menu.

Enter choice [16]: 1

Enter a value for BHAGAT's ALLOCLASS parameter [1]: [Return]
The configuration procedure has completed successfully.

BHAGAT has been enabled as a boot server. Disk serving and LAN capabilities are enabled automatically. If BHAGAT was not previously set up as a disk server, please run AUTOGEN to reboot BHAGAT:

```
$ @SYS$UPDATE:AUTOGEN GETDATA REBOOT
```

If you have changed BHAGAT's ALLOCLASS value, you must reconfigure the cluster, using the procedure described in the OpenVMS Cluster Systems manual.

Example 8.10. Sample Interactive CLUSTER_CONFIG_LAN.COM Session to Change a Satellite's Hardware Address

```
$ @CLUSTER_CONFIG_LAN.COM
      Cluster/IPCI Configuration Procedure
      CLUSTER_CONFIG_LAN Version V2.84
      Executing on an IA64 System
```

DECnet-Plus is installed on this node.

IA64 satellites will use TCP/IP BOOTP and TFTP services for downline loading.

TCP/IP is installed and running on this node.

Enter a "?" for help at any prompt. If you are familiar with the execution of this procedure, you may want to mute extra notes and explanations by invoking it with "@CLUSTER_CONFIG_LAN BRIEF".

BHAGAT is an IA64 system and currently a member of a cluster so the following functions can be performed:

MAIN Menu

1. ADD an IA64 node to the cluster.
2. REMOVE a node from the cluster.
3. CHANGE a cluster member's characteristics.
4. CREATE a duplicate system disk for BHAGAT.
5. MAKE a directory structure for a new root on a system disk.
6. DELETE a root from a system disk.
7. EXIT from this procedure.

Enter choice [7]: 3

CHANGE Menu

1. Enable BHAGAT as a boot server.
2. Disable BHAGAT as a boot server.
3. Enable a quorum disk for BHAGAT.
4. Disable a quorum disk for BHAGAT.
5. Enable BHAGAT as a disk server.
6. Disable BHAGAT as a disk server.
7. Change BHAGAT's ALLOCLASS value.
8. Enable BHAGAT as a tape server.
9. Disable BHAGAT as a tape server.
10. Change BHAGAT's TAPE_ALLOCLASS value.
11. Change an IA64 satellite node's LAN adapter hardware address.
12. Enable Cluster Communication using IP on BHAGAT.
13. Disable Cluster Communication using IP on BHAGAT.
14. Change BHAGAT's shared SCSI port allocation class value.
15. Reset an IA64 satellite node's boot environment file protections.
16. Return to MAIN menu.

Enter choice [16]: 11

What is the node's SCS node name? gomthi

Note: The current hardware address entry for GOMTHI is 00-30-6E-4C-BB-1A.

What is GOMTHI's new LAN adapter hardware address? 00-30-6E-4C-BA-2A

The configuration procedure has completed successfully.

Example 8.11. Sample Interactive CLUSTER_CONFIG_LAN.COM Session to Enable the Local Computer as a Tape Server

```
$ @CLUSTER_CONFIG_LAN.COM
      Cluster/IPCI Configuration Procedure
      CLUSTER_CONFIG_LAN Version V2.84
      Executing on an IA64 System
```

DECnet-Plus is installed on this node.

IA64 satellites will use TCP/IP BOOTP and TFTP services for downline loading.

TCP/IP is installed and running on this node.

Enter a "?" for help at any prompt. If you are familiar with the execution of this procedure, you may want to mute extra notes and explanations by invoking it with "@CLUSTER_CONFIG_LAN BRIEF".

BHAGAT is an IA64 system and currently a member of a cluster so the following functions can be performed:

MAIN Menu

1. ADD an IA64 node to the cluster.
2. REMOVE a node from the cluster.
3. CHANGE a cluster member's characteristics.
4. CREATE a duplicate system disk for BHAGAT.
5. MAKE a directory structure for a new root on a system disk.
6. DELETE a root from a system disk.
7. EXIT from this procedure.

Enter choice [7]: 3

CHANGE Menu

1. Enable BHAGAT as a boot server.
2. Disable BHAGAT as a boot server.
3. Enable a quorum disk for BHAGAT.
4. Disable a quorum disk for BHAGAT.
5. Enable BHAGAT as a disk server.
6. Disable BHAGAT as a disk server.
7. Change BHAGAT's ALLOCLASS value.
8. Enable BHAGAT as a tape server.
9. Disable BHAGAT as a tape server.
10. Change BHAGAT's TAPE_ALLOCLASS value.
11. Change an IA64 satellite node's LAN adapter hardware address.
12. Enable Cluster Communication using IP on BHAGAT.
13. Disable Cluster Communication using IP on BHAGAT.
14. Change BHAGAT's shared SCSI port allocation class value.
15. Reset an IA64 satellite node's boot environment file protections.
16. Return to MAIN menu.

Enter choice [16]: 8

Enter a value for BHAGAT's TAPE_ALLOCLASS parameter [0]: [Return]
Should BHAGAT serve any tapes it sees, local and remote [Y]? [Return]

BHAGAT has been enabled as a tape server. In MODPARAMS.DAT,
TMSCP_LOAD has been set to 1
TMSCP_SERVE_ALL has been set to 1

Please run AUTOGEN to reboot BHAGAT:

```
$ @SYS$UPDATE:AUTOGEN GETDATA REBOOT
```

If you have changed BHAGAT's TAPE_ALLOCLASS value, you must reconfigure

the cluster, using the procedure described in the "OpenVMS Cluster Systems" manual.

Example 8.12. Sample Interactive CLUSTER_CONFIG_LAN.COM Session to Change the Local Computer's TAPE_ALLOCLASS Value

```
$ @CLUSTER_CONFIG_LAN.COM
      Cluster/IPCI Configuration Procedure
      CLUSTER_CONFIG_LAN Version V2.84
      Executing on an IA64 System

DECnet-Plus is installed on this node.
IA64 satellites will use TCP/IP BOOTP and TFTP services for downline
loading.
TCP/IP is installed and running on this node.

      Enter a "?" for help at any prompt.  If you are familiar with
      the execution of this procedure, you may want to mute extra notes
      and explanations by invoking it with "@CLUSTER_CONFIG_LAN BRIEF".

BHAGAT is an IA64 system and currently a member of a cluster
so the following functions can be performed:

MAIN Menu

1. ADD an IA64 node to the cluster.
2. REMOVE a node from the cluster.
3. CHANGE a cluster member's characteristics.
4. CREATE a duplicate system disk for BHAGAT.
5. MAKE a directory structure for a new root on a system disk.
6. DELETE a root from a system disk.
7. EXIT from this procedure.

Enter choice [7]: 3

CHANGE Menu

1. Enable BHAGAT as a boot server.
2. Disable BHAGAT as a boot server.
3. Enable a quorum disk for BHAGAT.
4. Disable a quorum disk for BHAGAT.
5. Enable BHAGAT as a disk server.
6. Disable BHAGAT as a disk server.
7. Change BHAGAT's ALLOCLASS value.
8. Enable BHAGAT as a tape server.
9. Disable BHAGAT as a tape server.
10. Change BHAGAT's TAPE_ALLOCLASS value.
11. Change an IA64 satellite node's LAN adapter hardware address.
12. Enable Cluster Communication using IP on BHAGAT.
13. Disable Cluster Communication using IP on BHAGAT.
14. Change BHAGAT's shared SCSI port allocation class value.
15. Reset an IA64 satellite node's boot environment file protections.
16. Return to MAIN menu.

Enter choice [16]: 10

Enter a value for BHAGAT's TAPE_ALLOCLASS parameter [0]: 1
```

If you have changed BHAGAT's TAPE_ALLOCLASS value, you must reconfigure the cluster, using the procedure described in the "OpenVMS Cluster Systems" Manual. This includes running AUTOGEN for BHAGAT as shown below, before rebooting the cluster:

```
$ @SYS$UPDATE:AUTOGEN GETDATA REBOOT
```

If you have changed BHAGAT's TAPE_ALLOCLASS value, you must reconfigure the cluster, using the procedure described in the OpenVMS Cluster Systems manual.

Example 8.13. Sample Interactive CLUSTER_CONFIG_LAN.COM Session to Convert a Standalone Computer to a Cluster Boot Server

```
$ @CLUSTER_CONFIG_LAN.COM
IA64 platform support is in procedure CLUSTER_CONFIG_LAN.COM.
The currently running procedure, CLUSTER_CONFIG.COM, will call
it for you.
Cluster/IPCI Configuration Procedure
CLUSTER_CONFIG_LAN Version V2.84
Executing on an IA64 System

DECnet-Plus is installed on this node.
IA64 satellites will use TCP/IP BOOTP and TFTP services for downline
loading.
TCP/IP is installed and running on this node.

Enter a "?" for help at any prompt. If you are familiar with
the execution of this procedure, you may want to mute extra notes
and explanations by invoking it with "@CLUSTER_CONFIG_LAN BRIEF".

This IA64 node is not currently a cluster member.

MAIN Menu

1. ADD MOON to existing cluster, or form a new cluster.
2. MAKE a directory structure for a new root on a system disk.
3. DELETE a root from a system disk.
4. EXIT from this procedure.

Enter choice [4]: 1
Is the node to be a clustered node with a shared SCSI/FIBRE-CHANNEL bus (Y/
N)? N

What is the node's SCS node name? moon

DECnet is running on this node. Even though you are configuring a LAN-
based cluster, the DECnet database will provide some information and
may be updated.

Do you want to define a DECnet synonym [Y]? N
IA64 node, using LAN for cluster communications. PEDRIVER will be
loaded.
No other cluster interconnects are supported for IA64 nodes.
Enter this cluster's group number: 123
```

Enter this cluster's password:

Re-enter this cluster's password for verification:

Will MOON be a boot server [Y]? [Return]

TCP/IP BOOTP and TFTP services must be enabled on IA64 boot nodes.

Use SYS\$MANAGER:TCPIP\$CONFIG.COM on MOON to enable BOOTP and TFTP services after MOON has booted into the cluster.

Enter a value for MOON's ALLOCLASS parameter [0]:[Return]

Does this cluster contain a quorum disk [N]? [Return]

The EXPECTED_VOTES system parameter of members of a cluster indicates the total number of votes present when all cluster members are booted, and is used to determine the minimum number of votes (QUORUM) needed for cluster operation.

EXPECTED_VOTES value for this cluster: 1

Warning: Setting EXPECTED_VOTES to 1 allows this node to boot without being able to see any other nodes in the cluster. If there is another instance of the cluster in existence that is unreachable via SCS but shares common drives (such as a Fibrechannel fabric) this may result in severe disk corruption.

Do you wish to re-enter the value of EXPECTED_VOTES [Y]? N

The use of a quorum disk is recommended for small clusters to maintain cluster quorum if cluster availability with only a single cluster node is a requirement.

For complete instructions, check the section on configuring a cluster in the "OpenVMS Cluster Systems" manual.

WARNING: MOON will be a voting cluster member. EXPECTED_VOTES for this and every other cluster member should be adjusted at a convenient time before a reboot. For complete instructions, check the section on configuring a cluster in the "OpenVMS Cluster Systems" manual.

Execute AUTOGEN to compute the SYSGEN parameters for your configuration and reboot MOON with the new parameters. This is necessary before MOON can become a cluster member.

Do you want to run AUTOGEN now [Y]? [Return]

Running AUTOGEN -- Please wait.

%AUTOGEN-I-BEGIN, GETDATA phase is beginning.

.
.

8.5. Creating a Duplicate System Disk

As you continue to add Integrity servers running on a common Integrity common system disk, or Alpha computers running on an Alpha common system disk, you eventually reach the disk's storage or I/O capacity. In that case, you want to add one or more common system disks to handle the increased load.

Reminder: Remember that a system disk cannot be shared between two architectures. Furthermore, you cannot create a system disk for one architecture from a system disk of a different architecture.

8.5.1. Preparation

You can use either `CLUSTER_CONFIG_LAN.COM` or `CLUSTER_CONFIG.COM` to set up additional system disks. *After* you have coordinated cluster common files as described in Chapter 5, proceed as follows:

1. Locate an appropriate scratch disk for use as an additional system disk.
2. Log in as system manager.
3. Invoke either `CLUSTER_CONFIG_LAN.COM` or `CLUSTER_CONFIG.COM` and select the `CREATE` option.

8.5.2. Example

As shown in Example 8.14, the cluster configuration command procedure:

1. Prompts for the device names of the current and new system disks.
2. Backs up the current system disk to the new one.
3. Deletes all directory roots (except `SYS0`) from the new disk.
4. Mounts the new disk clusterwide.

Note: OpenVMS RMS error messages are displayed while the procedure deletes directory files. You can ignore these messages.

Example 8.14. Sample Interactive `CLUSTER_CONFIG_LAN.COM CREATE` Session

```
$ @CLUSTER_CONFIG_LAN.COM
      Cluster/IPCI Configuration Procedure
      CLUSTER_CONFIG_LAN Version V2.84
      Executing on an IA64 System

DECnet-Plus is installed on this node.
IA64 satellites will use TCP/IP BOOTP and TFTP services for downline
loading.
TCP/IP is installed and running on this node.

Enter a "?" for help at any prompt.  If you are familiar with
the execution of this procedure, you may want to mute extra notes
and explanations by invoking it with "@CLUSTER_CONFIG_LAN BRIEF".
```

BHAGAT is an IA64 system and currently a member of a cluster so the following functions can be performed:

MAIN Menu

1. ADD an IA64 node to the cluster.
2. REMOVE a node from the cluster.
3. CHANGE a cluster member's characteristics.
4. CREATE a duplicate system disk for BHAGAT.
5. MAKE a directory structure for a new root on a system disk.
6. DELETE a root from a system disk.
7. EXIT from this procedure.

Enter choice [7]: 4

The CREATE function generates a duplicate system disk.

- o It backs up the current system disk to the new system disk.
- o It then removes from the new system disk all system roots.

WARNING: Do not proceed unless you have defined appropriate logical names for cluster common files in SYLOGICALS.COM. For instructions, refer to the "OpenVMS Cluster Systems" manual.

Do you want to continue [N]? Y

This procedure will now ask you for the device name of the current system disk. The default device name (DISK\$BHAGAT_SYS:) is the logical volume name of SYS\$SYSDEVICE:.

What is the device name of the current system disk [DISK\$BHAGAT_SYS:]?
What is the device name of the new system disk?

.
.
.

8.6. Postconfiguration Tasks

Some configuration functions, such as adding or removing a voting member or enabling or disabling a quorum disk, require one or more additional operations.

These operations are listed in Table 8.10 and affect the integrity of the entire cluster. Follow the instructions in the table for the action you should take after executing either CLUSTER_CONFIG_LAN.COM or CLUSTER_CONFIG.COM to make major configuration changes.

Table 8.10. Actions Required to Reconfigure a Cluster

After running the cluster configuration procedure to...	You should...
Add or remove a voting member	Update the AUTOGEN parameter files and the current system parameter files for all nodes in the cluster, as described in Section 8.6.1.
Enable a quorum disk	Perform the following steps:

After running the cluster configuration procedure to...	You should...							
	<ol style="list-style-type: none"> 1. Update the AUTOGEN parameter files and the current system parameter files for all quorum watchers in the cluster, as described in Section 8.6.1. 2. Reboot the nodes that have been enabled as quorum disk watchers (Section 2.3.9). <p>Reference: See also Section 8.2.4 for more information about adding a quorum disk.</p>							
Disable a quorum disk	<p>Perform the following steps:</p> <p>Caution: Do not perform these steps until you are ready to reboot the entire OpenVMS Cluster system. Because you are reducing quorum for the cluster, the votes cast by the quorum disk being removed could cause cluster partitioning.</p> <ol style="list-style-type: none"> 1. Update the AUTOGEN parameter files and the current system parameter files for all quorum watchers in the cluster, as described in Section 8.6.1. 2. Evaluate whether or not quorum will be lost without the quorum disk: <table border="1" data-bbox="545 1034 1447 1482"> <thead> <tr> <th data-bbox="545 1034 842 1115">Quorum will not be lost</th> <th data-bbox="849 1034 1447 1115">THEN...</th> </tr> </thead> <tbody> <tr> <td data-bbox="545 1115 842 1438">Quorum will not be lost</td> <td data-bbox="849 1115 1447 1438"> <p>Perform these steps:</p> <ol style="list-style-type: none"> 1. Use the DCL command SET CLUSTER/EXPECTED_VOTES to reduce the value of quorum. 2. Reboot the nodes that have been disabled as quorum disk watchers. (Quorum disk watchers are described in Section 2.3.9.) </td> </tr> <tr> <td data-bbox="545 1438 842 1482">Quorum will be lost</td> <td data-bbox="849 1438 1447 1482">Shut down and reboot the entire cluster.</td> </tr> </tbody> </table> <p>Reference: Cluster shutdown is described in Section 8.6.2.</p> <p>Reference: See also Section 8.3.2 for more information about removing a quorum disk.</p>		Quorum will not be lost	THEN...	Quorum will not be lost	<p>Perform these steps:</p> <ol style="list-style-type: none"> 1. Use the DCL command SET CLUSTER/EXPECTED_VOTES to reduce the value of quorum. 2. Reboot the nodes that have been disabled as quorum disk watchers. (Quorum disk watchers are described in Section 2.3.9.) 	Quorum will be lost	Shut down and reboot the entire cluster.
Quorum will not be lost	THEN...							
Quorum will not be lost	<p>Perform these steps:</p> <ol style="list-style-type: none"> 1. Use the DCL command SET CLUSTER/EXPECTED_VOTES to reduce the value of quorum. 2. Reboot the nodes that have been disabled as quorum disk watchers. (Quorum disk watchers are described in Section 2.3.9.) 							
Quorum will be lost	Shut down and reboot the entire cluster.							
Add a satellite node	<p>Perform these steps:</p> <ul style="list-style-type: none"> • Update the volatile network databases on other cluster members (Section 8.6.4). • Optionally, alter the satellite's local disk label (Section 8.6.5). 							
Enable or disable the LAN or IP for cluster communications	Update the current system parameter files and reboot the node on which you have enabled or disabled the LAN or IP (Section 8.6.1).							
Change allocation class values	Update the current system parameter files and shut down and reboot the entire cluster (Sections 8.6.1 and 8.6.2).							

After running the cluster configuration procedure to...	You should...
Change the cluster group number or password	Shut down and reboot the entire cluster (Sections 8.6.2 and 8.6.7).

8.6.1. Updating Parameter Files

The cluster configuration command procedures (CLUSTER_CONFIG_LAN.COM or CLUSTER_CONFIG.COM) can be used to modify parameters in the AUTOGEN parameter file for the node on which it is run.

In some cases, such as when you add or remove a voting cluster member, or when you enable or disable a quorum disk, you must update the AUTOGEN files for all the other cluster members.

Use either of the methods described in the following table.

Method	Description
Update MODPARAMS.DAT files	<p>Edit MODPARAMS.DAT in all cluster members' [SYS x.SYSEXE] directories and adjust the value for the EXPECTED_VOTES system parameter appropriately.</p> <p>For example, if you add a voting member or if you enable a quorum disk, you must increment the value by the number of votes assigned to the new member (usually 1). If you add a voting member with one vote and enable a quorum disk with one vote on that computer, you must increment the value by 2.</p>
Update AGEN\$ files	<p>Update the parameter settings in the appropriate AGEN\$ include files:</p> <ul style="list-style-type: none"> • For satellites, edit SY\$MANAGER:AGEN\$NEW_SATELLITE_DEFAULTS.DAT. • For nonsatellites, edit SY\$MANAGER:AGEN\$NEW_NODE_DEFAULTS.DAT. <p>Reference: These files are described in Section 8.2.2.</p>

You must also update the current system parameter files (IA64VMSSYS.PAR or ALPHAVMSSYS.PAR, as appropriate) so that the changes take effect on the next reboot.

Use either of the methods described in the following table.

Method	Description
SYSMAN utility	<p>Perform the following steps:</p> <ol style="list-style-type: none"> 1. Log in as system manager. 2. Run the SYSMAN utility to update the EXPECTED_VOTES system parameter on all nodes in the cluster. For example: <pre>\$ RUN SYS\$SYSTEM:SYSMAN %SYSMAN-I-ENV, current command environment:</pre>

Method	Description
	<p>Clusterwide on local cluster Username SYSTEM will be used on nonlocal nodes</p> <pre> SYSMAN> SET ENVIRONMENT/CLUSTER SYSMAN> PARAM USE CURRENT SYSMAN> PARAM SET EXPECTED_VOTES 2 SYSMAN> PARAM WRITE CURRENT SYSMAN> EXIT </pre>
AUTOGEN utility	<p>Perform the following steps:</p> <ol style="list-style-type: none"> 1. Log in as system manager. 2. Run the AUTOGEN utility to update the EXPECTED_VOTES system parameter on all nodes in the cluster. For example: <pre> \$ RUN SYS\$SYSTEM:SYSMAN %SYSMAN-I-ENV, current command environment: Clusterwide on local cluster Username SYSTEM will be used on nonlocal nodes SYSMAN> SET ENVIRONMENT/CLUSTER SYSMAN> DO @SYS\$UPDATE:AUTOGEN GETDATA SETPARAMS SYSMAN> EXIT </pre> <p>Do <i>not</i> specify the SHUTDOWN or REBOOT option.</p> <p>Hints: If your next action is to shut down the node, you can specify SHUTDOWN or REBOOT (in place of SETPARAMS) in the DO @SYS\$UPDATE:AUTOGEN GETDATA command.</p>

Both of these methods propagate the values to the computer's ALPHAVMSSYS.PAR file on Alpha computers or to the IA64VMSSYS.PAR file on Integrity server systems. In order for these changes to take effect, continue with the instructions in either Section 8.6.2 to shut down the cluster or in Section 8.6.3 to shut down the node.

8.6.2. Shutting Down the Cluster

Using the SYSMAN utility, you can shut down the entire cluster from a single node in the cluster. Follow these steps to perform an orderly shutdown:

1. Log in to the system manager's account on any node in the cluster.
2. Run the SYSMAN utility and specify the SET ENVIRONMENT/CLUSTER command. Be sure to specify the /CLUSTER_SHUTDOWN qualifier to the SHUTDOWN NODE command. For example:

```

$ RUN SYS$SYSTEM:SYSMAN
SYSMAN> SET ENVIRONMENT/CLUSTER
%SYSMAN-I-ENV, current command environment:
Clusterwide on local cluster
Username SYSTEM will be used on nonlocal nodes

SYSMAN> SHUTDOWN NODE/CLUSTER_SHUTDOWN/MINUTES_TO_SHUTDOWN=5 -
SYSMAN> /AUTOMATIC_REBOOT/REASON="Cluster Reconfiguration"
%SYSMAN-I-SHUTDOWN, SHUTDOWN request sent to node

```

```
%SYSMAN-I-SHUTDOWN, SHUTDOWN request sent to node
SYSMAN>
```

```
SHUTDOWN message on BHAGAT from user SYSTEM at BHAGAT Batch 11:02:10
BHAGAT will shut down in 5 minutes; back up shortly via automatic reboot.
Please log off node BHAGAT.
```

```
Cluster Reconfiguration
```

```
SHUTDOWN message on BHAGAT from user SYSTEM at BHAGAT Batch 11:02:10
PLUTO will shut down in 5 minutes; back up shortly via automatic reboot.
Please log off node PLUTO.
```

```
Cluster Reconfiguration
```

For more information, see Section 10.6.

8.6.3. Shutting Down a Single Node

To stop a single node in an OpenVMS Cluster, you can use either the SYSMAN SHUTDOWN NODE command with the appropriate SETENVIRONMENT command or the SHUTDOWN command procedure. These methods are described in the following table.

Method	Description
SYSMAN utility	<p>Follow these steps:</p> <ol style="list-style-type: none"> 1. Log in to the system manager's account on any node in the OpenVMS Cluster. 2. Run the SYSMAN utility to shut down the node, as follows: <pre> \$ RUN SYS\$SYSTEM:SYSMAN SYSMAN> SET ENVIRONMENT/NODE=JUPITR Individual nodes: JUPITR Username SYSTEM will be used on nonlocal nodes SYSMAN> SHUTDOWN NODE/REASON="Maintenance" - _SYSMAN> /MINUTES_TO_SHUTDOWN=5 Hint: To shut down a subset of nodes in the cluster, you can enter several node names separated by commas on the SET ENVIRONMENT/NODE command. The following command shuts down nodes JUPITR and SATURN: <pre> SYSMAN> SET ENVIRONMENT/NODE=(JUPITR,SATURN) </pre> </pre>
SHUTDOWN command procedure	<p>Follow these steps:</p> <ol style="list-style-type: none"> 1. Log in to the system manager's account on the node to be shut down. 2. Invoke the SHUTDOWN command procedure as follows: <pre> \$ @SYS\$SYSTEM:SHUTDOWN </pre>

For more information, see Section 10.6.

8.6.4. Updating Network Data

Whenever you add a satellite, the cluster configuration command procedure you use (CLUSTER_CONFIG_LAN.COM or CLUSTER_CONFIG.COM) updates both the permanent and

volatile remote node network databases (NETNODE_REMOTE.DAT) on the boot server. However, the volatile databases on other cluster members are not automatically updated.

To share the new data throughout the cluster, you must update the volatile databases on all other cluster members. Log in as system manager, invoke the SYSMAN utility, and enter the following commands at the SYSMAN> prompt:

```
$ RUN SYS$SYSTEM:SYSMAN
SYSMAN> SET ENVIRONMENT/CLUSTER
%SYSMAN-I-ENV, current command environment:
      Clusterwide on local cluster
      Username SYSTEM          will be used on nonlocal nodes
SYSMAN> SET PROFILE/PRIVILEGES=(OPER,SYSPRV)
SYSMAN> DO MCR NCP SET KNOWN NODES ALL
%SYSMAN-I-OUTPUT, command execution on node X...
.
.
.
SYSMAN> EXIT
$
```

The file NETNODE_REMOTE.DAT must be located in the directory SYSS\$COMMON:[SYSEXE].

8.6.5. Altering Satellite Local Disk Labels

If you want to alter the volume label on a satellite node's local page and swap disk, follow these steps after the satellite has been added to the cluster:

Step	Action
1	Log in as system manager and enter a DCL command in the following format: SET VOLUME/LABEL=volume-label device-spec[:] Note: The SET VOLUME command requires write access (W) to the index file on the volume. If you are not the volume's owner, you must have either a system user identification code (UIC) or the SYSPRV privilege.
2	Update the [SYS n.SYSEXE]SATELLITE_PAGE.COM procedure on the boot server's system disk to reflect the new label.

8.6.6. Changing Allocation Class Values

If you must change allocation class values on any HSG or HSV subsystem, you must do so while the entire cluster is shut down.

Reference: To change allocation class values on computer systems, see Section 6.2.2.1.

8.6.7. Rebooting

The following table describes booting actions for satellite and storage subsystems:

For configurations with...	You must...
HSG and HSV subsystems	Reboot each computer after all HSG and HSV subsystems have been set and rebooted.

For configurations with...	You must...
Satellite nodes	<p>Reboot boot servers before rebooting satellites.</p> <p>Note that several new messages might appear. For example, if you have used the CLUSTER_CONFIG.COM CHANGE function to enable cluster communications over the LAN, one message reports that the LAN OpenVMS Cluster security database is being loaded.</p> <p>Reference: See also Section 9.3 for more information about booting satellites.</p>

For every disk-serving computer, a message reports that the MSCP server is being loaded.

To verify that all disks are being served in the manner in which you designed the configuration, at the system prompt (\$) of the node serving the disks, enter the SHOW DEVICE/SERVED command. For example, the following display represents a DSSI configuration:

```
$ SHOW DEVICE/SERVED
```

```
Device:  Status  Total Size  Current  Max  Hosts
$1$DIA0  Avail    1954050      0      0      0
$1$DIA2  Avail    1800020      0      0      0
```

Caution: If you boot a node into an existing OpenVMS Cluster using minimum startup (the system parameter STARTUP_P1 is set to MIN), a number of processes (for example, CACHE_SERVER, CLUSTER_SERVER, and CONFIGURE) are not started. VSI recommends that you start these processes manually if you intend to run the node in an OpenVMS Cluster system. Running a node without these processes enabled prevents the cluster from functioning properly.

Reference: Refer to the *VSI OpenVMS System Manager's Manual* for more information about starting these processes manually.

8.6.8. Rebooting Satellites Configured with OpenVMS on a Local Disk (Alpha only)

Satellite nodes can be set up to reboot automatically when recovering from system failures or power failures.

Reboot behavior varies from system to system. Many systems provide a console variable that allows you to specify which device to boot from by default. However, some systems have predefined boot “sniffers” that automatically detect a bootable device. The following table describes the rebooting conditions.

IF...	AND...	THEN...
If your system does not allow you to specify the boot device for automatic reboot (that is, it has a boot sniffer)	An operating system is installed on the system's local disk	That disk will be booted in preference to requesting a satellite MOP load. To avoid this, you should take one of the measures in the following list before allowing any operation that causes an automatic reboot—for example, executing SYSSYSTEM:SHUTDOWN.COM with the REBOOT option or using CLUSTER_CONFIG.COM to add that satellite to the cluster:

IF...	AND...	THEN...
		<ul style="list-style-type: none"> <li data-bbox="730 241 1430 450">• Rename the directory file <code>ddcu:[000000]SYS0.DIR</code> on the local disk to <code>ddcu:[000000]SYS x.DIR</code> (where <code>SYS x</code> is a root other than <code>SYS0</code>, <code>SYSE</code>, or <code>SYSF</code>). Then enter the DCL command <code>SET FILE/REMOVE</code> as follows to remove the old directory entry for the boot image <code>SYSBOOT.EXE</code>: <pre data-bbox="775 479 1350 600">\$ RENAME DUA0:[000000]SYS0.DIR DUA0: [000000]SYS1.DIR \$ SET FILE/REMOVE DUA0: [SYSEXE]SYSBOOT.EXE</pre> <li data-bbox="730 636 1390 768">• Disable the local disk. For instructions, refer to your computer-specific installation and operations guide. Note that this option is not available if the satellite's local disk is being used for paging and swapping.

8.7. Running AUTOGEN with Feedback

AUTOGEN includes a mechanism called **feedback**. This mechanism examines data collected during normal system operations, and it adjusts system parameters on the basis of the collected data whenever you run AUTOGEN with the feedback option. For example, the system records each instance of a disk server waiting for buffer space to process a disk request. Based on this information, AUTOGEN can size the disk server's buffer pool automatically to ensure that sufficient space is allocated.

Execute `SY$UPDATE:AUTOGEN.COM` manually as described in the *VSI OpenVMS System Manager's Manual*.

8.7.1. Advantages

To ensure that computers are configured adequately when they first join the cluster, you can run AUTOGEN with feedback automatically as part of the initial boot sequence. Although this step adds an additional reboot before the computer can be used, the computer's performance can be substantially improved.

VSI strongly recommends that you use the feedback option. Without feedback, it is difficult for AUTOGEN to anticipate patterns of resource usage, particularly in complex configurations. Factors such as the number of computers and disks in the cluster and the types of applications being run require adjustment of system parameters for optimal performance.

VSI also recommends using AUTOGEN with feedback rather than the `SYSGEN` utility to modify system parameters, because AUTOGEN:

- Uses parameter changes in `MODPARAMS.DAT` and `AGEN$` files. (Changes recorded in `MODPARAMS.DAT` are not lost during updates to the OpenVMS operating system).
- Reconfigures other system parameters to reflect changes.

8.7.2. Initial Values

When a computer is first added to an OpenVMS Cluster, system parameters that control the computer's system resources are normally adjusted in several steps, as follows:

1. The cluster configuration command procedure (`CLUSTER_CONFIG_LAN.COM` or `CLUSTER_CONFIG.COM`) sets initial parameters that are adequate to boot the computer in a minimum environment.
2. When the computer boots, `AUTOGEN` runs automatically to size the static operating system (without using any dynamic feedback data), and the computer reboots into the OpenVMS Cluster environment.
3. After the newly added computer has been subjected to typical use for a day or more, you should run `AUTOGEN` with feedback manually to adjust parameters for the OpenVMS Cluster environment.
4. At regular intervals, and whenever a major change occurs in the cluster configuration or production environment, you should run `AUTOGEN` with feedback manually to readjust parameters for the changes.

Because the first `AUTOGEN` operation (initiated by either `CLUSTER_CONFIG_LAN.COM` or `CLUSTER_CONFIG.COM`) is performed both in the minimum environment and without feedback, a newly added computer may be inadequately configured to run in the OpenVMS Cluster environment. For this reason, you might want to implement additional configuration measures like those described in Section 8.7.3 and Section 8.7.4.

8.7.3. Obtaining Reasonable Feedback

When a computer first boots into an OpenVMS Cluster, much of the computer's resource utilization is determined by the current OpenVMS Cluster configuration. Factors such as the number of computers, the number of disk servers, and the number of disks available or mounted contribute to a fixed minimum resource requirements. Because this minimum does not change with continued use of the computer, feedback information about the required resources is immediately valid.

Other feedback information, however, such as that influenced by normal user activity, is not immediately available, because the only “user” has been the system startup process. If `AUTOGEN` were run with feedback at this point, some system values might be set too low.

By running a simulated user load at the end of the first production boot, you can ensure that `AUTOGEN` has reasonable feedback information. The User Environment Test Package (UETP) supplied with your operating system contains a test that simulates such a load. You can run this test (the UETP LOAD phase) as part of the initial production boot, and then run `AUTOGEN` with feedback before a user is allowed to log in.

To implement this technique, you can create a command file like that in step1 of the procedure in Section 8.7.4, and submit the file to the computer's local batch queue from the cluster common `SYSTARTUP` procedure. Your command file conditionally runs the UETP LOAD phase and then reboots the computer with `AUTOGEN` feedback.

8.7.4. Creating a Command File to Run AUTOGEN

As shown in the following sample file, UETP lets you specify a typical user load to be run on the computer when it first joins the cluster. The UETP run generates data that `AUTOGEN` uses to set appropriate system parameter values for the computer when rebooting it with feedback. Note, however, that the default setting for the UETP user load assumes that the computer is used as a timesharing system. This calculation can produce system parameter values that might be excessive for a single-user workstation, especially if the workstation has large memory resources. Therefore, you might want to modify the default user load setting, as shown in the sample file.

Follow these steps:

1. Create a command file like the following:

```

$!
$! ***** SYS$COMMON:[SYSMGR]UETP_AUTOGEN.COM *****
$!
$! For initial boot only, run UETP LOAD phase and
$! reboot with AUTOGEN feedback.
$!
$ SET NOON
$ SET PROCESS/PRIVILEGES=ALL
$!
$! Run UETP to simulate a user load for a satellite
$! with 8 simultaneously active user processes. For a
$! CI connected computer, allow UETP to calculate the load.
$!
$ LOADS = "8"
$ IF F$GETDVI("PAA0:", "EXISTS") THEN LOADS = ""
$ @UETP LOAD 1 'loads'
$!
$! Create a marker file to prevent resubmission of
$! UETP_AUTOGEN.COM at subsequent reboots.
$!
$ CREATE SYS$SPECIFIC:[SYSMGR]UETP_AUTOGEN.DONE
$!
$! Reboot with AUTOGEN to set SYSGEN values.
$!
$ @SYS$UPDATE:AUTOGEN SAVPARAMS REBOOT FEEDBACK
$!
$ EXIT

```

2. Edit the cluster common SYSTARTUP file and add the following commands at the end of the file. Assume that queues have been started and that a batch queue is running on the newly added computer. Submit UETP_AUTOGEN.COM to the computer's local batch queue.

```

$!
$ NODE = F$GETSYI("NODE")
$ IF F$SEARCH ("SYS$SPECIFIC:[SYSMGR]UETP_AUTOGEN.DONE") .EQS. ""
$ THEN
$ SUBMIT /NOPRINT /NOTIFY /USERNAME=SYSTEST -
_ $ /QUEUE='NODE'_BATCH SYS$MANAGER:UETP_AUTOGEN

$ WAIT_FOR_UETP:
$ WRITE SYS$OUTPUT "Waiting for UETP and AUTOGEN... 'F$TIME()'"
$ WAIT 00:05:00.00 ! Wait 5 minutes
$ GOTO WAIT_FOR_UETP
$ ENDIF
$!

```

Note: UETP must be run under the user name SYSTEST.

3. Execute CLUSTER_CONFIG_LAN.COM or CLUSTER_CONFIG.COM to add the computer.

When you boot the computer, it runs UETP_AUTOGEN.COM to simulate the user load you have specified, and it then reboots with AUTOGEN feedback to set appropriate system parameter values.

Chapter 9. Building Large OpenVMS Cluster Systems

This chapter provides guidelines for building OpenVMS Cluster systems that include many computers—approximately 20 or more—and describes procedures that you might find helpful. (Refer to the OpenVMS Cluster Software *Software Product Description* (SPD) for configuration limitations.) Typically, such OpenVMS Cluster systems include a large number of satellites.

Note that the recommendations in this chapter also can prove beneficial in some clusters with fewer than 20 computers. Areas of discussion include:

- Booting
- Availability of MOP and disk servers
- Multiple system disks
- Shared resource availability
- Hot system files
- System disk space
- System parameters
- Network problems
- Cluster alias

9.1. Setting Up the Cluster

When building a new large cluster, you must be prepared to run AUTOGEN and reboot the cluster several times during the installation. The parameters that AUTOGEN sets for the first computers added to the cluster will probably be inadequate when additional computers are added. Readjustment of parameters is critical for boot and disk servers.

One solution to this problem is to run the UETP_AUTOGEN.COM command procedure (described in Section 8.7.4) to reboot computers at regular intervals as new computers or storage interconnects are added. For example, each time there is a 10% increase in the number of computers, storage, or interconnects, you should run UETP_AUTOGEN.COM. For best results, the last time you run the procedure should be as close as possible to the final OpenVMS Cluster environment.

To set up a new, large OpenVMS Cluster, follow these steps:

Step	Task
1	Configure boot and disk servers using the CLUSTER_CONFIG_LAN.COM or the CLUSTER_CONFIG.COM command procedure (described in Chapter 8).
2	Install all layered products and site-specific applications required for the OpenVMS Cluster environment, or as many as possible.

Step	Task
3	Prepare the cluster startup procedures so that they are as close as possible to those that will be used in the final OpenVMS Cluster environment.
4	Add a small number of satellites (perhaps two or three) using the cluster configuration command procedure.
5	Reboot the cluster to verify that the startup procedures work as expected.
6	After you have verified that startup procedures work, run UETP_AUTOGEN.COM on every computer's local batch queue to reboot the cluster again and to set initial production environment values. When the cluster has rebooted, all computers should have reasonable parameter settings. However, check the settings to be sure.
7	Add additional satellites to double their number. Then rerun UETP_AUTOGEN on each computer's local batch queue to reboot the cluster, and set values appropriately to accommodate the newly added satellites.
8	Repeat the previous step until all satellites have been added.
9	When all satellites have been added, run UETP_AUTOGEN a final time on each computer's local batch queue to reboot the cluster and to set new values for the production environment.

For best performance, do not run UETP_AUTOGEN on every computer simultaneously, because the procedure simulates a user load that is probably more demanding than that for the final production environment. A better method is to run UETP_AUTOGEN on several satellites (those with the least recently adjusted parameters) while adding new computers. This technique increases efficiency because little is gained when a satellite reruns AUTOGEN shortly after joining the cluster.

For example, if the entire cluster is rebooted after 30 satellites have been added, few adjustments are made to system parameter values for the 28th satellite added, because only two satellites have joined the cluster since that satellite ran UETP_AUTOGEN as part of its initial configuration.

9.2. General Booting Considerations

Two general booting considerations, concurrent booting and minimizing boot time, are described in this section.

9.2.1. Concurrent Booting

Concurrent booting occurs after a power or a site failure when all the nodes are rebooted simultaneously. This results insignificant I/O load on the interconnects. Also, results in network activity due to SCS traffic required for synchronizing. All satellites wait to reload operating system. As soon as the boot server is available, they begin to boot in parallel resulting in elapsed time during login.

9.2.2. Minimizing Boot Time

A large cluster needs to be carefully configured so that there is sufficient capacity to boot the desired number of nodes in the desired amount of time. The effect of 96 satellites rebooting could induce an I/O bottleneck that can stretch the OpenVMS Cluster reboot times into hours. The following list provides a few methods to minimize boot times.

- Careful configuration techniques

Guidelines for OpenVMS Cluster Configurations contains data on configurations and the capacity of the computers, system disks, and interconnects involved.

- Adequate system disk throughput

Achieving enough system disk throughput typically requires a combination of techniques. Refer to Section 9.7 for complete information.

- Sufficient network bandwidth

A single Gigabit Ethernet is unlikely to have sufficient bandwidth to meet the needs of a large OpenVMS cluster. Likewise, a single Gigabit Ethernet adapter may become a bottleneck, especially for a disk server during heavy application synchronizing. This results in high SCS traffic. Having more adapters for SCS helps in overcoming such bandwidth limitation.

Sufficient network bandwidth can also be provided using some of the techniques listed in step 1 of Table 9.2.

- Installation of only the required layered products and devices.

9.2.3. General Booting Considerations for Cluster over IP

OpenVMS clusters can use TCP/IP stack for communicating with other nodes in the cluster and passing SCS traffic. To be able to use TCP/IP for cluster communication a node has to be configured. For details on how to configure a node to use OpenVMS Cluster over IP, see Section 8.2.3.1. After enabling this feature, load TCP/IP stack early in the boot time during load. OpenVMS executive has been modified to load TCP/IP execlets early in the boot time so that the node can exchange SCS messages with other existing nodes of the cluster. This feature also uses configuration files which get loaded during boot time. It is necessary to ensure that these configuration files are correctly generated during the configuration. The following are some of considerations for booting.

- Ensure that the node has TCP/IP connectivity with other nodes of the cluster.
- Ensure that the IP multicast address used for cluster is able to be passed between the routers.
- If IP unicast is used, ensure that the nodes' IP address is present in all the existing nodes in the PE \$IP_CONFIG.DAT file. (MC SCACPRELOAD command can be used to load new IP address).

9.3. Booting Satellites

OpenVMS Cluster satellite nodes use a single LAN adapter for the initial stages of booting. If a satellite is configured with multiple LAN adapters, the system manager can specify with the console BOOT command which adapter to use for the initial stages of booting. Once the system is running, the OpenVMS Cluster uses all available LAN adapters. This flexibility allows you to work around broken adapters or network problems.

For Alpha and Integrity cluster satellites, the network boot device cannot be a prospective member of a LAN Failover Set. For example, if you create a LAN Failover Set, LLA consisting of EWA and EWB, to be active when the system boots, you cannot boot the system as a satellite over the LAN devices EWA or EWB.

The procedures and utilities for configuring and booting satellite nodes vary between Integrity servers and Alpha systems.

9.3.1. Differences between Alpha and Integrity server Satellites

Table 9.1 lists the differences between Alpha and Integrity server satellites.

Table 9.1. Differences Between Alpha and Integrity server Satellites

	Alpha	Integrity servers
Boot Protocol	MOP	PXE (BOOTP /DHCP /TFTP)
Crash Dumps	May crash to remote system disk or to local disk via Dump Off the System Disk (DOSD)	Requires DOSD. Crashing to the remote disk is not possible.
Error Log Buffers	Always written to the remote system disk	Error log buffers are written to the same disk as DOSD
File protections	No different than standard system disk	Requires that all loadable execlets are W:RE (the default case) and that certain files have ACL access via the VMS \$SATELLITE_ACCESS identifier

9.4. Configuring and Booting Satellite Nodes (Alpha)

Complete the items in the following Table 9.2 before proceeding with satellite booting.

Table 9.2. Checklist for Satellite Booting

Step	Action
1	<p>Configure disk server LAN adapters.</p> <p>Because disk-serving activity in an OpenVMS Cluster system can generate a substantial amount of I/O traffic on the LAN, boot and disk servers should use the highest-bandwidth LAN adapters in the cluster. The servers can also use multiple LAN adapters in a single system to distribute the load across the LAN adapters.</p> <p>The following list suggests ways to provide sufficient network bandwidth:</p> <ul style="list-style-type: none"> • Select network adapters with sufficient bandwidth. • Use switches to segregate traffic and to provide increased total bandwidth. • Use multiple LAN adapters on MOP and disk servers. • Use switches or higher speed LANs, fanning out to slower LAN segments. • Use multiple independent networks. • Provide sufficient MOP and disk server CPU capacity by selecting a computer with sufficient power and by configuring multiple server nodes to share the load.

Step	Action
2	If the MOP server node and system-disk server node are not already configured as cluster members, follow the directions in Section 8.4 for using the cluster configuration command procedure to configure each of the Alpha nodes. Include multiple boot and disk servers to enhance availability and distribute I/O traffic over several cluster nodes.
3	Configure additional memory for disk serving.
4	Run the cluster configuration procedure on the Alpha node for each satellite you want to boot into the OpenVMS Cluster.

9.4.1. Booting from a Single LAN Adapter

To boot a satellite, enter the following command:

```
>>> BOOT LAN-adapter-device-name
```

In the example, the *LAN-adapter-device-name* could be any valid LAN adapter name, for example EZA0 or XQB0.

If you need to perform a conversational boot, use the command shown. At the Alpha system console prompt (>>>), enter:

```
>>> b -flags 0,1 eza0
```

In this example, *-flags* stands for the flags command line qualifier, which takes two values:

- System root number

The “0” tells the console to boot from the system root [SYS0]. This is ignored when booting satellite nodes because the system root comes from the network database of the boot node.

- Conversational boot flag

The “1” indicates that the boot should be conversational.

The argument *eza0* is the LAN adapter to be used for booting.

Finally, notice that a load file is not specified in this boot command line. For satellite booting, the load file is part of the node description in the DECnet or LANCP database.

If the boot fails:

- If the configuration permits and the network database is properly set up, reenter the boot command using another LAN adapter (see Section 9.4.4).
- See Section C.2.5 for information about troubleshooting satellite booting problems.

9.4.2. Changing the Default Boot Adapter

To change the default boot adapter, you need the physical address of the alternate LAN adapter. You use the address to update the satellite's node definition in the DECnet or LANCP database on the MOP servers so that they recognize the satellite (described in Section 9.4.4). Use the SHOW CONFIG command to find the LAN address of additional adapters.

9.4.3. Booting from Multiple LAN Adapters (Alpha Only)

On Alpha systems, availability can be increased by using multiple LAN adapters for booting because access to the MOP server and disk server can occur via different LAN adapters. To use multiple adapter booting, perform the steps in the following table.

Step	Task
1	Obtain the physical addresses of the additional LAN adapters.
2	Use these addresses to update the node definition in the DECnet or LANCP database on some of the MOP servers so that they recognize the satellite (described in Section 9.4.4).
3	If the satellite is already defined in the DECnet database, skip to step 4. If the satellite is not defined in the DECnet database, specify the SYSSYSTEM:APB.EXE downline load file in the Alpha network database.
4	Specify multiple LAN adapters on the boot command line. (Use the SHOW DEVICE or SHOW CONFIG console command to obtain the names of adapters).

The following command line is the same as that used for booting from a single LAN adapter on an Alpha system (see Section 9.4.2) except that it lists two LAN adapters, eza0 and ezb0, as the devices from which to boot:

```
>>> b -flags 0,1 eza0, ezb0
```

In this command line:

Stage	What Happens
1	MOP booting is attempted from the first device (eza0). If that fails, MOP booting is attempted from the next device (ezb0). When booting from network devices, if the MOP boot attempt fails from all devices, then the console starts again from the first device.
2	Once the MOP load has completed, the boot driver starts the NISCA protocol on all of the LAN adapters. The NISCA protocol is used to access the system disk server and finish loading the operating system (see Appendix F).

9.4.4. Enabling Satellites to Use Alternate LAN Adapters for Booting

OpenVMS supports only one hardware address attribute per remote node definition in either a DECnet or LANCP database. To enable a satellite with multiple LAN adapters to use any LAN adapter to boot into the cluster, two different methods are available:

- Define a pseudonode for each additional LAN adapter.
- Create and maintain different node databases for different boot nodes.

Defining Pseudonodes for Additional LAN Adapters

When defining a pseudonode with a different DECnet or LANCP address:

- Make sure the address points to the same cluster satellite root directory as the existing node definition (to associate the pseudonode with the satellite).
- Specify the hardware address of the alternate LAN adapter in the pseudonode definition.

For DECnet, follow the procedure shown in Table 9.3. For LANCP, follow the procedure shown in Table 9.4.

Table 9.3. Procedure for Defining a Pseudonode Using DECnet MOP Services

Step	Procedure	Comments
1	Display the node's existing definition using the following NCP command: <pre>\$ RUN SYS\$SYSTEM:NCP NCP> SHOW NODE <i>node-name</i> CHARACTERISTICS</pre>	This command displays a list of the satellite's characteristics, such as its hardware address, load assist agent, load assist parameter, and more.
2	Create a pseudonode by defining a unique DECnet address and node name at the NCP command prompt, as follows: <pre>DEFINE NODE <i>pseudo-area.pseudo-number</i> - NAME <i>pseudo-node-name</i> - LOAD FILE APB.EXE - LOAD ASSIST AGENT SYS \$SHARE:NISCS LAA.EXE - LOAD ASSIST PARAMETER <i>disk\$sys:[<root.>]</i> - HARDWARE ADDRESS <i>xx-xx-xx-xx-xx-xx</i></pre>	This example is specific to an Alpha node.

Table 9.4. Procedure for Defining a Pseudonode Using LANCP MOP Services

Step	Procedure	Comments
1	Display the node's existing definition using the following LANCP command: <pre>\$ RUN SYS\$SYSTEM:LANCP LANCP> SHOW NODE <i>node-name</i></pre>	This command displays a list of the satellite's characteristics, such as its hardware address and root directory address.
2	Create a pseudonode by defining a unique LANCP address and node name at the LANCP command prompt, as follows: <pre>DEFINE NODE <i>pseudo-node-name</i> - /FILE= APB.EXE - /ROOT=<i>disk\$sys:[<root.>]</i> - /ADDRESS=<i>xx-xx-xx-xx-xx-xx</i></pre>	This example is specific to an Alpha node.

Creating Different Node Databases for Different Boot Nodes

When creating different DECnet or LANCP databases on different boot nodes:

- Set up the databases so that a system booting from one LAN adapter receives responses from a subset of the MOP servers. The same system booting from a different LAN adapter receives responses from a different subset of the MOP servers.

- In each database, list a different LAN address for the same node definition.

The procedures are similar for DECnet and LANCP, but the database file names, utilities, and commands differ. For the DECnet procedure, see Table 9.5. For the LANCP procedure, see Table 9.6.

Table 9.5. Procedure for Creating Different DECnet Node Databases

Step	Procedure	Comments
1	Define the logical name NETNODE_REMOTE to different values on different nodes so that it points to different files.	The logical NETNODE_REMOTE points to the working copy of the remote node file you are creating.
2	<p>Locate NETNODE_REMOTE.DAT files in the system-specific area for each node.</p> <p>On each of the various boot servers, ensure that the hardware address is defined as a unique address that matches one of the adapters on the satellite. Enter the following commands at the NCP command prompt:</p> <pre> DEFINE NODE <i>area.number</i> - NAME <i>node-name</i> - LOAD FILE APB.EXE - LOAD ASSIST AGENT SYS \$SHARE:NISCS_LAA.EXE - LOAD ASSIST PARAMETER <i>disk\$sys:[<root.>]</i> - HARDWARE ADDRESS <i>xx-xx-xx-xx-xx-xx</i> </pre>	<p>A NETNODE_REMOTE.DAT file located in [SYS0.SYSEX] overrides one located in [SYS0.SYSCOMMON.SYSEX] for a system booting from system root 0.</p> <p>If the NETNODE_REMOTE.DAT files are copies of each other, the node name, LOAD FILE, load assist agent, and load assist parameter are already set up. You need only specify the new hardware address.</p> <p>Because the default hardware address is stored in NETUPDATE.COM, you must also edit this file on the second boot server.</p>

Table 9.6. Procedure for Creating Different LANCP Node Databases

Step	Procedure	Comments
1	Define the logical name LAN\$NODE_DATABASE to different values on different nodes so that it points to different files.	The logical LAN \$NODE_DATABASE points to the working copy of the remote node file you are creating.
2	<p>Locate LAN\$NODE_DATABASE.DAT files in the system-specific area for each node.</p> <p>On each of the various boot servers, ensure that the hardware address is defined as a unique address that matches one of the adapters on the satellite. Enter the following commands at the LANCP command prompt:</p>	If the LAN \$NODE_DATABASE.DAT files are copies of each other, the node name and the FILE and ROOT qualifier values are already set up. You need only specify the new address.

Step	Procedure	Comments
	<pre> DEFINE NODE <i>node-name</i> - /FILE= APB.EXE - /ROOT=<i>disk\$sys:[<root.>]</i> - /ADDRESS=<i>xx-xx-xx-xx-xx-xx</i> </pre>	

Once the satellite receives the MOP downline load from the MOP server, the satellite uses the booting LAN adapter to connect to any node serving the system disk. The satellite continues to use the LAN adapters on the boot command line exclusively until after the run-time drivers are loaded. The satellite then switches to using the run-time drivers and starts the local area OpenVMS Cluster protocol on all of the LAN adapters.

For additional information about the NCP command syntax, refer to *VSI OpenVMS DECnet Network Management Utilities*.

For DECnet-Plus: On an OpenVMS Cluster running DECnet-Plus, you do not need to take the same actions in order to support a satellite with more than one LAN adapter. The DECnet-Plus support to downline load a satellite allows for an entry in the database that contains a list of LAN adapter addresses. See the DECnet-Plus documentation for complete information.

9.4.5. Configuring MOP Service

On a boot node, CLUSTER_CONFIG.COM enables the DECnet MOP downline load service on the first circuit that is found in the DECnet database.

On systems running DECnet for OpenVMS, display the circuit state and the service (MOP downline load service) state using the following command:

```

$ MCR NCP SHOW CHAR KNOWN CIRCUITS

      .
      .
      .
Circuit = SVA-0

State           = on
Service         = enabled
      .
      .
      .

```

This example shows that circuit SVA-0 is in the ON state with the MOP downline service enabled. This is the correct state to support MOP downline loading for satellites.

Enabling MOP service on additional LAN adapters (circuits) must be performed manually. For example, enter the following NCP commands to enable service for the circuit QNA-1:

```

$ MCR NCP SET CIRCUIT QNA-1 STATE OFF
$ MCR NCP SET CIRCUIT QNA-1 SERVICE ENABLED STATE ON
$ MCR NCP DEFINE CIRCUIT QNA-1 SERVICE ENABLED

```

Reference: For more details, refer to *DECnet-Plus for OpenVMS Network Management*.

9.4.6. Controlling Satellite Booting

You can control the satellite boot process in a number of ways. Table 9.7 shows examples specific to DECnet for OpenVMS. Refer to the DECnet-Plus documentation for equivalent information.

Table 9.7. Controlling Satellite Booting

Method		Comments
Disable MOP service on MOP servers temporarily		
Until the MOP server can complete its own startup operations, boot requests can be temporarily disabled by setting the DECnet Ethernet circuit to a "Service Disabled" state as shown:		This method prevents the MOP server from servicing the satellites; it does not prevent the satellites from requesting a boot from other MOP servers.
1	To disable MOP service during startup of a MOP server, enter the following commands: <pre>\$ MCR NCP DEFINE CIRCUIT MNA-1 - _ \$ SERVICE DISABLED \$ @SYS\$MANAGER:STARTNET \$ MCR NCP DEFINE CIRCUIT MNA-1 - _ \$ SERVICE ENABLED</pre>	If a satellite that is requesting a boot receives no response, it will make fewer boot requests over time. Thus, booting the satellite may take longer than normal once MOP service is reenabled. 1. MNA-1 represents the MOP service circuit. After entering these commands, service will be disabled in the volatile database. Do not disable service permanently.
2	To reenable MOP service later, enter the following commands in a command procedure so that they execute quickly and so that DECnet service to the users is not disrupted: <pre>\$ MCR NCP NCP> SET CIRCUIT MNA-1 STATE OFF NCP> SET CIRCUIT MNA-1 SERVICE ENABLED NCP> SET CIRCUIT MNA-1 STATE ON</pre>	2. Reenable service as shown.
Disable MOP service for individual satellites		
You can disable requests temporarily on a per-node basis in order to clear a node's information from the DECnet database. Clear a node's information from DECnet database on the MOP server using NCP, then reenable nodes as desired to control booting:		This method does not prevent satellites from requesting boot service from another MOP server.
1	To disable MOP service for a given node, enter the following command: <pre>\$ MCR NCP NCP> CLEAR NODE <i>satellite</i> HARDWARE ADDRESS</pre>	1. After entering the commands, service will be disabled in the volatile database. Do not disable service permanently.
2	To reenable MOP service for that node, enter the following command: <pre>\$ MCR NCP</pre>	2. Reenable service as shown.

Method	Comments
NCP> SET NODE <i>satellite</i> ALL	
Bring satellites to console prompt on shutdown	
Use any of the following methods to halt a satellite so that it halts (rather than reboots) upon restoration of power.	If you plan to use the DECnet Trigger operation, it is important to use a program to perform a HALT instruction that causes the satellite to enter console mode. This is because systems that support remote triggering only support it while the system is in console mode.
1 Use the VAXcluster Console System (VCS).	
2 Stop in console mode upon Halt or powerup: For Alpha computers: >>> (SET AUTO_ACTION HALT)	1. Some, but not all, satellites can be set up so they halt upon restoration of power or execution of a HALT instruction rather than automatically rebooting.
3 Set up a satellite so that it will stop in console mode when a HALT instruction is executed according to the instructions in the following list. a. Enter the following NCP commands so that a reboot will load an image that does a HALT instruction: \$ MCR NCP NCP> CLEAR NODE <i>node</i> LOAD ASSIST PARAMETER NCP> CLEAR NODE <i>node</i> LOAD ASSIST AGENT NCP> SET NODE <i>node</i> LOAD FILE - _ MOM\$LOAD:READ_ADDR.SYS b. Shut down the satellite, and specify an immediate reboot using the following SYSMAN command: \$ MCR SYSMAN SYSMAN> SET ENVIRONMENT/ NODE= <i>satellite</i> SYSMAN> DO @SYS \$UPDATE:AUTOGEN REBOOT c. When you want to allow the satellite to boot normally, enter the following NCP commands so that OpenVMS will be loaded later: \$ MCR NCP NCP> SET NODE <i>satellite</i> ALL	Note: You need to enter the SET commands only once on each system because the settings are saved in nonvolatile RAM. 2. The READ_ADDR.SYS program, which is normally used to find out the Ethernet address of a satellite node, also executes a HALT instruction upon its completion.

Important: When the SET HALT command is set up as described in Table 9.7, a power failure will cause the satellite to stop at the console prompt instead of automatically rebooting when power is

restored. This is appropriate for a mass power failure, but if someone trips over the power cord for a single satellite it can result in unnecessary unavailability.

You can provide a way to scan and trigger a reboot of satellites that go down this way by simply running a batch job periodically that performs the following tasks:

1. Uses the DCL lexical function F\$GETSYI to check each node that should be in the cluster.
2. Checks the CLUSTER_MEMBER lexical item.
3. Issues an NCP TRIGGER command for any satellite that is not currently a member of the cluster.

9.5. Configuring and Booting Satellite Nodes (Integrity servers)

Satellite

Any OpenVMS Version 8.3 system or a nPartition of a cell-based system can be used as a satellite. Support for nPartitions may require a firmware upgrade.

Satellite boot is supported over the core I/O LAN adapters only. All satellite systems must contain at least one local disk to support crash dumps and saving of the error log buffers across reboots. Diskless systems will not be able to take crash dumps in the event of abnormal software termination.

Boot Server

All Integrity server systems supported by OpenVMS Version 8.3 are supported as boot servers. At this time, VSI does not support cross-architecture booting for Integrity server satellite systems, so any cluster containing Integrity server satellite systems must have at least one Integrity server system to act as a boot node as well.

Required Software

- OpenVMS Version 8.3 or later
- TCP/IP Services for OpenVMS Version 5.6 or later

As with other satellite systems, the system software is read off of a disk served by one or more nodes to the cluster. The satellite system disk may be the same as the boot server's system disk but need not be. Unlike with Alpha satellites, where it was recommended but not required that the system disk be mounted on the boot server, Integrity server satellite systems require that the system disk be mounted on the boot server.

TCP/IP must be installed on the boot server's system disk. OpenVMS Version 8.3 must be installed on both the boot server's system disk and the satellite's system disk if different.

TCP/IP must be configured with BOOTP, TFTP and one or more interfaces enabled. At least one configured interface must be connected to a segment visible to the satellite systems. The boot server and all satellite systems will require an IP address. See the *VSI TCP/IP Services for OpenVMS Installation and Configuration* for details about configuring TCP/IP Services for OpenVMS.

9.5.1. Collecting Information from the Satellite System

If the satellite has a local disk with a version of OpenVMS installed, log in. If not, you may boot the installation DVD and select option 8 (Execute DCL commands and procedures) and execute the following commands:

```
$ LANCP := $LANCP
$ LANCP SHOW CONFIG
```

LAN Configuration:

Device	Parent	Medium/User	Version	Link	Speed	Duplex	Size	MAC	Address
Current	Address	Type							
EIB0		Ethernet	X-16	Up	1000	Full	1500	00-13-21-5B-86-49	
	00-13-21-5B-86-49	UTP	i82546						
EIA0		Ethernet	X-16	Up	1000	Full	1500	00-13-21-5B-86-48	
	00-13-21-5B-86-48	UTP	i82546						

Record the MAC address for the adapter you will use for booting. You will need it when defining the satellite system to the boot server. If the current address differs from the MAC address, use the MAC address.

9.5.2. Setting up the Satellite System for Booting and Crashing

If the satellite has a local disk with a version of OpenVMS installed, log in. If not, you may boot the installation DVD and select option 8 (Execute DCL commands and procedures.) Use `SYS $MANAGER:BOOT_OPTIONS.COM` to add a boot menu option for the network adapter from which you are booting. The procedure will ask you if this network entry is for a satellite boot and if so, it will set the Memory Disk boot option flag (0x200000) for that boot menu entry. The memory disk flag is required for satellite boot.

If you intended to use the system primarily for satellite boot, place the network boot option at position 1. The satellite system also requires DOSD (Dump Off the System Disk) for crash dumps and saving the unwritten error log buffers across reboots and crashes. `BOOT_OPTIONS.COM` may also be used to manage the DOSD device list. You may wish to create the DOSD device list at this time. See the *VSI OpenVMS System Manager's Manual* for information about setting up a DOSD device list.

9.5.3. Defining the Satellite System to the Boot Server

Integrity server Satellite systems boot via the PXE protocol. On OpenVMS, PXE is handled by BOOTP from the TCPIP product. If you are using more than one Integrity server system, which is a boot server in your cluster, be sure the BOOTP database is on a common disk. See the TCPIP documentation for information on configuring TCPIP components. TCPIP must be installed, configured and running before attempting to define a satellite system.

On an Integrity server system, which is a boot server, log in to the system manager's or other suitably privileged account. Execute the command procedure `SYS $MANAGER:CLUSTER_CONFIG_LAN.COM`. (`CLUSTER_CONFIG.COM`, which configures satellite nodes using DECnet, does not support Integrity server systems. It will, however, automatically invoke `CLUSTER_CONFIG_LAN` for Integrity server systems.) `CLUSTER_CONFIG_LAN` is a menu driven command procedure designed to help you configure satellite systems. The menus are context-sensitive and may vary depending on architecture and installed products. If you are unfamiliar with the procedure, please see refer to the System Management documentation for a more extensive overview of `CLUSTER_CONFIG_LAN`.

The essential information required to add an Integrity server satellite includes the node's SCS node name, SCS system ID, and hardware address. In addition, you will need to know the satellite's IP

address, network mask, and possibly gateway addresses. If you are unfamiliar with these concepts, please refer to the TCPIP documentation. The procedure will create a system root for the satellite.

CLUSTER_CONFIG_LAN should perform all steps required to make the satellite system bootable. If you choose local paging and swapping files, you will be prompted to boot the satellite system into the cluster so that the files may be created. If not, paging and swapping files will be created on the served system disk and you may boot the satellites at your convenience.

9.5.4. Booting the Satellite

If you have previously added an option to the boot menu, select that option. If you have not, see your hardware documentation for the steps required to boot from a network adapter. Be sure to set the environment variable VMS_FLAGS to include the memory disk boot flag (0x200000). The system will detail boot progress in the form of a system message when VMS_LOADER is obtained from the network, followed by one period character written to the console device for every file downloaded to start the boot sequence and last by a message indicating that IPB (the primary bootstrap image) has been loaded.

Note the following example:

```

Loading.: Satellite Boot EIA0 Mac(00-13-21-5b-86-48)
Running LoadFile()
CLIENT MAC ADDR: 00 13 21 5B 86 48
CLIENT IP: 16.116.43.79 MASK: 255.255.248.0 DHCP IP: 0.240.0.0
TSize.Running LoadFile()
Starting: Satellite Boot EIA0 Mac(00-13-21-5b-86-48)
Loading memory disk from IP 16.116.43.78
.....
Loading file: $13$DKA0:[SYS10.SYSCOMMON.SYSEXE]IPB.EXE from IP
 16.116.43.78
%IPB-I-SATSYSDIS, Satellite boot from system device $13$DKA0:
HP OpenVMS Industry Standard 64 Operating System, Version V8.3
© Copyright 1976-2006 Hewlett-Packard Development Company, L.P.
```

Upon first full boot, the satellite system will run AUTOGEN and reboot.

9.5.5. Additional Tasks on the Satellite System

If you had not done so previously, create the dump file for DOSD at this time. Edit the SYS \$STARTUP:SYCONFIG.COM file and add commands to mount the DOSD device. In order for the error log buffers to be recovered, the DOSD device must be mounted in SYCONFIG.

9.6. Booting Satellites with IP interconnect (Integrity servers, Alpha)

For Alpha satellite nodes, the satellite node and its boot server must exist in the same LAN segment. To select the interface to be used for satellite booting, assume that the satellite node does not have any disk running OpenVMS connected to it. If you are adding Alpha systems as satellite nodes, you can receive information from the ">>>>" prompt by executing the following command:

```

P00>>>show device
dga5245.1003.0.3.0          $1$DGA5245  COMPAQ HSV110 (C)COMPAQ  3028
dga5245.1004.0.3.0          $1$DGA5245  COMPAQ HSV110 (C)COMPAQ  3028
dga5890.1001.0.3.0          $1$DGA5890  COMPAQ HSV110 (C)COMPAQ  3028
```

```

dga5890.1002.0.3.0      $1$DGA5890   COMPAQ HSV110 (C)COMPAQ  3028
dka0.0.0.2004.0        DKA0         COMPAQ BD03685A24   HPB7
dka100.1.0.2004.0      DKA100       COMPAQ BD01864552   3B08
dka200.2.0.2004.0      DKA200       COMPAQ BD00911934   3B00
dqa0.0.0.15.0         DQA0         HL-DT-ST CD-ROM GCR-8480  2.11
dva0.0.0.1000.0        DVA0
eia0.0.0.2005.0        EIA0         00-06-2B-03-2D-7D
pga0.0.0.3.0          PGA0         WWN 1000-0000-c92a-78e9
pka0.7.0.2004.0        PKA0         SCSI Bus ID 7
pkb0.6.0.2.0          PKB0         SCSI Bus ID 6   5.57
P00>>>

```

From the output, the LAN interface will be EIA0 on which the IP address will be configured and used for Cluster configuration.

Note

The Alpha console uses the MOP protocol for network load of satellite systems. Since the MOP protocol is non-routable, the satellite boot server or servers and all satellites booting from them must reside in the same LAN. In addition, the boot server must have at least one LAN device enabled for cluster communications to permit the Alpha satellite nodes to access the system disk.

On Integrity server systems, the interface name will either start with EI or EW. If it is the first interface, it will be EIA0 or EWA0. Note the mac address of the interface that you want to use from the Shell prompt. To obtain the interface information on Integrity servers, execute the following command on the EFI Shell:

```
Shell> lanaddress
```

```
LAN Address Information
```

```

LAN Address      Path
-----
Mac(00306E4A133F)  Acpi(HWP0002,0)/Pci(3|0)/Mac(00306E4A133F)
*Mac(00306E4A02F9) Acpi(HWP0002,100)/Pci(2|0)/Mac(00306E4A02F9)

```

```
Shell>
```

Assuming that the active interface is EIA0, configure the satellite with EIA0, if it does not boot with EIA0 try with EWA0 subsequently. For more information about configuring a satellite node, see Section 8.2.3.4.

9.7. System-Disk Throughput

Achieving enough system-disk throughput requires some combination of the following techniques:

Technique	Reference
Avoid disk rebuilds at boot time.	Section 9.7.1
Offload work from the system disk.	Section 9.7.2
Configure multiple system disks.	Section 9.7.3
Use Volume Shadowing for OpenVMS.	Section 6.6

9.7.1. Avoiding Disk Rebuilds

The OpenVMS file system maintains a cache of preallocated file headers and disk blocks. When a disk is not properly dismounted, such as when a system fails, this preallocated space becomes temporarily unavailable. When the disk is mounted again, OpenVMS scans the disk to recover that space. This is called a **disk rebuild**.

A large OpenVMS Cluster system must ensure sufficient capacity to boot nodes in a reasonable amount of time. To minimize the impact of disk rebuilds at boot time, consider making the following changes:

Action	Result
Use the DCL command MOUNT/NOREBUILD for all user disks, at least on the satellite nodes. Enter this command into startup procedures that mount user disks.	It is undesirable to have a satellite node rebuild the disk, yet this is likely to happen if a satellite is the first to reboot after it or another node fails.
Set the system parameter ACP_REBLDSYSD to 0, at least for the satellite nodes.	This prevents a rebuild operation on the system disk when it is mounted implicitly by OpenVMS early in the boot process.
Avoid a disk rebuild during prime working hours by using the SET VOLUME/REBUILD command during times when the system is not so heavily used. Once the computer is running, you can run a batch job or a command procedure to execute the SET VOLUME/REBUILD command for each disk drive.	User response times can be degraded during a disk rebuild operation because most I/O activity on that disk is blocked. Because the SET VOLUME/REBUILD command determines whether a rebuild is needed, the job can execute the command for every disk. This job can be run during off hours, preferably on one of the more powerful nodes.

Caution: In large OpenVMS Cluster systems, large amounts of disk space can be preallocated to caches. If many nodes abruptly leave the cluster (for example, during a power failure), this space becomes temporarily unavailable. If your system usually runs with nearly full disks, do not disable rebuilds on the server nodes at boot time.

9.7.2. Offloading Work

In addition to the system disk throughput issues during an entire OpenVMS Cluster boot, access to particular system files even during steady-state operations (such as logging in, starting up applications, or issuing a PRINT command) can affect response times.

You can identify **hot** system files using a performance or monitoring tool (such as those listed in Section 1.5.2), and use the techniques in the following table to reduce hot file I/O activity on system disks:

Potential Hot Files	Methods to Help
Page and swap files	<p>When you run CLUSTER_CONFIG_LAN.COM or CLUSTER_CONFIG.COM to add computers to specify the sizes and locations of page and swap files, relocate the files as follows:</p> <ul style="list-style-type: none"> • Move page and swap files for computers off system disks. • Set up page and swap files for satellites on the satellites' local disks, if such disks are available.

Potential Hot Files	Methods to Help
Move these high-activity files off the system disk: <ul style="list-style-type: none"> • SYSUAF.DAT • NETPROXY.DAT • RIGHTSLIST.DAT • ACCOUNTNG.DAT • VMSMAIL_PROFILE.DATA • QMAN\$MASTER.DAT • Layered product and other application files 	Use any of the following methods: <ul style="list-style-type: none"> • Specify new locations for the files according to the instructions in Chapter 5. • Use caching in the HSC subsystem or in RF or RZ disks to improve the effective system-disk throughput. • Add a solid-state disk to your configuration. These devices have lower latencies and can handle a higher request rate than a regular magnetic disk. A solid-state disk can be used as a system disk or to hold system files. • Use DECram software to create RAM disks on MOP servers to hold copies of selected hot read-only files to improve boot times. A RAM disk is an area of main memory within a system that is set aside to store data, but it is accessed as if it were a disk.

Moving these files from the system disk to a separate disk eliminates most of the write activity to the system disk. This raises the read/write ratio and, if you are using Volume Shadowing for OpenVMS, maximizes the performance of shadowing on the system disk.

9.7.3. Configuring Multiple System Disks

Depending on the number of computers to be included in a large cluster and the work being done, you must evaluate the tradeoffs involved in configuring a single system disk or multiple system disks.

While a single system disk is easier to manage, a large cluster often requires more system disk I/O capacity than a single system disk can provide. To achieve satisfactory performance, multiple system disks may be needed. However, you should recognize the increased system management efforts involved in maintaining multiple system disks.

Consider the following when determining the need for multiple system disks:

- Concurrent user activity

In clusters with many satellites, the amount and type of user activity on those satellites influence system-disk load and, therefore, the number of satellites that can be supported by a single system disk. For example:

IF...	THEN...	Comments
Many users are active or run multiple applications simultaneously	The load on the system disk can be significant; multiple system disks may be required.	Some OpenVMS Cluster systems may need to be configured on the assumption that all users are constantly active. Such working conditions may require a larger, more expensive OpenVMS Cluster system that handles peak loads without performance degradation.

IF...	THEN...	Comments
Few users are active simultaneously	A single system disk might support a large number of satellites.	For most configurations, the probability is low that most users are active simultaneously. A smaller and less expensive OpenVMS Cluster system can be configured for these typical working conditions but may suffer some performance degradation during peak load periods.
Most users run a single application for extended periods	A single system disk might support a large number of satellites if significant numbers of I/O requests can be directed to application data disks.	Because each workstation user in an OpenVMS Cluster system has a dedicated computer, a user who runs large compute-bound jobs on that dedicated computer does not significantly affect users of other computers in the OpenVMS Cluster system. For clustered workstations, the critical shared resource is a disk server. Thus, if a workstation user runs an I/O-intensive job, its effect on other workstations sharing the same disk server might be noticeable.

- Concurrent booting activity

One of the few times when all OpenVMS Cluster computers are simultaneously active is during a cluster reboot. All satellites are waiting to reload the operating system, and as soon as a boot server is available, they begin to boot in parallel. This booting activity places a significant I/O load on the boot server, system disk, and interconnect.

Note: You can reduce overall cluster boot time by configuring multiple system disks and by distributing system roots for computers evenly across those disks. This technique has the advantage of increasing overall system disk I/O capacity, but it has the disadvantage of requiring additional system management effort. For example, installation of layered products or upgrades of the OpenVMS operating system must be repeated once for each system disk.

- System management

Because system management work load increases as separate system disks are added and does so in direct proportion to the number of separate system disks that need to be maintained, you want to minimize the number of system disks added to provide the required level of performance.

Volume Shadowing for OpenVMS is an alternative to creating multiple system disks. Volume shadowing increases the read I/O capacity of a single system disk and minimizes the number of separate system disks that have to be maintained because installations or updates need only be applied once to a volume-shadowed system disk. For clusters with substantial system disk I/O requirements, you can use multiple system disks, each configured as a shadow set.

Cloning the system disk is a way to manage multiple system disks. To clone the system disk:

- Create a system disk (or shadow set) with roots for all OpenVMS Cluster nodes.
- Use this as a master copy, and perform all software upgrades on this system disk.
- Back up the master copy to the other disks to create the *cloned* system disks.
- Change the volume names so they are unique.
- If you have not moved system files off the system disk, you must have the SYLOGICALS.COM startup file point to system files on the master system disk.
- Before an upgrade, be sure to save any changes you need from the cloned disks since the last upgrade, such as MODPARAMS.DAT and AUTOGEN feedback data, accounting files for billing, and password history.

9.8. Conserving System Disk Space

The essential files for a satellite root take up very little space, so that more than 96 roots can easily fit on a single system disk. However, if you use separate dump files for each satellite node or put page and swap files for all the satellite nodes on the system disk, you quickly run out of disk space.

9.8.1. Techniques

To avoid running out of disk space, set up common dump files for all the satellites or for groups of satellite nodes. For debugging purposes, it is best to have separate dump files for each MOP and disk server. Also, you can use local disks on satellite nodes to hold page and swap files, instead of putting them on the system disk. In addition, move page and swap files for MOP and disk servers off the system disk.

Reference: See Section 10.7 to plan a strategy for managing dump files.

9.9. Adjusting System Parameters

As an OpenVMS Cluster system grows, certain data structures within OpenVMS need to grow in order to accommodate the large number of nodes. If growth is not possible (for example, because of a shortage of nonpaged pool) this will induce intermittent problems that are difficult to diagnose. VSI recommends you to have a separate network for cluster communication. This can help avoid any user data interference with cluster traffic and suitable for environment that has high intra-cluster traffic.

You should run AUTOGEN with FEEDBACK frequently as a cluster grows, so that settings for many parameters can be adjusted. Refer to Section 8.7 for more information about running AUTOGEN.

In addition to running AUTOGEN with FEEDBACK, you should check and manually adjust the following parameters:

- SCSRESPCNT
- CLUSTER_CREDITS

SCS connections are now allocated and expanded only as needed, up to a limit of 65,000.

9.9.1. The SCSRESPCNT Parameter

Description: The SCSRESPCNT parameter controls the number of response descriptor table (RDT) entries available for system use. An RDT entry is required for every in-progress message exchange between two nodes.

Symptoms of entry shortages: A shortage of entries affects performance, since message transmissions must be delayed until a free entry is available.

How to determine a shortage of RDT entries: Use the SDA utility as follows to check each system for requests that waited because there were not enough free RDTs.

```
SDA> READ SYS$SYSTEM:SCSDEF
%SDA-I-READSYM, reading symbol table SYS$COMMON:[SYSEXE]SCSDEF.STB;1
SDA> EXAM @SCS$GL_RDT + RDT$L_QRDT_CNT
8044DF74: 00000000 "...."
SDA>
```

How to resolve shortages: If the SDA EXAMINE command displays a nonzero value, RDT waits have occurred. If you find a count that tends to increase over time under normal operations, increase SCSRESPCNT.

9.9.2. The CLUSTER_CREDITS Parameter

Description: The CLUSTER_CREDITS parameter specifies the number of per-connection buffers a node allocates to receiving VMS\$VAXcluster communications. This system parameter is not dynamic; that is, if you change the value, you must reboot the node on which you changed it.

Default: The default value is 10. The default value may be insufficient for a cluster that has very high locking rates.

Symptoms of cluster credit problem: A shortage of credits affects performance, since message transmissions are delayed until free credits are available. These are visible as credit waits in the SHOW CLUSTER display.

How to determine whether credit waits exist: Use the SHOW CLUSTER utility as follows:

1. Run SHOW CLUSTER/CONTINUOUS.
2. Type REMOVE SYSTEM/TYPE=HS.
3. Type ADD LOC_PROC, CR_WAIT.
4. Type SET CR_WAIT/WIDTH=10.
5. Check to see whether the number of CR_WAITS (credit waits) logged against the VMS \$VAXcluster connection for any remote node is incrementing regularly. Ideally, credit waits should not occur. However, occasional waits under very heavy load conditions are acceptable.

How to resolve incrementing credit waits:

If the number of CR_WAITS is incrementing more than once per minute, perform the following steps:

1. Increase the CLUSTER_CREDITS parameter on the node against which they are being logged by five. The parameter should be modified on the remote node, not on the node which is running SHOW CLUSTER.
2. Reboot the node.

Note that it is not necessary for the CLUSTER_CREDITS parameter to be the same on every node.

9.10. Minimize Network Instability

Network instability also affects OpenVMS Cluster operations. Table 9.8 lists techniques to minimize typical network problems.

Table 9.8. Techniques to Minimize Network Problems

Technique	Recommendation
Adjust the RECNXINTERVAL parameter.	<p>The RECNXINTERVAL system parameter specifies the number of seconds the OpenVMS Cluster system waits when it loses contact with a node, before removing the node from the configuration. Many large OpenVMS Cluster configurations operate with the RECNXINTERVAL parameter set to 40 seconds (the default value is 20 seconds).</p> <p>Raising the value of RECNXINTERVAL can result in longer perceived application pauses, especially when the node leaves the OpenVMS Cluster system abnormally. The pause is caused by the connection manager waiting for the number of seconds specified by RECNXINTERVAL.</p>
Protect the network	<p>For clusters connected on the LAN interconnect, treat the LAN as if it were a part of the OpenVMS Cluster system. For example, do not allow an environment in which a random user can disconnect a ThinWire segment to attach a new PC while 20 satellites hang.</p> <p>For Clusters running on IP interconnect, ensure that the IP network is protected using a VPN type of security.</p>
Choose your hardware and configuration carefully.	<p>Certain hardware is not suitable for use in a large OpenVMS Cluster system.</p> <ul style="list-style-type: none"> • Some network components can appear to work well with light loads, but are unable to operate properly under high traffic conditions. Improper operation can result in lost or corrupted packets that will require packet retransmissions. This reduces performance and can affect the stability of the OpenVMS Cluster configuration. • Beware of bridges that cannot filter and forward at full line rates and repeaters that do not handle congested conditions well. • Refer to <i>Guidelines for OpenVMS Cluster Configurations</i> to determine appropriate OpenVMS Cluster configurations and capabilities.
Use the LAVC \$FAILURE_ANALYSIS facility.	See Section D.5 for assistance in the isolation of network faults.

9.11. DECnet Cluster Alias

You should define a cluster alias name for the OpenVMS Cluster to ensure that remote access will be successful when at least one OpenVMS Cluster member is available to process the client program's requests.

The cluster alias acts as a single network node identifier for an OpenVMS Cluster system. Computers in the cluster can use the alias for communications with other computers in a DECnet network. Note that it is possible for nodes running DECnet for OpenVMS to have a unique and separate cluster alias from nodes running DECnet-Plus. In addition, clusters running DECnet-Plus can have one cluster alias for VAX, one for Alpha, and another for both.

Note: A single cluster alias can include nodes running either DECnet for OpenVMS or DECnet-Plus, but not both. Also, an OpenVMS Cluster running both DECnet for OpenVMS and DECnet-Plus requires multiple system disks (one for each).

Reference: See Chapter 4 for more information about setting up and using a cluster alias in an OpenVMS Cluster system.

Chapter 10. Maintaining an OpenVMS Cluster System

Once your cluster is up and running, you can implement routine, site-specific maintenance operations—for example, backing up disks or adding user accounts, performing software upgrades and installations, running AUTOGEN with the feedback option on a regular basis, and monitoring the system for performance.

You should also maintain records of current configuration data, especially any changes to hardware or software components. If you are managing a cluster that includes satellite nodes, it is important to monitor LAN activity.

From time to time, conditions may occur that require the following special maintenance operations:

- Restoring cluster quorum after an unexpected computer failure
- Executing conditional shutdown operations
- Performing security functions in LAN and mixed-interconnect clusters

10.1. Backing Up Data and Files

As a part of the regular system management procedure, you should copy operating system files, application software files, and associated files to an alternate device using the OpenVMS Backup utility.

Some backup operations are the same in an OpenVMS Cluster as they are on a single OpenVMS system. For example, an incremental back up of a disk while it is in use, or the backup of a nonshared disk.

Backup tools for use in a cluster include those listed in Table 10.1.

Table 10.1. Backup Methods

Tool	Usage
Online backup	Use from a running system to back up: <ul style="list-style-type: none">• The system's local disks• Cluster-shareable disks other than system disks• The system disk or disks <p>Caution: Files open for writing at the time of the backup procedure may not be backed up correctly.</p>
Menu-driven	If you have access to the OpenVMS Alpha distribution CD-ROM, back up your system using the menu system provided on that disc. This menu system, which is displayed automatically when you boot the CD-ROM, allows you to: <ul style="list-style-type: none">• Enter a DCL environment, from which you can perform backup and restore operations on the system disk (instead of using standalone BACKUP).

Tool	Usage
	<ul style="list-style-type: none"> Install or upgrade the operating system and layered products, using the POLYCENTER Software Installation utility. <p>Reference: For more detailed information about using the menu-driven procedure, see the <i>OpenVMS Upgrade and Installation Manual</i> and the <i>VSI OpenVMS System Manager's Manual</i>.</p>

Plan to perform the backup process regularly, according to a schedule that is consistent with application and user needs. This may require creative scheduling so that you can coordinate backups with times when user and application system requirements are low.

Reference: See the *VSI OpenVMS System Management Utilities Reference Manual* for complete information about the OpenVMS Backup utility.

10.2. Updating the OpenVMS Operating System

When updating the OpenVMS operating system, follow the steps in Table 10.2.

Table 10.2. Upgrading the OpenVMS Operating System

Step	Action
1	Back up the system disk.
2	Perform the update procedure once for each system disk.
3	Install any mandatory updates.
4	Run AUTOGEN on each node that boots from that system disk.
5	Run the user environment test package (UETP) to test the installation.
6	Use the OpenVMS Backup utility to make a copy of the new system volume.

Reference: See the appropriate OpenVMS upgrade and installation manual for complete instructions.

10.2.1. Rolling Upgrades

The OpenVMS operating system allows an OpenVMS Cluster system running on multiple system disks to continue to provide service while the system software is being upgraded. This process is called a **rolling upgrade** because each node is upgraded and rebooted in turn, until all the nodes have been upgraded.

If you must first migrate your system from running on one system disk to running on two or more system disks, follow these steps:

Step	Action
1	Follow the procedures in Section 8.5 to create a duplicate disk.
2	Follow the instructions in Section 5.8 for information about coordinating system files.

These sections help you add a system disk and prepare a common user environment on multiple system disks to make the shared system files such as the queue database, rights lists, proxies, mail, and other files available across the OpenVMS Cluster system.

10.3. LAN Network Failure Analysis

The OpenVMS operating system provides a sample program to help you analyze OpenVMS Cluster network failures on the LAN. You can edit and use the `SYS$EXAMPLES:LAVC $FAILURE_ANALYSIS.MAR` program to detect and isolate failed network components. Using the network failure analysis program can help reduce the time required to detect and isolate a failed network component, thereby providing a significant increase in cluster availability.

Reference: For a description of the network failure analysis program, refer to Appendix D.

10.4. Recording Configuration Data

To maintain an OpenVMS Cluster system effectively, you must keep accurate records about the current status of all hardware and software components and about any changes made to those components. Changes to cluster components can have a significant effect on the operation of the entire cluster. If a failure occurs, you may need to consult your records to aid problem diagnosis.

Maintaining current records for your configuration is necessary both for routine operations and for eventual troubleshooting activities.

10.4.1. Record Information

At a minimum, your configuration records should include the following information:

- A diagram of your physical cluster configuration. (Appendix D includes a discussion of keeping a LAN configuration diagram).
- `SCSNODE` and `SCSSYSTEMID` parameter values for all computers.
- `VOTES` and `EXPECTED_VOTES` parameter values.
- DECnet names and addresses for all computers.
- Current values for cluster-related system parameters, especially `ALLOCLASS` and `TAPE_ALLOCLASS` values for HSC subsystems and computers.

Reference: Cluster system parameters are described in Appendix A.

- Names and locations of default bootstrap command procedures for all computers connected with the CI.
- Names of cluster disk and tape devices.
- In LAN and mixed-interconnect clusters, LAN hardware addresses for satellites.
- Names of LAN adapters.
- Names of LAN segments or rings.
- Names of LAN bridges and switches and port settings.
- Names of wiring concentrators or of `DELNI` or `DEMPR` adapters.

- Serial numbers of all hardware components.
- Changes to any hardware or software components (including site-specific command procedures), along with dates and times when changes were made.

10.4.2. Satellite Network Data

The first time you execute `CLUSTER_CONFIG.COM` to add a satellite, the procedure creates the file `NETNODE_UPDATE.COM` in the boot server's `SYSS$SPECIFIC:[SYSMGR]` directory. (For a common-environment cluster, you must rename this file to the `SYSS$COMMON:[SYSMGR]` directory, as described in Section 5.8.2). This file, which is updated each time you add or remove a satellite or change its Ethernet hardware address, contains all essential network configuration data for the satellite.

If an unexpected condition at your site causes configuration data to be lost, you can use `NETNODE_UPDATE.COM` to restore it. You can also read the file when you need to obtain data about individual satellites. Note that you may want to edit the file occasionally to remove obsolete entries.

Example 10.1 shows the contents of the file after satellites `EUROPA` and `GANYMD` have been added to the cluster.

Example 10.1. Sample `NETNODE_UPDATE.COM` File

```
$ RUN SYS$SYSTEM:NCP
  define node EUROPA address 2.21
  define node EUROPA hardware address 08-00-2B-03-51-75
  define node EUROPA load assist agent sys$share:niscs_laa.exe
  define node EUROPA load assist parameter $1$DGALL:<SYS10.>
  define node EUROPA tertiary loader sys$system:tertiary_vmb.exe
  define node GANYMD address 2.22
  define node GANYMD hardware address 08-00-2B-03-58-14
  define node GANYMD load assist agent sys$share:niscs_laa.exe
  define node GANYMD load assist parameter $1$DGALL:<SYS11.>
  define node GANYMD tertiary loader sys$system:tertiary_vmb.exe
```

Reference: See the DECnet-Plus documentation for equivalent NCL command information.

10.5. Controlling OPCOM Messages

When a satellite joins the cluster, the Operator Communications Manager (OPCOM) has the following default states:

- For all systems in an OpenVMS Cluster configuration except workstations:
 - `OPA0`: is enabled for all message classes.
 - The log file `SYSS$MANAGER:OPERATOR.LOG` is opened for all classes.
- For workstations in an OpenVMS Cluster configuration, even though the OPCOM process is running:
 - `OPA0`: is not enabled.

- No log file is opened.

10.5.1. Overriding OPCOM Defaults

Table 10.3 shows how to define the following system logical names in the command procedure SYS\$MANAGER:SYLOGICALS.COM to override the OPCOM default states.

Table 10.3. OPCOM System Logical Names

System Logical Name	Function
OPC\$OPA0_ENABLE	If defined to be true, OPA0: is enabled as an operator console. If defined to be false, OPA0: is not enabled as an operator console. DCL considers any string beginning with T or Y or any odd integer to be true, all other values are false.
OPC\$OPA0_CLASSES	<p>Defines the operator classes to be enabled on OPA0:. The logical name can be a search list of the allowed classes, a list of classes, or a combination of the two. For example:</p> <pre> \$ DEFINE/SYSTEM OP\$OPA0_CLASSES CENTRAL,DISKS,TAPE \$ DEFINE/SYSTEM OP\$OPA0_CLASSES "CENTRAL,DISKS,TAPE" \$ DEFINE/SYSTEM OP\$OPA0_CLASSES "CENTRAL,DISKS",TAPE </pre> <p>You can define OPC\$OPA0_CLASSES even if OPC\$OPA0_ENABLE is not defined. In this case, the classes are used for any operator consoles that are enabled, but the default is used to determine whether to enable the operator console.</p>
OPC\$LOGFILE_ENABLE	If defined to be true, an operator log file is opened. If defined to be false, no log file is opened.
OPC\$LOGFILE_CLASSES	Defines the operator classes to be enabled for the log file. The logical name can be a search list of the allowed classes, a comma-separated list, or a combination of the two. You can define this system logical even when the OPC\$LOGFILE_ENABLE system logical is not defined. In this case, the classes are used for any log files that are open, but the default is used to determine whether to open the log file.
OPC\$LOGFILE_NAME	Supplies information that is used in conjunction with the default name SY\$MANAGER:OPERATOR.LOG to define the name of the log file. If the log file is directed to a disk other than the system disk, you should include commands to mount that disk in the SYLOGICALS.COM command procedure.

10.5.2. Example

The following example shows how to use the OPC\$OPA0_CLASSES system logical to define the operator classes to be enabled. The following command prevents SECURITY class messages from being displayed on OPA0.

```
$ DEFINE/SYSTEM OPC$OPA0_CLASSES CENTRAL,PRINTER,TAPES,DISKS,DEVICES, -
```

```
_$ CARDS , NETWORK , CLUSTER , LICENSE , OPER1 , OPER2 , OPER3 , OPER4 , OPER5 , -  
_$ OPER6 , OPER7 , OPER8 , OPER9 , OPER10 , OPER11 , OPER12
```

In large clusters, state transitions (computers joining or leaving the cluster) generate many multiline OPCOM messages on a boot server's console device. You can avoid such messages by including the DCL command `REPLY/DISABLE=CLUSTER` in the appropriate site-specific startup command file or by entering the command interactively from the system manager's account.

10.6. Shutting Down a Cluster

The `SHUTDOWN` command of the `SYSMAN` utility provides five options for shutting down OpenVMS Cluster computers:

- `NONE` (the default)
- `REMOVE_NODE`
- `CLUSTER_SHUTDOWN`
- `REBOOT_CHECK`
- `SAVE_FEEDBACK`

These options are described in the following sections.

10.6.1. The `NONE` Option

If you select the default `SHUTDOWN` option `NONE`, the shutdown procedure performs the normal operations for shutting down a standalone computer. If you want to shut down a computer that you expect will rejoin the cluster shortly, you can specify the default option `NONE`. In that case, cluster quorum is not adjusted because the operating system assumes that the computer will soon rejoin the cluster.

In response to the “Shutdown options [NONE]:” prompt, you can specify the `DISABLE_AUTOSTART= n` option, where `n` is the number of minutes before autostart queues are disabled in the shutdown sequence. For more information about this option, see Section 7.13.

10.6.2. The `REMOVE_NODE` Option

If you want to shut down a computer that you expect will not rejoin the cluster for an extended period, use the `REMOVE_NODE` option. For example, a computer may be waiting for new hardware, or you may decide that you want to use a computer for standalone operation indefinitely.

When you use the `REMOVE_NODE` option, the active quorum in the remainder of the cluster is adjusted downward to reflect the fact that the removed computer's votes no longer contribute to the quorum value. The shutdown procedure readjusts the quorum by issuing the `SET CLUSTER/EXPECTED_VOTES` command, which is subject to the usual constraints described in Section 10.11.

Note: The system manager is still responsible for changing the `EXPECTED_VOTES` system parameter on the remaining OpenVMS Cluster computers to reflect the new configuration.

10.6.3. The `CLUSTER_SHUTDOWN` Option

When you choose the `CLUSTER_SHUTDOWN` option, the computer completes all shut down activities up to the point where the computer would leave the cluster in a normal shutdown situation.

At this point the computer waits until all other nodes in the cluster have reached the same point. When all nodes have completed their shutdown activities, the entire cluster dissolves in one synchronized operation. The advantage of this is that individual nodes do not complete shutdown independently, and thus do not trigger state transitions or potentially leave the cluster without quorum.

When performing a CLUSTER_SHUTDOWN you must specify this option on every OpenVMS Cluster computer. If any computer is not included, clusterwide shutdown cannot occur.

10.6.4. The REBOOT_CHECK Option

When you choose the REBOOT_CHECK option, the shutdown procedure checks for the existence of basic system files that are needed to reboot the computer successfully and notifies you if any files are missing. You should replace such files before proceeding. If all files are present, the following informational message appears:

```
%SHUTDOWN-I-CHECKOK, Basic reboot consistency check completed.
```

Note: You can use the REBOOT_CHECK option separately or in conjunction with either the REMOVE_NODE or the CLUSTER_SHUTDOWN option. If you choose REBOOT_CHECK with one of the other options, you must specify the options in the form of a comma-separated list.

10.6.5. The SAVE_FEEDBACK Option

Use the SAVE_FEEDBACK option to enable the AUTOGEN feedback operation.

Note: Select this option only when a computer has been running long enough to reflect your typical work load.

Reference: For detailed information about AUTOGEN feedback, see the *VSI OpenVMS System Manager's Manual*.

10.6.6. Shutting Down TCP/IP

Where clusters use IP as the interconnect, shutting down the TCP/IP connection results in loss of connection between the node and the existing members of the cluster. As a result, the Quorum of the cluster hangs, leading to the CLUEXIT crash. Therefore, ensure that all software applications are closed before shutting down TCP/IP

Shut down TCP/IP as shown:

```
$@SYSS$MANAGER:TCPIPCONFIG
Checking TCP/IP Services for OpenVMS configuration database files.
```

```
HP TCP/IP Services for OpenVMS Configuration Menu
```

```
Configuration options:
```

```

1 - Core environment
2 - Client components
3 - Server components
4 - Optional components
5 - Shutdown HP TCP/IP Services for OpenVMS
6 - Startup HP TCP/IP Services for OpenVMS
7 - Run tests
A - Configure options 1 - 4
[E] - Exit configuration procedure
```

```
Enter configuration option: 5
Begin Shutdown...
```

```
TCPIP$SHUTDOWN has detected the presence of IPCI configuration
file: SYS$SYSROOT:[SYSEXE]TCPIP$CLUSTER.DAT;
```

```
If you are using TCP/IP as your only cluster communication
channel, then stopping TCP/IP will cause this system to
CLUEXIT. Remote systems may also CLUEXIT.
```

```
Non-interactive. Continuing with TCP/IP shutdown ...
```

10.7. Dump Files

Whether your OpenVMS Cluster system uses a single common system disk or multiple system disks, you should plan a strategy to manage dump files.

10.7.1. Controlling Size and Creation

Dump-file management is especially important for large clusters with a single system disk. For example, on a 1 GB OpenVMS Alpha computer, AUTOGEN creates a dump file in excess of 350,000 blocks.

In the event of a software-detected system failure, each computer normally writes the contents of memory as a compressed selective dump file on its system disk for analysis. AUTOGEN calculates the size of the file based on the size of physical memory and the number of processes. If system disk space is limited (as is probably the case if a single system disk is used for a large cluster), you may want to specify that no dump file be created for satellites.

You can control dump-file size and creation for each computer by specifying appropriate values for the AUTOGEN symbols DUMPSTYLE and DUMPFIL in the computer's MODPARAMS.DAT file. For example, specify dump files as shown in Table 10.4.

Table 10.4. AUTOGEN Dump-File Symbols

Value Specified	Result
DUMPSTYLE = 9	Compressed selective dump file created (default)
DUMPFIL = 0	No dump file created
DUMPFIL = <i>n</i>	Dump file of size <i>n</i> created

Refer to the *VSI OpenVMS System Manager's Manual* for more information on dump files and Dump Off System Disk (DOSD).

Caution: Although you can configure computers without dump files, the lack of a dump file can make it difficult or impossible to determine the cause of a system failure.

The recommended method for controlling dump file size and location is using AUTOGEN and MODPARAMS.DAT. However, if necessary, the SYSGEN utility can be used explicitly. The following example shows the use of SYSGEN to modify the system dump-file size on large-memory systems:

```
$ MCR SYSGEN
SYSGEN> USE CURRENT
```

```

SYSGEN> SET DUMPSTYLE 9
SYSGEN> WRITE CURRENT
SYSGEN> CREATE SYS$SYSTEM:SYSDUMP.DMP/SIZE=350000
SYSGEN> EXIT
$ @SHUTDOWN

```

The dump-file size of 35,000 blocks is sufficient to cover about 1 GB of memory. This size is usually large enough to encompass the information needed to analyze a system failure.

After the system reboots, you can purge SYSDUMP.DMP.

10.7.2. Sharing Dump Files

Another option for saving dump-file space is to share a single dump file among multiple computers. While this technique makes it possible to analyze isolated computer failures, dumps will be lost if multiple computers fail at the same time or if a second computer fails before you can analyze the first failure. Because boot server failures have a greater impact on cluster operation than do failures of other computers you should configure dump files on boot servers to help ensure speedy analysis of problems.

Dump files cannot be shared between architectures. However, you can share a single dump file among multiple Alpha computers, and another single dump file among multiple Integrity computers and another single dump file among VAX computers. Follow these steps for each operating system:

Step	Action
1	Decide whether to use full or selective dump files. (Selective recommended).
2	Determine the size of the largest dump file needed by any satellite.
3	Select a satellite whose memory configuration is the largest of any in the cluster and do the following: <ol style="list-style-type: none"> 1. Specify DUMPSTYLE = 9 (or DUMPSTYLE = 8) in that satellite's MODPARAMS.DAT file. 2. Remove any DUMPFIL symbol from the satellite's MODPARAMS.DAT file. 3. Run AUTOGEN on that satellite to create a dump file.
4	Rename the dump file to SYS\$COMMON:[SYSEXE]SYSDUMP-COMMON.DMP or create a new dump file named SYSDUMP-COMMON.DMP in SYS\$COMMON:[SYSEXE].
5	Rename the old system-specific dump file on each system that has its own dump file: <pre>\$ RENAME SYS\$SYSDEVICE:[SYSn.SYSEXE]SYSDUMP.DMP .OLD</pre> <p>The value of n in the command line is the root for each system (for example, SYS0 or SYS1). Rename the file so that the operating system software does not use it as the dump file when the system is rebooted.</p>
6	For each satellite that is to share the dump file, do the following: <ol style="list-style-type: none"> 1. Create a file synonym entry for the dump file in the system-specific root. For example, to create a synonym for the satellite using root SYS1E, enter a command like the following: <pre>\$ SET FILE SYS\$COMMON:[SYSEXE]SYSDUMP-COMMON.DMP -</pre>

Step	Action
	<pre>_\$ /ENTER=SYS\$SYSDEVICE:[SYS1E.SYSEXE]SYSDUMP.DMP</pre> <p>2. Add the following lines to the satellite's MODPARAMS.DAT file:</p> <pre>DUMPFIL = 0DUMPSTYLE = 0 (or DUMPSTYLE = 1)</pre>
7	Reboot each node so it can map to the new common dump file. The operating system software cannot use the new file for a crash dump until you reboot the system.
8	After you reboot, delete the SYSDUMP.OLD file in each system-specific root. Do not delete any file called SYSDUMP.DMP; instead, rename it, reboot, and then delete it as described in steps 5 and 7.

10.8. Maintaining the Integrity of OpenVMS Cluster Membership

Because multiple LAN and mixed-interconnect clusters coexist on a single extended LAN, the operating system provides mechanisms to ensure the integrity of individual clusters and to prevent access to a cluster by an unauthorized computer.

The following mechanisms are designed to ensure the integrity of the cluster:

- A cluster authorization file (SYS\$COMMON:[SYSEXE]CLUSTER_AUTHORIZE.DAT), which is initialized during installation of the operating system or during execution of the CLUSTER_CONFIG.COM CHANGE function. The file is maintained with the SYSMAN utility.
- Control of conversational bootstrap operations on satellites.

The purpose of the cluster group number and password is to prevent accidental access to the cluster by an unauthorized computer. Under normal conditions, the system manager specifies the cluster group number and password either during installation or when you run CLUSTER_CONFIG.COM (see Example 8.13) to convert a standalone computer to run in an OpenVMS Cluster system.

OpenVMS Cluster systems use these mechanisms to protect the integrity of the cluster in order to prevent problems that could otherwise occur under circumstances like the following:

- When setting up a new cluster, the system manager specifies a group number identical to that of an existing cluster on the same Ethernet.
- A satellite user with access to a local system disk tries to join a cluster by executing a conversational SYSBOOT operation at the satellite's console.

Reference: These mechanisms are discussed in Section 10.8.1 and Section 8.2.1, respectively.

10.8.1. Cluster Group Data

The cluster authorization file, SYS\$COMMON:[SYSEXE]CLUSTER_AUTHORIZE.DAT, contains the cluster group number and (in scrambled form) the cluster password. The CLUSTER_AUTHORIZE.DAT file is accessible only to users with the SYSPRV privilege.

Under normal conditions, you need not alter records in the CLUSTER_AUTHORIZE.DAT file interactively. However, if you suspect a security breach, you may want to change the cluster password. In that case, you use the SYSMAN utility to make the change.

To change the cluster password, follow these instructions:

Step	Action
1	Invoke the SYSMAN utility.
2	Log in as system manager on a boot server.
3	Enter the following command: <pre>\$ RUN SYS\$SYSTEM:SYSMAN SYSMAN></pre>
4	At the SYSMAN> prompt, enter any of the CONFIGURATION commands in the following list. <ul style="list-style-type: none"> • CONFIGURATION SET CLUSTER_AUTHORIZATION Updates the cluster authorization file, CLUSTER_AUTHORIZE.DAT, in the directory SYS\$COMMON:[SYSEXE]. The SET command creates this file if it does not already exist. You can include the following qualifiers on this command: <ul style="list-style-type: none"> • /GROUP_NUMBER—Specifies a cluster group number. Group number must be in the range from 1 to 4095 or 61440 to 65535. • /PASSWORD—Specifies a cluster password. Password may be from 1 to 31 characters in length and may include alphanumeric characters, dollar signs (\$), and underscores (_). • CONFIGURATION SHOW CLUSTER_AUTHORIZATION Displays the cluster group number. • HELP CONFIGURATION SET CLUSTER_AUTHORIZATION Explains the command's functions.
5	If your configuration has multiple system disks, each disk must have a copy of CLUSTER_AUTHORIZE.DAT. You must run the SYSMAN utility to update all copies.
Caution: If you change either the group number or the password, you must reboot the entire cluster. For instructions, see Section 8.6.	

10.8.2. Example

Example 10.2 illustrates the use of the SYSMAN utility to change the cluster password.

Example 10.2. Sample SYSMAN Session to Change the Cluster Password

```
$ RUN SYS$SYSTEM:SYSMAN
SYSMAN> SET ENVIRONMENT/CLUSTER
%SYSMAN-I-ENV, current command environment:
    Clusterwide on local cluster
    Username SYSTEM          will be used on nonlocal nodes
SYSMAN> SET PROFILE/PRIVILEGES=SYSPRV
SYSMAN> CONFIGURATION SET CLUSTER_AUTHORIZATION/PASSWORD=NEWPASSWORD
%SYSMAN-I-CAFOLDGROUP, existing group will not be changed
%SYSMAN-I-CAFREBOOT, cluster authorization file updated
```

```
The entire cluster should be rebooted.  
SYSMAN> EXIT  
$
```

10.9. Adjusting Packet Size for LAN or IP Configurations

You can adjust the maximum packet size for LAN configurations with the `NISCS_MAX_PKTSZ` system parameter.

10.9.1. System Parameter Settings for LANs and IPs

Starting with OpenVMS Version 7.3, the operating system (PEdriver) automatically detects the maximum packet size of all the virtual circuits to which the system is connected. If the maximum packet size of the system's interconnects is smaller than the default packet-size setting, PEdriver automatically reduces the default packet size.

Starting with OpenVMS 8.4, OpenVMS can make use of HP TCP/IP services for cluster communications using the UDP protocol. `NISCS_MAX_PKTSZ` will only affect the LAN channel payload size. To affect the IP channel payload size use the `NISCS_UDP_PKTSZ` parameter. For more information about the `NISCS_UDP_PKTSZ` parameter, see `HELP`.

10.9.2. How to Use `NISCS_MAX_PKTSZ`

To obtain this parameter's current, default, minimum, and maximum values, use the following command:

```
$ MC SYSGEN SHOW NISCS_MAX_PKTSZ
```

You can use the `NISCS_MAX_PKTSZ` parameter to reduce packet size, which in turn can reduce memory consumption. However, reducing packet size can also increase CPU utilization for block data transfers, because more packets will be required to transfer a given amount of data. Lock message packets are smaller than the minimum value, so the `NISCS_MAX_PKTSZ` setting will not affect locking performance.

You can also use `NISCS_MAX_PKTSZ` to force use of a common packet size on all LAN paths by bounding the packet size to that of the LAN path with the smallest packet size. Using a common packet size can avoid VC closure due to packet size reduction when failing down to a slower, smaller packet size network.

If a memory-constrained system, such as a workstation, has adapters to a network path with large-size packets, such as FDDI or Gigabit Ethernet with jumbo packets, then you may want to conserve memory by reducing the value of the `NISCS_MAX_PKTSZ` parameter.

10.9.3. How to Use `NISCS_UDP_PKTSZ`

This parameter specifies the upper limit on the size, in bytes, of the user data area in the largest packet sent by NISCA on any IP network.

`NISCS_UDP_PKTSZ` allows the system manager to change the packet size used for cluster communications over IP on network communication paths.

PEdriver uses NISCS_UDP_PKTSZ to compute the maximum amount of data to transmit in any packet.

Currently, the maximum payload over an IP channel is defined by one of the following three parameters. The least of the 3 values will be in effect.

- NISCS_UDP_PKTSZ SYSGEN parameter
- 1500 bytes
- IP_MTU of the interface supported by TCP/IP stack

Note

This parameter only affects the IP channel payload and not the LAN channel payload. LAN channel payload is controlled by the NISCS_MAX_PKTSZ parameter.

10.9.4. Editing Parameter Files

If you decide to change the value of the NISCS_MAX_PKTSZ or NISCS_UDP_PKTSZ parameter, edit the SYSSSPECIFIC:[SYSEXE]MODPARAMS.DAT file to permit AUTOGEN to factor the changed packet size into its calculations.

10.10. Determining Process Quotas

On Alpha systems, process quota default values in SYSUAF.DAT are often higher than the SYSUAF.DAT defaults on VAX systems. How, then, do you choose values for processes that could run on Alpha systems or on VAX systems in an OpenVMS Cluster? Understanding how a process is assigned quotas when the process is created in a dual-architecture OpenVMS Cluster configuration will help you manage this task.

10.10.1. Quota Values

The quotas to be used by a new process are determined by the OpenVMS LOGINOUT software. LOGINOUT works the same on OpenVMS Alpha and OpenVMS VAX systems. When a user logs in and a process is started, LOGINOUT uses the *larger* of:

- The value of the quota defined in the process's SYSUAF.DAT record
- The *current* value of the corresponding PQL_M *quota* system parameter on the host node in the OpenVMS Cluster

Example: LOGINOUT compares the value of the account's ASTLM process limit (as defined in the common SYSUAF.DAT) with the value of the PQL_MASTLM system parameter on the host Alpha system or on the host VAX system in the OpenVMS Cluster.

10.10.2. PQL Parameters

The letter M in PQL_M means minimum. The PQL_M *quotas* system parameters set a minimum value for the quotas. In the Current and Default columns of the following edited SYSMAN display, note how the current value of each PQL_M *quota* parameter exceeds its system-defined default value in most cases.

```
$ RUN SYS$SYSTEM:SYSMAN
```

```
SYSMAN> PARAMETER SHOW/PQL
```

```
%SYSMAN-I-USEACTNOD, a USE ACTIVE has been defaulted on node I64MOZ
```

```
Node I64MOZ: Parameters in use: ACTIVE
```

Parameter Name	Current	Default	Minimum	Maximum	Unit	Dynamic
PQL_DASTLM	24	24	4	-1	Ast	D
PQL_MASTLM	4	4	4	-1	Ast	D
PQL_DBIOLM	32	32	4	-1	I/O	D
PQL_MBIOLM	4	4	4	-1	I/O	D
PQL_DBYTLM	262144	262144	128000	-1	Bytes	D
PQL_MBYTLM	128000	128000	128000	-1	Bytes	D
PQL_DCPULM	0	0	0	-1	10Ms	D
PQL_MCPULM	0	0	0	-1	10Ms	D
PQL_DDIOLM	32	32	4	-1	I/O	D
PQL_MDIOLM	4	4	4	-1	I/O	D
PQL_DFILLM	128	128	2	-1	Files	D
PQL_MFILLM	2	2	2	-1	Files	D
PQL_DPGFLQUOTA	700000	700000	512000	-1	Pagelets	D
PQL_MPGFLQUOTA	512000	512000	512000	-1	Pagelets	D
PQL_DPRCLM	32	32	0	-1	Processes	D
PQL_MPRCLM	0	0	0	-1	Processes	D
PQL_DTQELM	16	16	0	-1	Timers	D
PQL_MTQELM	0	0	0	-1	Timers	D
PQL_DWSDEFAULT	53417	32768	16384	-1	Pagelets	
PQL_MWSDEFAULT	53417	16384	16384	-1	Pagelets	
PQL_DWSQUOTA	106834	65536	32768	-1	Pagelets	D
PQL_MWSQUOTA	106834	32768	32768	-1	Pagelets	D
PQL_DWSEXTENT	1619968	131072	65536	-1	Pagelets	D
PQL_MWSEXTENT	1619968	65536	65536	-1	Pagelets	D
PQL_DENQLM	2048	2048	64	-1	Locks	D
PQL_MENQLM	64	64	64	-1	Locks	D
PQL_DJTQUOTA	8192	8192	0	-1	Bytes	D
PQL_MJTQUOTA	0	0	0	-1	Bytes	D

In this display, the values for many PQL_M *quota* parameters increased from the defaults to their current values. Typically, this happens over time when the AUTOGEN feedback is run periodically on your system. The PQL_M *quota* values also can change, of course, when you modify the values in MODPARAMS.DAT or in SYSMAN. If you plan to use a common SYSUAF.DAT in an OpenVMS Cluster, with both Integrity servers and Alpha computers, remember the dynamic nature of the PQL_M *quota* parameters.

10.10.3. Examples

The following table summarizes common SYSUAF.DAT scenarios and probable results on Integrity servers and Alpha computers in an OpenVMS Cluster system.

Table 10.5. Common SYSUAF.DAT Scenarios and Probable Results

WHEN you set values at...	THEN a process that starts on...	Will result in...
Alpha level	An Alpha node	Execution with the values you deemed appropriate.
	Integrity server node	LOGINOUT ignoring the typically lower Integrity server level values in the SYSUAF and instead use the value of each quota's current

WHEN you set values at...	THEN a process that starts on...	Will result in...
		PQL_M <i>quota</i> values on the Alpha system. Monitor the current values of PQL_M <i>quota</i> system parameters if you choose to try this approach. Increase as necessary the appropriate PQL_M <i>quota</i> values on the Alpha system in MODPARAMS.DAT.
Integrity server level	Integrity server node	Execution with the values you deemed appropriate.
	An Alpha node	LOGINOUT ignoring the typically lower Integrity server level values in the SYSUAF and instead use the value of each quota's current PQL_M <i>quota</i> values on the Alpha system. Monitor the current values of PQL_M <i>quota</i> system parameters if you choose to try this approach. Increase as necessary the appropriate PQL_M <i>quota</i> values on the Alpha system in MODPARAMS.DAT.

You might decide to experiment with the higher process-quota values that usually are associated with an OpenVMS Alpha system's SYSUAF.DAT as you determine values for a common SYSUAF.DAT in an OpenVMS Cluster environment. The higher Alpha-level process quotas might be appropriate for processes created on host Integrity server nodes in the OpenVMS Cluster if the Integrity server systems have large available memory resources.

You can determine the values that are appropriate for processes on your Integrity server and Alpha systems by experimentation and modification over time. Factors in your decisions about appropriate limit and quota values for each process will include the following:

- Amount of available memory
- CPU processing power
- Average work load of the applications
- Peak work loads of the applications

10.11. Restoring Cluster Quorum

During the life of an OpenVMS Cluster system, computers join and leave the cluster. For example, you may need to add more computers to the cluster to extend the cluster's processing capabilities, or a computer may shut down because of a hardware or fatal software error. The connection management software coordinates these cluster transitions and controls cluster operation.

When a computer shuts down, the remaining computers, with the help of the connection manager, reconfigure the cluster, excluding the computer that shut down. The cluster can survive the failure of the computer and continue process operations as long as the cluster votes total is greater than the cluster quorum value. If the cluster votes total falls below the cluster quorum value, the cluster suspends the execution of all processes.

10.11.1. Restoring Votes

For process execution to resume, the cluster votes total must be restored to a value greater than or equal to the cluster quorum value. Often, the required votes are added as computers join or rejoin the cluster. However, waiting for a computer to join the cluster and increasing the votes value is not always a simple or convenient remedy. An alternative solution, for example, might be to shut down and reboot all the computers with a reduce quorum value.

After the failure of a computer, you may want to run the Show Cluster utility and examine values for the VOTES, EXPECTED_VOTES, CL_VOTES, and CL_QUORUM fields. (See the *VSI OpenVMS System Management Utilities Reference Manual* for a complete description of these fields.) The VOTES and EXPECTED_VOTES fields show the settings for each cluster member; the CL_VOTES and CL_QUORUM fields show the cluster votes total and the current cluster quorum value.

To examine these values, enter the following commands:

```
$ SHOW CLUSTER/CONTINUOUS
COMMAND> ADD CLUSTER
```

Note: If you want to enter SHOW CLUSTER commands interactively, you must specify the /CONTINUOUS qualifier as part of the SHOW CLUSTER command string. If you do not specify this qualifier, SHOW CLUSTER displays cluster status information returned by the DCL command SHOW CLUSTER and returns you to the DCL command level.

If the display from the Show Cluster utility shows the CL_VOTES value equal to the CL_QUORUM value, the cluster cannot survive the failure of any remaining voting member. If one of these computers shuts down, all process activity in the cluster stops.

10.11.2. Reducing Cluster Quorum Value

To prevent the disruption of cluster process activity, you can reduce the cluster quorum value as described in Table 10.6.

Table 10.6. Reducing the Value of Cluster Quorum

Technique	Description
Use the DCL command SET CLUSTER/EXPECTED_VOTES to adjust the cluster quorum to a value you specify.	<p>If you do not specify a value, the operating system calculates an appropriate value for you. You need to enter the command on only one computer to propagate the new value throughout the cluster. When you enter the command, the operating system reports the new value.</p> <p>Suggestion: Normally, you use the SET CLUSTER/EXPECTED_VOTES command only after a computer has left the cluster for an extended period. (For more information about this command, see the <i>VSI OpenVMS DCL Dictionary</i>).</p> <p>Example: For example, if you want to change expected votes to set the cluster quorum to 2, enter the following command:</p> <pre>\$ SET CLUSTER/EXPECTED_VOTES=3</pre> <p>The resulting value for quorum is $(3 + 2)/2 = 2$.</p> <p>Note: No matter what value you specify for the SETCLUSTER/EXPECTED_VOTES command, you cannot increase quorum to a value that is greater than the number of the votes present, nor can you reduce quorum to a value that is half or fewer of the votes present.</p>

Technique	Description
	When a computer that previously was a cluster member is ready to rejoin, you must reset the EXPECTED_VOTES system parameter to its original value in MODPARAMS.DAT on all computers and then reconfigure the cluster according to the instructions in Section 8.6. You do not need to use the SETCLUSTER/EXPECTED_VOTES command to increase cluster quorum, because the quorum value is increased automatically when the computer rejoins the cluster.
Use the IPC Q command to recalculate the quorum.	Refer to the <i>VSI OpenVMS System Manager's Manual</i> for a description of the Q command.
Select one of the cluster-related shutdown options.	Refer to Section 10.6 for a description of the shutdown options.

10.12. Cluster Performance

Sometimes performance issues involve monitoring and tuning applications and the system as a whole. Tuning involves collecting and reporting on system and network processes to improve performance. A number of tools can help you collect information about an active system and its applications.

10.12.1. Using the SHOW Commands

The following table briefly describes the SHOW commands available with the OpenVMS operating system. Use the SHOW DEVICE commands and qualifiers shown in the table.

Command	Purpose
SHOW DEVICE/FULL	Shows the complete status of a device, including: <ul style="list-style-type: none"> • Whether the disk is available to the cluster • Whether the disk is MSCP served or dual ported • The name and type of the primary and secondary hosts • Whether the disk is mounted on the system where you enter the command • The systems in the cluster on which the disk is mounted
SHOW DEVICE/FILES	Displays a list of the names of all files open on a volume and their associated process name and process identifier (PID). The command: <ul style="list-style-type: none"> • Lists files opened only on this node. • Finds all open files on a disk. You can use either the SHOW DEVICE/FILES command or SYSMAN commands on each node that has the disk mounted.
SHOW DEVICE/SERVED	Displays information about disks served by the MSCP server on the node where you enter the command. Use the following qualifiers to customize the information:

Command	Purpose
	<ul style="list-style-type: none"> • /HOST displays the names of processors that have devices online through the local MSCP server, and the number of devices. • /RESOURCE displays the resources available to the MSCP server, total amount of nonpaged dynamic memory available for I/O buffers, and number of I/O request packets. • /COUNT displays the number of each size and type of I/O operation the MSCP server has performed since it was started. • /ALL displays all of the information listed for the SHOW DEVICE/SERVED command.

The SHOW CLUSTER command displays a variety of information about the OpenVMS Cluster system. The display output provides a view of the cluster as seen from a single node, rather than a complete view of the cluster.

Reference: The *VSI OpenVMS System Management Utilities Reference Manual* contains complete information about all the SHOW commands and the Show Cluster utility.

10.12.2. Using the Monitor Utility

The following table describes using the OpenVMS Monitor utility to locate disk I/O bottlenecks. I/O bottlenecks can cause the OpenVMS Cluster system to appear to hang.

Step	Action
1	<p>To determine which clusterwide disks may be problem disks:</p> <ol style="list-style-type: none"> 1. Create a node-by-node summary of disk I/O using the MONITOR/NODE command 2. Adjust the “row sum” column for MSCP served disks as follows: <ul style="list-style-type: none"> • I/O rate on serving node includes local requests and all requests from other nodes • I/O rate on other nodes includes requests generated from that node • Requests from remote nodes are counted twice in the row sum column 3. Note disks with the row sum more than 8 I/Os per second 4. Eliminate from the list of cluster problem disks the disks that are: <ul style="list-style-type: none"> • Not shared • Dedicated to an application • In the process of being backed up
2	<p>For each node, determine the impact of potential problem disks:</p> <ul style="list-style-type: none"> • If a disproportionate amount of a disk's I/O comes from a particular node, the problem is most likely specific to the node.

Step	Action
	<ul style="list-style-type: none"> • If a disk's I/O is spread evenly over the cluster, the problem may be clusterwide overuse. • If the average queue length for a disk on a given node is less than 0.2, then the disk is having little impact on the node.
3	<p>For each problem disk, determine whether:</p> <ul style="list-style-type: none"> • Page and swap files from any node are on the disk. • Commonly used programs or data files are on the disk (use the SHOWDEVICE/FILES command). • Users with default directories on the disk are causing the problem.

10.12.3. Using VSI Availability Manager

HP Availability Manager is a real-time monitoring, diagnostic, and correction tools used by system managers to improve the availability and throughput of a system. Availability Manager runs on OpenVMS Integrity servers and OpenVMS Alpha servers and on a Windows node.

These products, which are included with the operating system, help system managers correct system resource utilization problems for CPU usage, low memory, lock contention, hung or runaway processes, I/O, disks, page files, and swap files.

Availability Manager enables you to monitor one or more OpenVMS nodes on an extended LAN from either an OpenVMS Alpha or a Windows node. Availability Manager collects system and process data from multiple OpenVMS nodes simultaneously. It analyzes the data and displays the output using a native Java GUI.

DECcmds collects and analyzes data from multiple nodes (VAX and Alpha) simultaneously, directing all output to a centralized DECwindows display. DECcmds helps you observe and troubleshoot availability problems, as follows:

- Alerts users to resource availability problems, suggests paths for further investigation, and recommends actions to improve availability.
- Centralizes management of remote nodes within an extended LAN.
- Allows real-time intervention, including adjustment of node and process parameters, even when remote nodes are hung.
- Adjusts to site-specific requirements through a wide range of customization options.

Reference: For more information about Availability Manager, see the *Availability Manager User's Guide*.

For more information about DECcmds, see the *DECcmds User's Guide*.

10.12.4. Monitoring LAN Activity

It is important to monitor LAN activity on a regular basis. Using the SCACP, you can monitor LAN activity as well as set and show default ports, start and stop LAN devices, and assign priority values to channels.

Reference: For more information about SCACP, see the *VSI OpenVMS System Management Utilities Reference Manual*.

Using NCP commands like the following, you can set up a convenient monitoring procedure to report activity for each 12-hour period. Note that DECnet event logging for event 0.2 (automatic line counters) must be enabled.

Reference: For detailed information on DECnet for OpenVMS event logging, refer to the *VSI OpenVMS DECnet Network Management Utilities* manual.

In these sample commands, BNA-0 is the line ID of the Ethernet line.

```
NCP> DEFINE LINE BNA-0 COUNTER TIMER 43200
NCP> SET LINE BNA-0 COUNTER TIMER 43200
```

At every timer interval (in this case, 12 hours), DECnet will create an event that sends counter data to the DECnet event log. If you experience a performance degradation in your cluster, check the event log for increases in counter values that exceed normal variations for your cluster. If all computers show the same increase, there may be a general problem with your Ethernet configuration. If, on the other hand, only one computer shows a deviation from usual values, there is probably a problem with that computer or with its Ethernet interface device.

The following layered products can be used in conjunction with one of VSI's LAN bridges to monitor the LAN traffic levels: RBMS, DECelms, DECmcc, and LAN Traffic Monitor (LTM).

Note that some of these products are no longer supported by VSI.

10.12.5. LAN or PEDRIVER Fast Path Settings

Save the LAN DEVICE and PEDRIVER, on which SCS communication is enabled, on the same CPU by executing the following command:

```
SET DEVICE EWA/Pref=1
SET DEVICE PEA0/Pref=1
```

If a node uses IP as the interconnect for cluster communication, ensure that LAN, BG, and PE devices are in the same CPU. If the CPU is saturated, set off load devices on to a different CPU.

Appendix A. Cluster System Parameters

For systems to boot properly into a cluster, certain system parameters must be set on each cluster computer. Table A.1 lists system parameters used in cluster configurations.

A.1. Values

Some system parameters are in units of pagelets, whereas others are in pages. AUTOGEN determines the hardware page size and records it in the PARAMS.DAT file.

Caution: When reviewing AUTOGEN recommended values or when setting system parameters with SYSGEN, note carefully which units are required for each parameter.

Table A.1 describes system parameters that are specific to OpenVMS Cluster configurations that may require adjustment in certain configurations. Table A.2 describes OpenVMS Cluster specific system parameters that are reserved for OpenVMS use.

Reference: System parameters, including cluster and volume shadowing system parameters, are documented in the *VSI OpenVMS System Management Utilities Reference Manual*.

Table A.1. Adjustable Cluster System Parameters

Parameter	Description
ALLOCLASS	Specifies a numeric value from 0 to 255 to be assigned as the disk allocation class for the computer. The default value is 0.
CHECK_CLUSTER	Serves as a VAXCLUSTER parameter sanity check. When CHECK_CLUSTER is set to 1, SYSBOOT outputs a warning message and forces a conversational boot if it detects the VAXCLUSTER parameter is set to 0.
CLUSTER_CREDITS	<p>Specifies the number of per-connection buffers a node allocates to receiving VMS\$VAXcluster communications.</p> <p>If the SHOW CLUSTER command displays a high number of credit waits for the VMS\$VAXcluster connection, you might consider increasing the value of CLUSTER_CREDITS on the other node. However, in large cluster configurations, setting this value unnecessarily high will consume a large quantity of nonpaged pool. Each receive buffer is at least SCSMAXMSG bytes in size but might be substantially larger depending on the underlying transport.</p> <p>It is not required that all nodes in the cluster have the same value for CLUSTER_CREDITS. For small or memory-constrained systems, the default value of CLUSTER_CREDITS should be adequate.</p>
CWCREPRC_ENABLE	Controls whether an unprivileged user can create a process on another OpenVMS Cluster node. The default value of 1 allows an unprivileged user to create a detached process with the same UIC on another node. A value of 0 requires that a user have DETACH or CMKRNL privilege to create a process on another node.

Parameter	Description																
DISK_QUORUM	The physical device name, in ASCII, of an optional quorum disk. ASCII spaces indicate that no quorum disk is being used. DISK_QUORUM must be defined on one or more cluster computers capable of having a direct (not MSCP served) connection to the disk. These computers are called quorum disk watchers . The remaining computers (computers with a blank value for DISK_QUORUM) recognize the name defined by the first watcher computer with which they communicate.																
DR_UNIT_BASE	Specifies the base value from which unit numbers for DR devices (StorageWorks RAID Array 200 Family logical RAID drives) are counted. DR_UNIT_BASE provides a way for unique RAID device numbers to be generated. DR devices are numbered starting with the value of DR_UNIT_BASE and then counting from there. For example, setting DR_UNIT_BASE to 10 will produce device names such as \$1\$DRA10, \$1\$DRA11, and so on. Setting DR_UNIT_BASE to appropriate, nonoverlapping values on all cluster members that share the same (nonzero) allocation class will ensure that no two RAID devices are given the same name.																
EXPECTED_VOTES	Specifies a setting that is used to derive the initial quorum value. This setting is the sum of all VOTES held by potential cluster members. By default, the value is 1. The connection manager sets a quorum value to a number that will prevent cluster partitioning (see Section 2.3). To calculate quorum, the system uses the following formula: $\text{estimated quorum} = (\text{EXPECTED_VOTES} + 2) / 2$																
LAN_FLAGS (Integrity servers and Alpha)	LAN_FLAGS is a bit mask used to enable features in the local area networks port drivers and support code. The default value for LAN_FLAGS is 0. The bit definitions as follows: <table border="1" data-bbox="443 1317 1355 2038"> <thead> <tr> <th>Bit</th> <th>Description</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>The default value of zero indicates that ATM devices run in the SONET mode. If set to 1, this bit indicates that ATM devices run in the SDH mode.</td> </tr> <tr> <td>1</td> <td>If set, this bit enables a subset of the ATM trace and debug messages in the LAN port drivers and support code.</td> </tr> <tr> <td>2</td> <td>If set, this bit enables all ATM trace and debug messages in the LAN port drivers and support code.</td> </tr> <tr> <td>3</td> <td>If set, this bit runs UNI 3.0 over all ATM adapters. Auto-sensing of the ATM UNI version is enabled if both bit 3 and bit 4 are off (0).</td> </tr> <tr> <td>4</td> <td>If set, this bit runs UNI 3.1 over all ATM adapters. Auto-sensing of the ATM UNI version is enabled if both bit 3 and bit 4 are off (0).</td> </tr> <tr> <td>5</td> <td>If set, this bit disables auto-negotiation over all Gigabit Ethernet Adapters.</td> </tr> <tr> <td>6</td> <td>If set, this bit enables the use of jumbo frames over all Gigabit Ethernet Adapters.</td> </tr> </tbody> </table>	Bit	Description	0	The default value of zero indicates that ATM devices run in the SONET mode. If set to 1, this bit indicates that ATM devices run in the SDH mode.	1	If set, this bit enables a subset of the ATM trace and debug messages in the LAN port drivers and support code.	2	If set, this bit enables all ATM trace and debug messages in the LAN port drivers and support code.	3	If set, this bit runs UNI 3.0 over all ATM adapters. Auto-sensing of the ATM UNI version is enabled if both bit 3 and bit 4 are off (0).	4	If set, this bit runs UNI 3.1 over all ATM adapters. Auto-sensing of the ATM UNI version is enabled if both bit 3 and bit 4 are off (0).	5	If set, this bit disables auto-negotiation over all Gigabit Ethernet Adapters.	6	If set, this bit enables the use of jumbo frames over all Gigabit Ethernet Adapters.
Bit	Description																
0	The default value of zero indicates that ATM devices run in the SONET mode. If set to 1, this bit indicates that ATM devices run in the SDH mode.																
1	If set, this bit enables a subset of the ATM trace and debug messages in the LAN port drivers and support code.																
2	If set, this bit enables all ATM trace and debug messages in the LAN port drivers and support code.																
3	If set, this bit runs UNI 3.0 over all ATM adapters. Auto-sensing of the ATM UNI version is enabled if both bit 3 and bit 4 are off (0).																
4	If set, this bit runs UNI 3.1 over all ATM adapters. Auto-sensing of the ATM UNI version is enabled if both bit 3 and bit 4 are off (0).																
5	If set, this bit disables auto-negotiation over all Gigabit Ethernet Adapters.																
6	If set, this bit enables the use of jumbo frames over all Gigabit Ethernet Adapters.																

Parameter	Description
	7 Reserved.
	8 If set, this bit disables the use of flow control over all LAN adapters that support flow control.
	9 Reserved.
	10 Reserved.
	11 If set, this bit disables the logging of error log entries by LAN drivers.
	12 If set, this bit enables a fast timeout on transmit requests, usually between 1 and 1.2 seconds instead of 3 to 4seconds for most LAN drivers.
	13 If set, this bit transmits that are given to the LAN device and never completed by the device (transmit timeout condition) are completed with error status (SS\$_ABORT) rather than success status (SS\$_NORMAL).
LOCKDIRWT	Lock manager directory system weight. Determines the portion of lock manager directory to be handled by this system. The default value is adequate for most systems.
LOCKRMWT	Lock manager remaster weight. This parameter, in conjunction with the lock remaster weight from a remote node, determines the level of activity necessary for remastering a lock tree.
MC_SERVICES_P0 (dynamic)	<p>Controls whether other MEMORY CHANNEL nodes in the cluster continue to run if this node bugchecks or shuts down.</p> <p>A value of 1 causes other nodes in the MEMORY CHANNEL cluster to fail with bugcheck code MC_FORCED_CRASH if this node bug checks or shuts down.</p> <p>The default value is 0. A setting of 1 is intended only for debugging purposes; the parameter should otherwise be left at its default state.</p>
MC_SERVICES_P2 (static)	<p>Specifies whether to load the PMDRIVER (PMA0) MEMORY CHANNEL cluster port driver. PMDRIVER is a new driver that serves as the MEMORY CHANNEL cluster port driver. It works together with MCDRIVER (the MEMORY CHANNEL device driver and device interface)to provide MEMORY CHANNEL clustering. If PMDRIVER is not loaded, cluster connections will not be made over the MEMORY CHANNEL interconnect.</p> <p>The default for MC_SERVICES_P2 is 1. This default value causes PMDRIVER to be loaded when you boot the system.</p> <p>VSI recommends that this value not be changed. This parameter value must be the same on all nodes connected by MEMORY CHANNEL.</p>
MC_SERVICES_P3 (dynamic)	<p>Specifies the maximum number of tags supported. The maximum value is 2048 and the minimum value is 100.</p> <p>The default value is 800. VSI recommends that this value not be changed.</p>

Parameter	Description								
	This parameter value must be the same on all nodes connected by MEMORY CHANNEL.								
MC_SERVICES_P4 (static)	<p>Specifies the maximum number of regions supported. The maximum value is 4096 and the minimum value is 100.</p> <p>The default value is 200. VSI recommends that this value not be changed.</p> <p>This parameter value must be the same on all nodes connected by MEMORY CHANNEL.</p>								
MC_SERVICES_P6 (static)	<p>Specifies MEMORY CHANNEL message size, the body of an entry in a free queue, or a work queue. The maximum value is 65536 and the minimum value is 544. The default value is 992, which is suitable in all cases except systems with highly constrained memory.</p> <p>For such systems, you can reduce the memory consumption of MEMORY CHANNEL by slightly reducing the default value of 992. This value must always be equal to or greater than the result of the following calculation:</p> <ol style="list-style-type: none"> 1. Select the larger of SCS_MAXMSG and SCS_MAXDG. 2. Round that value to the next quadword. <p>This parameter value must be the same on all nodes connected by MEMORY CHANNEL.</p>								
MC_SERVICES_P7 (dynamic)	<p>Specifies whether to suppress or display messages about cluster activities on this node. Can be set to a value of 0, 1, or 2. The meanings of these values are:</p> <table border="1"> <thead> <tr> <th>Value</th> <th>Meaning</th> </tr> </thead> <tbody> <tr> <td>0</td> <td>Nonverbose mode – no informational messages will appear on the console or in the error log.</td> </tr> <tr> <td>1</td> <td>Same as verbose mode plus PMDRIVER stalling and recovery messages.</td> </tr> <tr> <td>2</td> <td>Same as verbose mode plus PMDRIVER stalling and recovery messages.</td> </tr> </tbody> </table> <p>The default value is 0. VSI recommends that this value not be changed except for debugging MEMORY CHANNEL problems or adjusting the MC_SERVICES_P9 parameter.</p>	Value	Meaning	0	Nonverbose mode – no informational messages will appear on the console or in the error log.	1	Same as verbose mode plus PMDRIVER stalling and recovery messages.	2	Same as verbose mode plus PMDRIVER stalling and recovery messages.
Value	Meaning								
0	Nonverbose mode – no informational messages will appear on the console or in the error log.								
1	Same as verbose mode plus PMDRIVER stalling and recovery messages.								
2	Same as verbose mode plus PMDRIVER stalling and recovery messages.								
MC_SERVICES_P9 (static)	<p>Specifies the number of initial entries in a single channel's free queue. The maximum value is 2048 and the minimum value is 10.</p> <p>Note that MC_SERVICES_P9 is not a dynamic parameter; you must reboot the system after each change in order for the change to take effect.</p> <p>The default value is 150. VSI recommends that this value not be changed.</p> <p>This parameter value must be the same on all nodes connected by MEMORY CHANNEL.</p>								
MPDEV_AFB_INTVL (disks only)	Specifies the automatic failback interval in seconds. The automatic failback interval is the minimum number of seconds that must elapse before the								

Parameter	Description
	<p>system will attempt another failback from an MSCP path to a direct path on the same device.</p> <p>MPDEV_POLLER must be set to ON to enable automatic failback. You can disable automatic failback without disabling the poller by setting MPDEV_AFB_INTVL to 0. The default is 300 seconds.</p>
MPDEV_D1 (disks only)	Reserved for use by the operating system.
MPDEV_D2 (disks only)	Reserved for use by the operating system.
MPDEV_D3 (disks only)	Reserved for use by the operating system.
MPDEV_D4 (disks only)	Reserved for use by the operating system.
MPDEV_ENABLE	<p>Enables the formation of multipath sets when set to ON (1). When set to OFF (0), the formation of additional multipath sets and the addition of new paths to existing multipath sets is disabled. However, existing multipath sets remain in effect. The default is ON.</p> <p>MPDEV_REMOTE and MPDEV_AFB_INTVL have no effect when MPDEV_ENABLE is set to OFF.</p>
MPDEV_LCRETRIES (disks only)	Controls the number of times the system retries the direct paths to the controller that the logical unit is online to, before moving on to direct paths to the other controller, or to an MSCP served path to the device. The valid range for retries is 1 through 256. The default is 1.
MPDEV_POLLER	<p>Enables polling of the paths to multipath set members when set to ON (1). Polling allows early detection of errors on inactive paths. If a path becomes unavailable or returns to service, the system manager is notified with an OPCOM message. When set to OFF (0), multipath polling is disabled. The default is ON. Note that this parameter must be set to ON to use the automatic failback feature.</p>
MPDEV_REMOTE (disks only)	<p>Enables MSCP served disks to become members of a multipath set when set to ON (1). When set to OFF (0), only local paths to a SCSI or Fibre Channel device are used in the formation of additional multipath sets. MPDEV_REMOTE is enabled by default. However, setting this parameter to OFF has no effect on existing multipath sets that have remote paths.</p> <p>To use multipath failover to a served path, MPDEV_REMOTE must be enabled on all systems that have direct access to shared SCSI/Fibre Channel devices. The first release to provide this feature is OpenVMS Alpha Version 7.3-1. Therefore, all nodes on which MPDEV_REMOTE is enabled must be running OpenVMS Alpha Version 7.3-1 (or later).</p> <p>If MPDEV_ENABLE is set to OFF (0), the setting of MPDEV_REMOTE has no effect because the addition of all new paths to multipath sets is disabled. The default is ON.</p>
MSCP_BUFFER	This buffer area is the space used by the server to transfer data between client systems and local disks.

Parameter	Description
	<p>On VAX systems, MSCP_BUFFER specifies the number of pages to be allocated to the MSCP server's local buffer area.</p> <p>MSCP_BUFFER specifies the number of pagelets to be allocated the MSCP server's local buffer area.</p>
MSCP_CMD_TMO	<p>Specifies the time in seconds that the OpenVMS MSCP server uses to detect MSCP command timeouts. The MSCP server must complete the command within a built-in time of approximately 40 seconds plus the value of the MSCP_CMD_TMO parameter.</p> <p>An MSCP_CMD_TMO value of 0 is normally adequate. A value of 0 provides the same behavior as in previous releases of OpenVMS (which did not have an MSCP_CMD_TMO system parameter). A nonzero setting increases the amount of time before an MSCP command times out.</p> <p>If command timeout errors are being logged on client nodes, setting the parameter to a nonzero value on OpenVMS servers reduces the number of errors logged. Increasing the value of this parameter reduces the number of client MSCP command timeouts and increases the time it takes to detect faulty devices.</p> <p>If you need to decrease the number of command timeout errors, set an initial value of 60. If timeout errors continue to be logged, you can increase this value in increments of 20 seconds.</p>
MSCP_CREDITS	<p>Specifies the number of outstanding I/O requests that can be active from one client system.</p>
MSCP_LOAD	<p>Controls whether the MSCP server is loaded. Specify 1 to load the server, and use the default CPU load rating. A value greater than 1 loads the server and uses this value as a constant load rating. By default, the value is set to 0 and the server is not loaded.</p>
MSCP_SERVE_ALL	<p>Controls the serving of disks. The settings take effect when the system boots. You cannot change the settings when the system is running.</p> <p>Starting with OpenVMS Version 7.2, the serving types are implemented as a bit mask. To specify the type of serving your system will perform, locate the type you want in the following table and specify its value. For some systems, you may want to specify two serving types, such as serving the system disk and serving locally attached disks. To specify such a combination, add the values of each type, and specify the sum.</p> <p>In a mixed-version cluster that includes any systems running OpenVMS Version 7.1-x or earlier, serving all available disks is restricted to serving all disks except those whose allocation class does not match the system's node allocation class (pre-Version 7.2 meaning). To specify this type of serving, use the value 9 (which sets bit 0 and bit 3).</p> <p>The following table describes the serving type controlled by each bit and its decimal value.</p>

Parameter	Description	
	Bit and Value When Set	Description
	Bit 0 (1)	Serve all available disks (locally attached and those connected to HS <i>x</i> and DSSI controllers). Disks with allocation classes that differ from the system's allocation class (set by the ALLOCLASS parameter) are also served if bit 3 is not set.
	Bit 1 (2)	Serve locally attached (non-HS <i>x</i> and non-DSSI) disks.
	Bit 2 (4)	Serve the system disk. This is the default setting. This setting is important when other nodes in the cluster rely on this system being able to serve its system disk. This setting prevents obscure contention problems that can occur when a system attempts to complete I/O to a remote system disk whose system has failed.
	Bit 3 (8)	Restrict the serving specified by bit 0. All disks except those with allocation classes that differ from the system's allocation class (set by the ALLOCLASS parameter) are served. This is pre-Version 7.2 behavior. If your cluster includes systems running Open 7.1- <i>x</i> or earlier, and you want to serve all available disks, you must specify 9, the result of setting this bit and bit 0.
	Bit 4 (15)	By default, the bit 4 is not set, hence the DUDRIVER will accept the devices with unit number greater than 9999. On the client side, if bit 4 is set (10000 binary) in the MSCP_SERVE_ALL parameter, the client will reject devices with unit number greater than 9999 and retains the earlier behavior.
	<p>Although the serving types are now implemented as a bit mask, the values of 0, 1, and 2, specified by bit 0 and bit 1, retain their original meanings:</p> <ul style="list-style-type: none"> • 0 — Do not serve any disks (the default for earlier versions of OpenVMS). • 1 — Serve all available disks. • 2 — Serve only locally attached (non-HS <i>x</i> and non-DSSI) disks. <p>If the MSCP_LOAD system parameter is 0, MSCP_SERVE_ALL is ignored. For more information about this system parameter, see Section 6.3.1.</p>	
NISCS_CONV_BOOT	During booting as an OpenVMS Cluster satellite, specifies whether conversational bootstraps are enabled on the computer. The default value of 0 specifies that conversational bootstraps are disabled. A value of 1 enables conversational bootstraps.	
NISCS_LAN_OVRHD	Starting with OpenVMS Version 7.3, this parameter is obsolete. This parameter was formerly provided to reserve space in a LAN packet for	

Parameter	Description
	<p>encryption fields applied by external encryption devices. PEDRIVER now automatically determines the maximum packet size a LAN path can deliver, including any packet-size reductions required by external encryption devices.</p>
NISCS_LOAD_PEA0	<p>Specifies whether the port driver (PEDRIVER) must be loaded to enable cluster communications over the local area network (LAN) or IP. The default value of 0 specifies that the driver is not loaded.</p> <p>Caution: If the NISCS_LOAD_PEA0 parameter is set to 1, the VAXCLUSTER system parameter must be set to 2. This ensures coordinated access to shared resources in the OpenVMS Cluster and prevents accidental data corruption.</p>
NISCS_MAX_PKTSZ	<p>Specifies an upper limit, in bytes, on the size of the user data area in the largest packet sent by NISCA on any LAN network.</p> <p>NISCS_MAX_PKTSZ allows the system manager to change the packet size used for cluster communications on network communication paths. PEDRIVER automatically allocates memory to support the largest packet size that is usable by any virtual circuit connected to the system up to the limit set by this parameter. Its default values are different for OpenVMS Integrity servers and Alpha.</p> <p>On Integrity servers and Alpha, to optimize performance, the default value is the largest packet size currently supported by OpenVMS.</p> <p>PEDRIVER uses NISCS_MAX_PKTSZ to compute the maximum amount of data to transmit in any LAN or IP packet:</p> <p>LAN packet size \leq LAN header (padded Ethernet format)</p> <ul style="list-style-type: none"> • NISCS_MAX_PKTSZ • NISCS checksum (only if data checking is enabled) • LAN CRC or FCS <p>The actual packet size automatically used by PEDRIVER might be smaller than the NISCS_MAX_PKTSZ limit for any of the following reasons:</p> <ul style="list-style-type: none"> • On a per-LAN-path basis, if PEDRIVER determines that the LAN path between two nodes, including the local and remote LAN adapters and intervening LAN equipment, can convey only a lesser size. <p>In other words, only nodes with large-packet LAN adapters connected end-to-end by large-packet LAN equipment can use large packets. Nodes connected to large-packet LANs but having an end-to-end path that involves an Ethernet segment restrict packet size to that of an Ethernet packet 1498 bytes.</p> <ul style="list-style-type: none"> • For performance reasons, PEDRIVER might further limit the upper bound on packet size so that the packets can be allocated from a look aside list in the nonpaged pool.

Parameter	Description								
	<p>The actual memory allocation includes the required data structure overhead used by PEDRIVER and the LAN drivers, in addition to the actual LAN packet size.</p> <p>The following table shows the minimum NISCS_MAX_PKTSZ value required to use the maximum packet size supported by specified LAN types.</p> <table border="1" data-bbox="544 488 1449 741"> <thead> <tr> <th data-bbox="544 488 719 530">Type of LAN</th> <th data-bbox="724 488 1449 530">Minimum Value for NISCS_MAX_PKTSZ</th> </tr> </thead> <tbody> <tr> <td data-bbox="544 537 719 580">Ethernet</td> <td data-bbox="724 537 1449 580">1498</td> </tr> <tr> <td data-bbox="544 586 719 658">Gigabit Ethernet</td> <td data-bbox="724 586 1449 658">8192</td> </tr> <tr> <td data-bbox="544 665 719 741">10 Gigabit Ethernet</td> <td data-bbox="724 665 1449 741">8192</td> </tr> </tbody> </table> <p>Note that the maximum packet size for some Gigabit Ethernet adapters is larger than the maximum value of NISCS_MAX_PKTSZ (8192 bytes). For information on how to enable jumbo frames on Gigabit Ethernet (packet sizes larger than those noted for Ethernet), see the LAN_FLAGS parameter.</p> <p>OpenVMS Alpha Version 7.3-2 or later supports the DEGXA Gigabit Ethernet adapter, which is a BroadcomBCM5703 chip (TIGON3) network interface card (NIC). The introduction of the DEGXA Gigabit Ethernet adapter continues the existing Gigabit Ethernet support as both a LAN device as well as a cluster interconnect device.</p> <p>Note that starting with OpenVMS Version 8.4, OpenVMS can use HP TCP/IP services for cluster communications using UDP protocol. NISCS_MAX_PKTSZ will only affect the LAN channel payload size. To affect the IP channel payload size see NISCS_UDP_PKTSZ system parameter.</p>	Type of LAN	Minimum Value for NISCS_MAX_PKTSZ	Ethernet	1498	Gigabit Ethernet	8192	10 Gigabit Ethernet	8192
Type of LAN	Minimum Value for NISCS_MAX_PKTSZ								
Ethernet	1498								
Gigabit Ethernet	8192								
10 Gigabit Ethernet	8192								
NISCS_PORT_SERV	<p>NISCS_PORT_SERV provides flag bits for PEDRIVER port services. Setting bits 0 and 1 (decimal value 3) enables data checking. The remaining bits are reserved for future use. Starting with OpenVMS Version 7.3-1, you can use the SCACP command SET VC/CHECKSUMMING to specify data checking on the VCs to certain nodes. You can do this on a running system. (Refer to the SCACP documentation in the <i>VSI OpenVMS System Management Utilities Reference Manual</i> for more information).</p> <p>On the other hand, changing the setting of NISCS_PORT_SERV requires a reboot. Furthermore, this parameter applies to all virtual circuits between the node on which it is set and other nodes in the cluster.</p> <p>NISCS_PORT_SERV has the AUTOGEN attribute.</p> <p>NISCS_PORT_SERV can be used for enabling PEdriver data compression. The SCACP SET VC command now includes a /COMPRESSION (or/NOCOMPRESSION) qualifier, which enables or disables sending compressed data by the specified PEdriver VCs. The default is /NOCOMPRESSION.</p>								

Parameter	Description
	<p>You can also enable the VC use of compression by setting bit 3 of the NISCS_PORT_SERV system parameter. The /NOCOMPRESSION qualifier does not override compression enabled by setting bit 2 of NISCS_PORT_SERV. For more information, see the SCACP utility chapter, and NISCS_PORT_SERV in the <i>HP OpenVMS System Management Utilities Reference Manual</i> and the <i>HP OpenVMS Availability Manager User's Guide</i>.</p>
NISCS_UDP_PKTSZ	<p>This parameter specifies an upper limit on the size, in bytes, of the user data area in the largest packet sent by NISCA on any IP network.</p> <p>NISCS_UDP_PKTSZ allows the system manager to change the packet size used for cluster communications over IP on network communication paths.</p> <p>PEDRIVER uses NISCS_UDP_PKTSZ to compute the maximum amount of data to transmit in any packet.</p> <p>Currently, the maximum payload over an IP channel is defined by one of the following three parameters. The least of the three values will be in effect.</p> <ul style="list-style-type: none"> • NISCS_UDP_PKTSZ • 1500 bytes • IP_MTU of the interface supported by TCP/IP stack <p>Note that this parameter only affects the IP channel payload and not the LAN channel payload. The LAN channel payload is controlled by NISCS_MAX_PKTSZ.</p>
NISCS_USE_UDP	<p>If NISCS_USE_UDP is set to 1, the PEdriver uses IP in addition to the LAN driver for cluster communication. The bit setting of 1 loads the IPCI configuration information in the configuration files, which are loaded during the boot sequence. SYSSYSTEM:PE\$IP_CONFIG.DAT and SYSSYSTEM:TCPIPCLUSTER.DAT are the two configuration files used for IP Cluster interconnect.</p>
PASTDGBUF	<p>Specifies the number of datagram receive buffers to queue initially for the cluster port driver's configuration poller. The initial value is expanded during system operation, if needed.</p> <p>MEMORY CHANNEL devices ignore this parameter.</p>
QDSKINTERVAL	<p>Specifies, in seconds, the disk quorum polling interval. The maximum is 32767, the minimum is 1, and the default is 3. Lower values trade increased overhead cost for greater responsiveness.</p> <p>This parameter should be set to the same value on each cluster computer.</p>
QDSKVOTES	<p>Specifies the number of votes contributed to the cluster votes total by a quorum disk. The maximum is 127, the minimum is 0, and the default is 1. This parameter is used only when DISK_QUORUM is defined.</p>
RECNXINTERVAL	<p>Specifies, in seconds, the interval during which the connection manager attempts to reconnect a broken connection to an other computer. If a new</p>

Parameter	Description
	<p>connection cannot be established during this period, the connection is declared irrevocably broken, and either this computer or the other must leave the cluster. This parameter trades faster response to certain types of system failures for the ability to survive transient faults of increasing duration.</p> <p>This parameter should be set to the same value on each cluster computer. This parameter also affects the tolerance of the OpenVMS Cluster system for LAN bridge failures (see Section 3.2.10).</p>
SCSBUFFCNT	<p>On Integrity server systems and Alpha systems, the SCS buffers are allocated as needed, and SCSBUFFCNT is reserved for OpenVMS use, only.</p>
SCSCONNCNT	<p>The initial number of SCS connections that are configured for use by all system applications, including the one used by Directory Service Listen. The initial number will be expanded by the system if needed.</p> <p>If no SCS ports are configured on your system, this parameter is ignored. The default value is adequate for all SCS hardware combinations.</p> <p>Note: AUTOGEN provides feedback for this parameter on VAX systems only.</p>
SCSNODE ¹	<p>Specifies the name of the computer. This parameter is not dynamic.</p> <p>Specify SCSNODE as a string of up to six characters. Enclose the string in quotation marks.</p> <p>If the computer is in an OpenVMS Cluster, specify a value that is unique within the cluster. Do not specify the null string.</p> <p>If the computer is running DECnet for OpenVMS, the value must be the same as the DECnet node name.</p>
SCSRESPCNT	<p>SCSRESPCNT is the total number of response descriptor table entries (RDTEs) configured for use by all system applications.</p> <p>If no SCS or DSA port is configured on your system, this parameter is ignored.</p>
SCSSYSTEMID ¹	<p>Specifies a number that identifies the computer. This parameter is not dynamic. SCSSYSTEMID is the low-order 32 bits of the 48-bit system identification number.</p> <p>If the computer is in an OpenVMS Cluster, specify a value that is unique within the cluster.</p> <p>If the computer is running DECnet for OpenVMS, calculate the value from the DECnet address using the following formula:</p> $\text{SCSSYSTEMID} = (\text{DECnet-area-number} * 1024) + \text{DECnet-node-number}$ <p>Example: If the DECnet address is 2.211, calculate the value as follows:</p> $\text{SCSSYSTEMID} = (2 * 1024) + 211 = 2259$

Parameter	Description															
SCSSYSTEMIDH	Specifies the high-order 16 bits of the 48-bit system identification number. This parameter must be set to 0. It is reserved by OpenVMS for future use.															
TAPE_ALLOCLASS	Specifies a numeric value from 0 to 255 to be assigned as the tape allocation class for tape devices connected to the computer. The default value is 0.															
TIMVCFAIL	Specifies the time required for a virtual circuit failure to be detected. VSI recommends that you use the default value. VSI further recommends that you decrease this value only in OpenVMS Cluster systems of three or fewer CPUs, use the same value on each computer in the cluster, and use dedicated LAN segments for cluster I/O.															
TMSCP_LOAD	Controls whether the TMSCP server is loaded. Specify a value of 1 to load the server and set all available TMSCP tapes served. By default, the value is set to 0, and the server is not loaded.															
TMSCP_SERVE_ALL	<p>Controls the serving of tapes. The settings take effect when the system boots. You cannot change the settings when the system is running.</p> <p>Starting with OpenVMS Version 7.2, the serving types are implemented as a bit mask. To specify the type of serving your system will perform, locate the type you want in the following table and specify its value. For some systems, you may want to specify two serving types, such as serving all tapes except those whose allocation class does not match. To specify such a combination, add the values of each type, and specify the sum.</p> <p>In a mixed-version cluster that includes any systems running OpenVMS Version 7.1- <i>x</i> or earlier, serving all available tapes is restricted to serving all tapes except those whose allocation class does not match the system's allocation class (pre-Version 7.2 meaning). To specify this type of serving, use the value 9, which sets bit 0 and bit 3.</p> <p>The following table describes the serving type controlled by each bit and its decimal value.</p> <table border="1"> <thead> <tr> <th>Bit</th> <th>Value When Set</th> <th>Description</th> </tr> </thead> <tbody> <tr> <td>Bit 0</td> <td>1</td> <td>Serve all available tapes (locally attached and those connected to HS <i>x</i> and DSSI controllers). Tapes with allocation classes that differ from the system's allocation class (set by the ALLOCLASS parameter) are also served if bit 3 is not set.</td> </tr> <tr> <td>Bit 1</td> <td>2</td> <td>Serve locally attached (non-HS <i>x</i> and non-DSSI) tapes.</td> </tr> <tr> <td>Bit 2</td> <td>n/a</td> <td>Reserved.</td> </tr> <tr> <td>Bit 3</td> <td>8</td> <td>Restrict the serving specified by bit 0. All tapes except those with allocation classes that differ from the system's allocation class (set by the ALLOCLASS parameter) are served.</td> </tr> </tbody> </table>	Bit	Value When Set	Description	Bit 0	1	Serve all available tapes (locally attached and those connected to HS <i>x</i> and DSSI controllers). Tapes with allocation classes that differ from the system's allocation class (set by the ALLOCLASS parameter) are also served if bit 3 is not set.	Bit 1	2	Serve locally attached (non-HS <i>x</i> and non-DSSI) tapes.	Bit 2	n/a	Reserved.	Bit 3	8	Restrict the serving specified by bit 0. All tapes except those with allocation classes that differ from the system's allocation class (set by the ALLOCLASS parameter) are served.
Bit	Value When Set	Description														
Bit 0	1	Serve all available tapes (locally attached and those connected to HS <i>x</i> and DSSI controllers). Tapes with allocation classes that differ from the system's allocation class (set by the ALLOCLASS parameter) are also served if bit 3 is not set.														
Bit 1	2	Serve locally attached (non-HS <i>x</i> and non-DSSI) tapes.														
Bit 2	n/a	Reserved.														
Bit 3	8	Restrict the serving specified by bit 0. All tapes except those with allocation classes that differ from the system's allocation class (set by the ALLOCLASS parameter) are served.														

Parameter	Description	
		This is pre-Version 7.2 behavior. If your cluster includes systems running OpenVMS Version 7.1- <i>x</i> or earlier, and you want to serve all available tapes, you must specify 9, the result of setting this bit and bit 0.
	Bit 4	15 By default, the bit 4 is not set, hence the TUDRIVER will accept the devices with unit number greater than 9999. On the client side, if bit 4 is set (10000 binary) in the TMSCP_SERVE_ALL parameter, the client will reject devices with unit number greater than 9999 and retains the earlier behavior.
	<p>Although the serving types are now implemented as a bit mask, the values of 0, 1, and 2, specified by bit 0 and bit 1, retain their original meanings:</p> <ul style="list-style-type: none"> • 0 — Do not serve any disks (the default for earlier versions of OpenVMS). • 1 — Serve all available disks. • 2 — Serve only locally attached (non-HS <i>x</i> and non-DSSI) disks. <p>If the TMSCP_LOAD system parameter is 0, TMSCP_SERVE_ALL is ignored.</p>	
VAXCLUSTER	<p>Controls whether the computer should join or form a cluster. This parameter accepts the following three values:</p> <ul style="list-style-type: none"> • 0 — Specifies that the computer will not participate in a cluster. • 1 — Specifies that the computer should participate in a cluster if hardware supporting SCS (CI or DSSI) is present or if NISCS_LOAD_PEA0 is set to 1, indicating that cluster communications is enabled over the local area network (LAN) or IP. • 2 — Specifies that the computer should participate in a cluster. <p>You should always set this parameter to 2 on computers intended to run in a cluster, to 0 on computers that boot from a UDA disk controller and are not intended to be part of a cluster, and to 1 (the default) otherwise.</p> <p>Caution: If the NISCS_LOAD_PEA0 system parameter is set to 1, the VAXCLUSTER parameter must be set to 2. This ensures coordinated access to shared resources in the OpenVMS Cluster system and prevents accidental data corruption. Data corruption may occur on shared resources if the NISCS_LOAD_PEA0 parameter is set to 1 and the VAXCLUSTER parameter is set to 0.</p>	
VOTES	Specifies the number of votes toward a quorum to be contributed by the computer. The default is 1.	

¹Once a computer has been recognized by another computer in the cluster, you cannot change the SCSSYSTEMID or SCSNODE parameter without either changing both or rebooting the entire cluster.

Table A.2 lists system parameters that should not require adjustment at any time. These parameters are provided for use in system debugging. VSI recommends that you do not change these parameters unless you are advised to do so by your VSI support representative. Incorrect adjustment of these parameters can result in cluster failures.

Table A.2. Cluster System Parameters Reserved for OpenVMS Use Only (Integrity servers and Alpha)

Parameter	Description																				
MC_SERVICES_P1 (dynamic)	The value of this parameter must be the same on all nodes connected by MEMORYCHANNEL.																				
MC_SERVICES_P5 (dynamic)	This parameter must remain at the default value of 8000000. This parameter value must be the same on all nodes connected by MEMORY CHANNEL.																				
MC_SERVICES_P8 (static)	This parameter must remain at the default value of 0. This parameter value must be the same on all nodes connected by MEMORY CHANNEL.																				
MPDEV_D1	A multipath system parameter.																				
PE4	<p>PE4 SYSGEN parameter can be used to tune the important parameters of PEDRIVER driver. The PE4 value comprises of the following parameters:</p> <table border="1"> <thead> <tr> <th>Parameter</th> <th>PE4 Bits</th> <th>Default</th> <th>Units</th> </tr> </thead> <tbody> <tr> <td>Listen Timeout</td> <td><7 : 0></td> <td>8</td> <td>Seconds</td> </tr> <tr> <td>HELLO Interval</td> <td><15 : 8></td> <td>30</td> <td>0.1 Sec (100ms)</td> </tr> <tr> <td>CC Ticks/Second</td> <td><23 : 16></td> <td>50</td> <td></td> </tr> <tr> <td>Piggyback Ack Delay</td> <td><31 : 24></td> <td>10</td> <td>0.01 Sec (10ms)</td> </tr> </tbody> </table> <p>VSI recommends to retain the default values for these parameters. Any changes to these parameters should be done with the guidance of VSI support.</p>	Parameter	PE4 Bits	Default	Units	Listen Timeout	<7 : 0>	8	Seconds	HELLO Interval	<15 : 8>	30	0.1 Sec (100ms)	CC Ticks/Second	<23 : 16>	50		Piggyback Ack Delay	<31 : 24>	10	0.01 Sec (10ms)
Parameter	PE4 Bits	Default	Units																		
Listen Timeout	<7 : 0>	8	Seconds																		
HELLO Interval	<15 : 8>	30	0.1 Sec (100ms)																		
CC Ticks/Second	<23 : 16>	50																			
Piggyback Ack Delay	<31 : 24>	10	0.01 Sec (10ms)																		
PRCPOLINTERVAL	<p>Specifies, in seconds, the polling interval used to look for SCS applications, such as the connection manager and MSCP disks, on other computers. Each computer is polled, at most, once each interval.</p> <p>This parameter trades polling overhead against quick recognition of new computers or servers as they appear.</p>																				
SCSMAXMSG	<p>The maximum number of bytes of system application data in one sequenced message. The amount of physical memory consumed by one message is SCSMAXMSG plus the overhead for buffer management.</p> <p>If an SCS port is not configured on your system, this parameter is ignored.</p>																				
SCSMAXDG	<p>Specifies the maximum number of bytes of application data in one datagram.</p> <p>If an SCS port is not configured on your system, this parameter is ignored.</p>																				

Parameter	Description
SCSFLOWCUSH	<p data-bbox="609 246 1410 521">Specifies the lower limit for receive buffers at which point SCS starts to notify the remote SCS of new receive buffers. For each connection, SCS tracks the number of receive buffers available. SCS communicates this number to the SCS at the remote end of the connection. However, SCS does not need to do this for each new receive buffer added. Instead, SCS notifies the remote SCS of new receive buffers if the number of receive buffers falls as low as the SCSFLOWCUSH value.</p> <p data-bbox="609 555 1385 620">If an SCS port is not configured on your system, this parameter is ignored.</p>

Appendix B. Building Common Files

This appendix provides guidelines for building a common user authorization file (UAF) from computer-specific files. It also describes merging RIGHTS.LIST.DAT files.

For more detailed information about how to set up a computer-specific authorization file, see the descriptions in the *VSI OpenVMS Guide to System Security*.

B.1. Building a Common SYSUAF.DAT File

To build a common SYSUAF.DAT file, follow the steps in Table B.1.

Table B.1. Building a Common SYSUAF.DAT File

Step	Action
1	<p>Print a listing of SYSUAF.DAT on each computer. To print this listing, invoke AUTHORIZE and specify the AUTHORIZE command LIST as follows:</p> <pre>\$ SET DEF SYS\$SYSTEM \$ RUN AUTHORIZE UAF> LIST/FULL [*,*]</pre>
2	<p>Use the listings to compare the accounts from each computer. On the listings, mark any necessary changes. For example:</p> <ul style="list-style-type: none">• Delete any accounts that you no longer need.• Make sure that UICs are set appropriately:<ul style="list-style-type: none">• User UICs<p>Check each user account in the cluster to see whether it should have a unique user identification code UIC. For example, OpenVMS Cluster member VENUS may have a user account JONES that has the same UIC as user account SMITH on computer MARS. When computers VENUS and MARS are joined to form a cluster, accounts JONES and SMITH will exist in the cluster environment with the same UIC. If the UICs of these accounts are not differentiated, each user will have the same access rights to various objects in the cluster. In this case, you should assign each account a unique UIC.</p>• Group UICs <p>Make sure that accounts that perform the same type of work have the same group UIC. Accounts in a single-computer environment probably follow this convention. However, there may be groups of users on each computer that will perform the same work in the cluster but that have group UICs unique to their local computer. As a rule, the group UIC for any given work category should be the same on each computer in the cluster. For example, data entry accounts on VENUS should have the same group UIC as data entry accounts on MARS.</p> <p>Note: If you change the UIC for a particular user, you should also change the owner UICs for that user's existing files and directories. You can use the DCL</p>

Step	Action
	commands SET FILE and SET DIRECTORY to make these changes. These commands are described in detail in the <i>VSI OpenVMS DCL Dictionary</i> .
3	<p>Choose the SYSUAF.DAT file from one of the computers to be a master SYSUAF.DAT.</p> <p>Note: See <i>A Comparison of System Management on OpenVMS AXP and OpenVMS VAX</i> for information about setting the number of SYSUAF process limits and quotas on an Alpha computer.</p>
4	<p>Merge the SYSUAF.DAT files from the other computers to the master SYSUAF.DAT by running the Convert utility (CONVERT) on the computer that owns the master SYSUAF.DAT. (See the <i>VSI OpenVMS Record Management Utilities Reference Manual</i> for a description of CONVERT.) To use CONVERT to merge the files, each SYSUAF.DAT file must be accessible to the computer that is running CONVERT.</p> <p>Syntax: To merge the UAFs into the master SYSUAF.DAT file, specify the CONVERT command in the following format:</p> <pre>CONVERT SYSUAF1, SYSUAF2, . . . SYSUAFn MASTER_SYSUAF</pre> <p>Note that if a given user name appears in more than one source file, only the first occurrence of that name appears in the merged file.</p> <p>Example: The following command sequence example creates a new SYSUAF.DAT file from the combined contents of the two input files:</p> <pre>\$ SET DEFAULT SYS\$SYSTEM \$ CONVERT/MERGE [SYS1.SYSEXE]SYSUAF.DAT, - _ \$ [SYS2.SYSEXE]SYSUAF.DAT SYSUAF.DAT</pre> <p>The CONVERT command in this example adds the records from the files [SYS1.SYSEXE]SYSUAF.DAT and [SYS2.SYSEXE]SYSUAF.DAT to the file SYSUAF.DAT on the local computer.</p> <p>After you run CONVERT, you have a master SYSUAF.DAT that contains records from the other SYSUAF.DAT files.</p>
5	Use AUTHORIZE to modify the accounts in the master SYSUAF.DAT according to the changes you marked on the initial listings of the SYSUAF.DAT files from each computer.
6	Place the master SYSUAF.DAT file in SYS\$COMMON:[SYSEXE].
7	Remove all node-specific SYSUAF.DAT files.

B.2. Merging RIGHTSLIST.DAT Files

If you need to merge RIGHTSLIST.DAT files, you can use a command sequence like the following:

```
$ ACTIVE_RIGHTSLIST = F$PARSE("RIGHTSLIST", "SYS$SYSTEM:.DAT")
$ CONVERT/SHARE/STAT 'ACTIVE_RIGHTSLIST' RIGHTSLIST.NEW
$ CONVERT/MERGE/STAT/EXCEPTION=RIGHTSLIST_DUPLICATES.DAT -
_ $ [SYS1.SYSEXE]RIGHTSLIST.DAT, [SYS2.SYSEXE]RIGHTSLIST.DAT RIGHTSLIST.NEW
$ DUMP/RECORD RIGHTSLIST_DUPLICATES.DAT
$ CONVERT/NOSORT/FAST/STAT RIGHTSLIST.NEW 'ACTIVE_RIGHTSLIST'
```

The commands in this example add the RIGHTSLIST.DAT files from two OpenVMS Cluster computers to the master RIGHTSLIST.DAT file in the current default directory. For detailed information about creating and maintaining RIGHTSLIST.DAT files, see the security guide for your system.

Appendix C. Cluster Troubleshooting

C.1. Diagnosing Computer Failures

This appendix contains information to help you perform troubleshooting operations for the following:

- Failures of computers to boot or to join the cluster
- Cluster hangs
- CLUEXIT bugchecks
- Port device problems

C.1.1. Preliminary Checklist

Before you initiate diagnostic procedures, be sure to verify that these conditions are met:

- All cluster hardware components are correctly connected and checked for proper operation.
- OpenVMS Cluster computers and mass storage devices are configured according to requirements specified in the VSI OpenVMS Cluster Software *Software Product Description*.
- When attempting to add a satellite to a cluster, you must verify that the LAN is configured according to requirements specified in the OpenVMS Cluster Software SPD. You must also verify that you have correctly configured and started the network, following the procedures described in Chapter 4.

If, after performing preliminary checks and taking appropriate corrective action, you find that a computer still fails to boot or to join the cluster, you can follow the procedures in Sections C.2 through C.3 to attempt recovery.

C.1.2. Sequence of Booting Events

To perform diagnostic and recovery procedures effectively, you must understand the events that occur when a computer boots and attempts to join the cluster. This section outlines those events and shows typical messages displayed at the console.

Note that events vary, depending on whether a computer is the first to boot in a new cluster or whether it is booting in an active cluster. Note also that some events (such as loading the cluster database containing the password and group number) occur only in OpenVMS Cluster systems on a LAN or IP.

The normal sequence of events is shown in Table C.1.

Table C.1. Sequence of Booting Events

Step	Action
1	<p>The computer boots. If the computer is a satellite, a message like the following shows the name and LAN address of the MOP server that has downline loaded the satellite. At this point, the satellite has completed communication with the MOP server and further communication continues with the system disk server, using OpenVMS Cluster communications.</p> <pre>%VAXcluster-I-SYSLOAD, system loaded from Node X...</pre>

Step	Action
	<p>For any booting computer, the OpenVMS “banner message” is displayed in the following format:</p> <pre>operating-system Version n.n dd-mmm-yyyy hh:mm.ss</pre>
2	<p>The computer attempts to form or join the cluster, and the following message appears:</p> <pre>waiting to form or join an OpenVMS Cluster system</pre> <p>If the computer is a member of an OpenVMS Cluster based on the LAN, the cluster security database (containing the cluster password and group number) is loaded. Optionally, the MSCP server, and TMSCP server can be loaded:</p> <pre>%VAXcluster-I-LOADSECDB, loading the cluster security database %MSCPLOAD-I-LOADMSCP, loading the MSCP disk server (TMSCPLOAD-I-LOADTMSCP, loading the TMSCP tape server</pre> <p>If the computer is a member of an OpenVMS Cluster based on IP, the IP configuration file is also loaded along with the cluster security database, the MSCP server and the TMSCP server:</p> <pre>%VMScluster-I-LOADIPCICFG, loading the IP cluster configuration file (TMSCPLOAD-S-LOADEDIPCICFG, Successfully loaded IP cluster configuration file</pre> <p>If the computer is a member of an OpenVMS Cluster based on IP, the IP configuration file is also loaded along with the cluster security database, the MSCP server and the TMSCP server:</p> <pre>%VMScluster-I-LOADIPCICFG, loading the IP cluster configuration file (TMSCPLOAD-S-LOADEDIPCICFG, Successfully loaded IP cluster configuration file</pre> <p>For IP-based cluster communication, the IP interface and TCP/IP services are enabled. The multicast and unicast addresses are added to the list of IP bus, WE0 and sends the Hello packet:</p> <pre>%PEA0, Configuration data for IP clusters found (TMSCPLOAD, IP Multicast enabled for cluster communication, Multicast address, 224.0.0.3 (TMSCPLOAD, Cluster communication enabled on IP interface, WE0 (TMSCPLOAD, Successfully initialized with TCP/IP services (TMSCPLOAD, Remote node Address, 16.138.185.68, added to unicast list of IP bus, WE0 (TMSCPLOAD, Remote node Address, 15.146.235.222, added to unicast list of IP bus, WE0 (TMSCPLOAD, Remote node Address, 15.146.239.192, added to unicast list of IP bus, WE0 (TMSCPLOAD, Hello sent on IP bus WE0 (TMSCPLOAD, Cluster communication successfully initialized on IP</pre>

Step	Action
	interface , WE0
3	<p>If the computer discovers a cluster, the computer attempts to join it. If a cluster is found, the connection manager displays one or more messages in the following format:</p> <pre data-bbox="427 421 1406 450">%CNXMAN, Sending VAXcluster membership request to system X...</pre> <p>Otherwise, the connection manager forms the cluster when it has enough votes to establish quorum (that is, when enough voting computers have booted).</p>
4	<p>As the booting computer joins the cluster, the connection manager displays a message in the following format:</p> <pre data-bbox="427 656 1177 685">%CNXMAN, now a VAXcluster member -- system X...</pre> <p>Note that if quorum is lost while the computer is booting, or if a computer is unable to join the cluster within 2 minutes of booting, the connection manager displays messages like the following:</p> <pre data-bbox="427 853 1198 1003">%CNXMAN, Discovered system X... %CNXMAN, Deleting CSB for system X... %CNXMAN, Established "connection" to quorum disk %CNXMAN, Have connection to system X... %CNXMAN, Have "connection" to quorum disk</pre> <p>The last two messages show any connections that have already been formed.</p>
5	<p>If the cluster includes a quorum disk, you may also see messages like the following:</p> <pre data-bbox="427 1144 1246 1200">%CNXMAN, Using remote access method for quorum disk %CNXMAN, Using local access method for quorum disk</pre> <p>The first message indicates that the connection manager is unable to access the quorum disk directly, either because the disk is unavailable or because it is accessed through the MSCP server. Another computer in the cluster that can access the disk directly must verify that a reliable connection to the disk exists.</p> <p>The second message indicates that the connection manager can access the quorum disk directly and can supply information about the status of the disk to computers that cannot access the disk directly.</p> <p>Note: The connection manager may not see the quorum disk initially because the disk may not yet be configured. In that case, the connection manager first uses remote access, then switches to local access.</p>
6	<p>Once the computer has joined the cluster, normal startup procedures execute. One of the first functions is to start the OPCOM process:</p> <pre data-bbox="427 1756 1342 1939">%%%%%%%%%% OPCOM 15-JAN-1994 16:33:55.33 %%%%%%%%%%% Logfile has been initialized by operator _X...\$OPA0: Logfile is SYS\$SYSROOT:[SYSMGR]OPERATOR.LOG;17 %%%%%%%%%% OPCOM 15-JAN-1994 16:33:56.43 %%%%%%%%%%% 16:32:32.93 Node X... (csid 0002000E) is now a VAXcluster member</pre>
7	As other computers join the cluster, OPCOM displays messages like the following:

Step	Action
	<pre> %%%%%%%% OPCOM 15-JAN-1994 16:34:25.23 %%%% (from node X...) 16:34:24.42 Node X... (csid 000100F3) received VAXcluster membership request from node X... </pre>

As startup procedures continue, various messages report startup events.

Hint: For troubleshooting purposes, you can include in your site-specific startup procedures messages announcing each phase of the startup process—for example, mounting disks or starting queues.

C.2. Satellite Fails to Boot

To boot successfully, a satellite must communicate with a MOP server over the LAN or IP. You can use the DECnet event logging feature to verify this communication. Perform the following procedure:

Step	Action						
1	Log in as system manager on the MOP server.						
2	<p>If event logging for management-layer events is not already enabled, enter the following NCP commands to enable it:</p> <pre> NCP> SET LOGGING MONITOR EVENT 0.* NCP> SET LOGGING MONITOR STATE ON </pre>						
3	<p>Enter the following DCL command to enable the terminal to receive DECnet messages reporting downline load events:</p> <pre> \$ REPLY/ENABLE=NETWORK </pre>						
4	<p>Boot the satellite. If the satellite and the MOP server can communicate and all boot parameters are correctly set, messages like the following are displayed at the MOP server's terminal:</p> <pre> DECnet event 0.3, automatic line service From node 2.4 (URANUS), 15-JAN-1994 09:42:15.12 Circuit QNA-0, Load, Requested, Node = 2.42 (OBERON) File = SYS\$SYSDEVICE:<SYS10.>, Operating system Ethernet address = 08-00-2B-07-AC-03 DECnet event 0.3, automatic line service From node 2.4 (URANUS), 15-JAN-1994 09:42:16.76 Circuit QNA-0, Load, Successful, Node = 2.44 (ARIEL) File = SYS\$SYSDEVICE:<SYS11.>, Operating system Ethernet address = 08-00-2B-07-AC-13 </pre> <table border="1"> <thead> <tr> <th>WHEN...</th> <th>THEN...</th> </tr> </thead> <tbody> <tr> <td>The satellite cannot communicate with the MOP server (VAX or Alpha).</td> <td>No message for that satellite appears. There may be a problem with a LAN cable connection or adapter service.</td> </tr> <tr> <td>The satellite's data in the DECnet database is incorrectly specified (for example, if the hardware address is incorrect).</td> <td> <p>A message like the following displays the correct address and indicates that a load was requested:</p> <pre> DECnet event 0.7, aborted service request From node 2.4 (URANUS) 15- JAN-1994 Circuit QNA-0, Line open error </pre> </td> </tr> </tbody> </table>	WHEN...	THEN...	The satellite cannot communicate with the MOP server (VAX or Alpha).	No message for that satellite appears. There may be a problem with a LAN cable connection or adapter service.	The satellite's data in the DECnet database is incorrectly specified (for example, if the hardware address is incorrect).	<p>A message like the following displays the correct address and indicates that a load was requested:</p> <pre> DECnet event 0.7, aborted service request From node 2.4 (URANUS) 15- JAN-1994 Circuit QNA-0, Line open error </pre>
WHEN...	THEN...						
The satellite cannot communicate with the MOP server (VAX or Alpha).	No message for that satellite appears. There may be a problem with a LAN cable connection or adapter service.						
The satellite's data in the DECnet database is incorrectly specified (for example, if the hardware address is incorrect).	<p>A message like the following displays the correct address and indicates that a load was requested:</p> <pre> DECnet event 0.7, aborted service request From node 2.4 (URANUS) 15- JAN-1994 Circuit QNA-0, Line open error </pre>						

Step	Action
	<p>Ethernet address=08-00-2B-03-29-99</p> <p>Note the absence of the node name, node address, and system root.</p>

Sections C.2.2 through C.2.5 provide more information about satellite boot troubleshooting and often recommend that you ensure that the system parameters are set correctly.

C.2.1. Displaying Connection Messages

To enable the display of connection messages during a conversational boot, perform the following steps:

Step	Action
1	<p>Enable conversational booting by setting the satellite's NISCS_CONV_BOOT system parameter to 1. On Integrity servers and Alpha systems, update the ALPHAVMSSYS.PAR file and on Integrity server systems update the IA64VMSSYS.PAR file in the system root on the disk server.</p>
2	<p>Perform a conversational boot.</p> <p>On Integrity servers and Alpha systems, enter the following command at the console:</p> <pre>>>> b -flags 0,1</pre> <p>On VAX systems, set bit <0> in register R5. For example, on a VAXstation 3100 system, enter the following command on the console:</p> <pre>>>> B/1</pre>
3	<p>Observe connection messages.</p> <p>Display connection messages during a satellite boot to determine which system in a large cluster is serving the system disk to a cluster satellite during the boot process. If booting problems occur, you can use this display to help isolate the problem with the system that is currently serving the system disk. Then, if your server system has multiple LAN adapters, you can isolate specific LAN adapters.</p>
4	<p>Isolate LAN adapters.</p> <p>Isolate a LAN adapter by methodically rebooting with only one adapter connected. That is, disconnect all but one of the LAN adapters on the server system and reboot the satellite. If the satellite boots when it is connected to the system disk server, then follow the same procedure using a different LAN adapter. Continue these steps until you have located the bad adapter.</p>

Reference: See also Appendix C for help with troubleshooting satellite booting problems.

C.2.2. General OpenVMS Cluster Satellite-Boot Troubleshooting

If a satellite fails to boot, use the steps outlined in this section to diagnose and correct problems in OpenVMS Cluster systems.

Step	Action						
1	Verify that the boot device is available. This check is particularly important for clusters in which satellites boot from multiple system disks.						
2	Verify that the DECnet network is up and running.						
3	Check the cluster group code and password. The cluster group code and password are set using the CLUSTER_CONFIG.COM procedure.						
4	Verify that you have installed the correct OpenVMS Integrity server, OpenVMS Alpha, and OpenVMS VAX licenses.						
5	<p>Verify system parameter values on each satellite node, as follows:</p> <pre>VAXCLUSTER = 2 NISCS_LOAD_PEA0 = 1 NISCS_USE_UDP=0 or 1 based on LAN or IP interconnect NISCS_LAN_OVRHD = 0 NISCS_MAX_PKTSZ = 1498¹ SCSNODE is the name of the computer. SCSSYSTEMID is a number that identifies the computer. VOTES = 0</pre> <p>The SCS parameter values are set differently depending on your system configuration.</p> <p>Reference: Appendix A describes how to set these SCS parameters.</p> <p>To check system parameter values on a satellite node that cannot boot, invoke the SYSGEN utility on a running system in the OpenVMS Cluster that has access to the satellite node's local root. (Note that you must invoke the SYSGEN utility from a node that is running the same type of operating system—for example, to troubleshoot an Alpha satellite node, you must run the SYSGEN utility on an Alpha system.) Check system parameters as follows:</p> <table border="1"> <thead> <tr> <th>Step</th> <th>Action</th> </tr> </thead> <tbody> <tr> <td>A</td> <td> <p>Find the local root of the satellite node on the system disk. The following example is from an Alpha system running DECnet for OpenVMS:</p> <pre>\$ MCR NCP SHOW NODE HOME CHARACTERISTICS Node Volatile Characteristics as of 10-JAN-1994 09:32:56 Remote node = 63.333 (HOME) Hardware address = 08-00-2B-30-96-86 Load file = APB.EXE Load Assist Agent = SYS\$SHARE:NISCS_LAA.EXE Load Assist Parameter = ALPHA\$SYSD:[SYS17.]</pre> <p>The local root in this example is ALPHA\$SYSD:[SYS17.].</p> <p>Reference: Refer to the DECnet-Plus documentation for equivalent information using NCL commands.</p> </td> </tr> <tr> <td>B</td> <td> <p>Enter the SHOW LOGICAL command at the system prompt to translate the logical name for ALPHA\$SYSD.</p> <pre>\$ SHO LOG ALPHA\$SYSD</pre> </td> </tr> </tbody> </table>	Step	Action	A	<p>Find the local root of the satellite node on the system disk. The following example is from an Alpha system running DECnet for OpenVMS:</p> <pre>\$ MCR NCP SHOW NODE HOME CHARACTERISTICS Node Volatile Characteristics as of 10-JAN-1994 09:32:56 Remote node = 63.333 (HOME) Hardware address = 08-00-2B-30-96-86 Load file = APB.EXE Load Assist Agent = SYS\$SHARE:NISCS_LAA.EXE Load Assist Parameter = ALPHA\$SYSD:[SYS17.]</pre> <p>The local root in this example is ALPHA\$SYSD:[SYS17.].</p> <p>Reference: Refer to the DECnet-Plus documentation for equivalent information using NCL commands.</p>	B	<p>Enter the SHOW LOGICAL command at the system prompt to translate the logical name for ALPHA\$SYSD.</p> <pre>\$ SHO LOG ALPHA\$SYSD</pre>
Step	Action						
A	<p>Find the local root of the satellite node on the system disk. The following example is from an Alpha system running DECnet for OpenVMS:</p> <pre>\$ MCR NCP SHOW NODE HOME CHARACTERISTICS Node Volatile Characteristics as of 10-JAN-1994 09:32:56 Remote node = 63.333 (HOME) Hardware address = 08-00-2B-30-96-86 Load file = APB.EXE Load Assist Agent = SYS\$SHARE:NISCS_LAA.EXE Load Assist Parameter = ALPHA\$SYSD:[SYS17.]</pre> <p>The local root in this example is ALPHA\$SYSD:[SYS17.].</p> <p>Reference: Refer to the DECnet-Plus documentation for equivalent information using NCL commands.</p>						
B	<p>Enter the SHOW LOGICAL command at the system prompt to translate the logical name for ALPHA\$SYSD.</p> <pre>\$ SHO LOG ALPHA\$SYSD</pre>						

Step	Action
	"ALPHA\$SYSD" = "\$69\$DUA121:" (LNM\$SYSTEM_TABLE)
C	<p>Invoke the SYSGEN utility on the system from which you can access the satellite's local disk. (This example invokes the SYSGEN utility on an Integrity server system or Alpha system using the parameter file IA64VMSSYS.PAR or ALPHAVMSSYS.PAR appropriately.) The following example illustrates how to enter the SYSGEN command USE with the system parameter file on the local root for the satellite node and then enter the SHOW command to query the parameters in question.</p> <pre> \$ MCR SYSGEN SYSGEN> USE \$69\$DUA121: [SYS17.SYSEXE]ALPHAVMSSYS.PAR SYSGEN> SHOW VOTES Parameter Name Current Default Min. Max. Unit Dynamic ----- VOTES 0 1 0 127 Votes SYSGEN> EXIT </pre>

¹For Ethernet adapters, the value of NISCS_MAX_PKT SZ is 1498. For Gigabit Ethernet and 10 Gb Ethernet adapters, the value is 8192.

C.2.3. MOP Server Troubleshooting

To diagnose and correct problems for MOP servers, follow the steps outlined in this section.

Step	Action
1	Perform the steps outlined in Section C.2.2.
2	<p>Verify the NCP circuit state is on and the service is enabled. Enter the following commands to run the NCP utility and check the NCP circuit state.</p> <pre> \$ MCR NCP NCP> SHOW CIRCUIT ISA-0 CHARACTERISTICS Circuit Volatile Characteristics as of 12-JAN-1994 10:08:30 Circuit = ISA-0 State = on Service = enabled Designated router = 63.1021 Cost = 10 Maximum routers allowed = 33 Router priority = 64 Hello timer = 15 Type = Ethernet Adjacent node = 63.1021 Listen timer = 45 </pre>
3	<p>If service is not enabled, you can enter NCP commands like the following to enable it</p> <pre> NCP> SET CIRCUIT circuit-id STATE OFF NCP> DEFINE CIRCUIT circuit-id SERVICE ENABLED NCP> SET CIRCUIT circuit-id SERVICE ENABLED STATE ON </pre> <p>The DEFINE command updates the permanent database and ensures that service is enabled the next time you start the network. Note that DECnet traffic is interrupted while the circuit is off.</p>

Step	Action
4	Verify that the load assist parameter points to the system disk and the system root for the satellite.
5	Verify that the satellite's system disk is mounted on the MOP server node.
6	On Integrity server systems and Alpha systems, verify that the load file is APB.EXE.
7	For MOP booting, the satellite node's parameter file, (ALPHAVMSYS.PAR for Integrity and Alpha computers and VAXVMSSYS.PAR for VAX computers) must be located in the [SYSEXE] directory of the satellite system root.
8	Ensure that the file CLUSTER_AUTHORIZE.DAT is located in the [SYSCOMMON.SYSEXE] directory of the satellite system root.

C.2.4. Disk Server Troubleshooting

To diagnose and correct problems for disk servers, follow the steps outlined in this section.

Step	Action
1	Perform the steps in Section C.2.2.
2	For each satellite node, verify the following system parameter values: MSCP_LOAD = 1 MSCP_SERVE_ALL = 1
3	The disk servers for the system disk must be connected directly to the disk.

C.2.5. Satellite Booting Troubleshooting

To diagnose and correct problems for satellite booting, follow the steps outlined in this section.

Step	Action				
1	Perform the steps in Sections C.2.2, C.2.3, and C.2.4.				
2	For each satellite node, verify that the VOTES system parameter is set to 0.				
3	<p>Verify the DECnet network database on the MOP servers by running the NCP utility and entering the following commands to display node characteristics. The following example displays information about an Alpha node named UTAH:</p> <pre>\$ MCR NCP NCP> SHOW NODE UTAH CHARACTERISTICS Node Volatile Characteristics as of 15-JAN-1994 10:28:09 Remote node = 63.227 (UTAH) Hardware address = 08-00-2B-2C-CE-E3 Load file = APB.EXE Load Assist Agent = SYS\$SHARE:NISCS_LAA.EXE Load Assist Parameter = \$69\$DUA100:[SYS17.]</pre> <p>The load file must be APB.EXE. In addition, when booting Alpha nodes, for each LAN adapter specified on the boot command line, the load assist parameter must point to the same system disk and root number.</p>				
4	Verify the following information in the NCP display:				
	<table border="1"> <thead> <tr> <th>Step</th> <th>Action</th> </tr> </thead> <tbody> <tr> <td></td> <td></td> </tr> </tbody> </table>	Step	Action		
Step	Action				

Step	Action
	<p>A Verify the DECnet address for the node.</p> <p>B Verify the load assist agent is SYSSHARE:NISCS_LAA.EXE.</p> <p>C Verify the load assist parameter points to the satellite system disk and correct root.</p> <p>D Verify that the hardware address matches the satellite's Ethernet address. At the satellite's console prompt, use the information shown in Table 8.3 to obtain the satellite's current LAN hardware address.</p> <p>Compare the hardware address values displayed by NCP and at the satellite's console. The values should be identical and should also match the value shown in the SYS\$MANAGER:NETNODE_UPDATE.COM file. If the values do not match, you must make appropriate adjustments. For example, if you have recently replaced the satellite's LAN adapter, you must execute CLUSTER_CONFIG.COM CHANGE function to update the network database and NETNODE_UPDATE.COM on the appropriate MOP server.</p>
5	<p>Perform a conversational boot to determine more precisely why the satellite is having trouble booting. The conversational boot procedure displays messages that can help you solve network booting problems. The messages provide information about the state of the network and the communications process between the satellite and the system disk server.</p> <p>Reference: Section C.2.6 describes booting messages for Alpha systems.</p>

C.2.6. Alpha Booting Messages (Alpha Only)

On Alpha systems, the messages are displayed as shown in Table C.2.

Table C.2. Alpha Booting Messages (Alpha Only)

Message	Comments
%VMScLuster-I-MOPSERVER, MOP server for downline load was node UTAH	
This message displays the name of the system providing the DECnet MOP downline load. This message acknowledges that control was properly transferred from the console performing the MOP load to the image that was loaded.	If this message is not displayed, either the MOP load failed or the wrong file was MOP downline loaded.
%VMScLuster-I-BUSONLINE, LAN adapter is now running 08-00-2B-2C-CE-E3	
This message displays the LAN address of the Ethernet or FDDI adapter specified in the boot command. Multiple lines can be displayed if multiple LAN devices were specified in the boot command line. The booting satellite can now attempt to locate the system disk by sending a message to the cluster multicast address.	If this message is not displayed, the LAN adapter is not initialized properly. Check the physical network connection. For FDDI, the adapter must be on the ring.
%VMScLuster-I-VOLUNTEER, System disk service volunteered by node EUROPA AA-00-04-00-4C-FD	
This message displays the name of a system claiming to serve the satellite system disk. This system has responded to the multicast	If this message is not displayed, one or more of the following

Message	Comments
message sent by the booting satellite to locate the servers of the system disk.	<p>situations may be causing the problem:</p> <ul style="list-style-type: none"> • The network path between the satellite and the boot server either is broken or is filtering the local area OpenVMS Cluster multicast messages. • The system disk is not being served. • The CLUSTER_AUTHORIZE.DAT file on the system disk does not match the other cluster members.
%VMSccluster-I-CREATECH, Creating channel to node EUROPA 08-00-2B-2C-CE-E2 08-00-2B-12-AE-A2	
This message displays the LAN address of the local LAN adapter (first address) and of the remote LAN adapter (second address) that form a communications path through the network. These adapters can be used to support a NISCA virtual circuit for booting. Multiple messages can be displayed if either multiple LAN adapters were specified on the boot command line or the system serving the system disk has multiple LAN adapters.	If you do not see as many of these messages as you expect, there may be network problems related to the LAN adapters whose addresses are not displayed. Use the Local Area OpenVMS Cluster Network Failure Analysis Program for better troubleshooting (see Section D.5).
%VMSccluster-I-OPENVC, Opening virtual circuit to node EUROPA	
This message displays the name of a system that has established an NISCA virtual circuit to be used for communications during the boot process. Booting uses this virtual circuit to connect to the remote MSCP server.	
%VMSccluster-I-MSCPCONN, Connected to a MSCP server for the system disk, node EUROPA	
This message displays the name of a system that is actually serving the satellite system disk.	If this message is not displayed, the system that claimed to serve the system disk could not serve the disk. Check the OpenVMS Cluster configuration.
%VMSccluster-W-SHUTDOWNCH, Shutting down channel to node EUROPA 08-00-2B-2C-CE-E3 08-00-2B-12-AE-A2	
This message displays the LAN address of the local LAN adapter (first address) and of the remote LAN adapter (second address) that have just lost communications. Depending on the type of failure, multiple messages may be displayed if either the booting system or the system serving the system disk has multiple LAN adapters.	

Message	Comments
%VMSccluster-W-CLOSEVC, Closing virtual circuit to node EUROPA	
This message indicates that NISCA communications have failed to the system whose name is displayed.	
%VMSccluster-I-RETRY, Attempting to reconnect to a system disk server	
This message indicates that an attempt will be made to locate another system serving the system disk. The LAN adapters will be reinitialized and all communications will be restarted.	
%VMSccluster-W-PROTOCOL_TIMEOUT, NISCA protocol timeout	
Either the booting node has lost connections to the remote system or the remote system is no longer responding to requests made by the booting system. In either case, the booting system has declared a failure and will reestablish communications to a boot server.	

C.3. Computer Fails to Join the Cluster

If a computer fails to join the cluster, follow the procedures in this section to determine the cause.

C.3.1. Verifying OpenVMS Cluster Software Load

To verify that OpenVMS Cluster software has been loaded, follow these instructions:

Step	Action
1	Look for connection manager (%CNXMAN) messages like those shown in Section C.1.2.
2	If no such messages are displayed, OpenVMS Cluster software probably was not loaded at boot time. Reboot the computer in conversational mode. At the SYSBOOT> prompt, set the VAXCLUSTER parameter to 2.
3	For OpenVMS Cluster systems communicating over the LAN or mixed interconnects, set NISCS_LOAD_PEA0 to 1 and VAXCLUSTER to 2. These parameters should also be set in the computer's MODPARAMS.DAT file. (For more information about booting a computer in conversational mode, consult your installation and operations guide).
4	For OpenVMS Cluster systems on the LAN, verify that the cluster security database file (SYS\$COMMON:CLUSTER_AUTHORIZE.DAT) exists and that you have specified the correct group number for this cluster (see Section 10.8.1).

C.3.2. Verifying Boot Disk and Root

To verify that the computer has booted from the correct disk and system root, follow these instructions:

Step	Action
1	If %CNXMAN messages are displayed, and if, after the conversational reboot, the computer still does not join the cluster, check the console output on all active computers and look for messages indicating that one or more computers found a

Step	Action
	remote computer that conflicted with a known or local computer. Such messages suggest that two computers have booted from the same system root.
3	<p>If you find it necessary to modify the computer's bootstrap command procedure (console media), you may be able to do so on another processor that is already running in the cluster.</p> <p>Replace the running processor's console media with the media to be modified, and use the Exchange utility and a text editor to make the required changes. Consult the appropriate processor-specific installation and operations guide for information about examining and editing boot command files.</p>

C.3.3. Verifying SCSNODE and SCSSYSTEMID Parameters

To be eligible to join a cluster, a computer must have unique SCSNODE and SCSSYSTEMID parameter values.

Step	Action						
1	Check that the current values do not duplicate any values set for existing OpenVMS Cluster computers. To check values, you can perform a conversational bootstrap operation.						
2	<p>If the values of SCSNODE or SCSSYSTEMID are not unique, do either of the following:</p> <ul style="list-style-type: none"> Alter <i>both</i> values. Reboot all other computers. <p>Note: To modify values, you can perform a conversational bootstrap operation. However, for reliable future bootstrap operations, specify appropriate values for these parameters in the computer's MODPARAMS.DAT file.</p> <table border="1"> <thead> <tr> <th>WHEN you change...</th> <th>THEN...</th> </tr> </thead> <tbody> <tr> <td>The SCSNODE parameter</td> <td>Change the DECnet node name too, because both names must be the same.</td> </tr> <tr> <td>Either the SCSNODE parameter or the SCSSYSTEMID parameter on a node that was previously an OpenVMS Cluster member</td> <td>Change the DECnet node number, too, because both numbers must be the same. Reboot the entire cluster.</td> </tr> </tbody> </table>	WHEN you change...	THEN...	The SCSNODE parameter	Change the DECnet node name too, because both names must be the same.	Either the SCSNODE parameter or the SCSSYSTEMID parameter on a node that was previously an OpenVMS Cluster member	Change the DECnet node number, too, because both numbers must be the same. Reboot the entire cluster.
WHEN you change...	THEN...						
The SCSNODE parameter	Change the DECnet node name too, because both names must be the same.						
Either the SCSNODE parameter or the SCSSYSTEMID parameter on a node that was previously an OpenVMS Cluster member	Change the DECnet node number, too, because both numbers must be the same. Reboot the entire cluster.						

C.3.4. Verifying Cluster Security Information

To verify the cluster group code and password, follow these instructions:

Step	Action
1	Verify that the database file SYS \$COMMON:CLUSTER_AUTHORIZE.DAT exists.
2	For clusters with multiple system disks, ensure that the correct (same) group number and password were specified for each.

Step	Action
	Reference: See Section 10.8 to view the group number and to reset the password in the CLUSTER_AUTHORIZE.DAT file using the SYSMAN utility.

C.4. Startup Procedures Fail to Complete

If a computer boots and joins the cluster but appears to hang before startup procedures complete—that is, before you are able to log in to the system—be sure that you have allowed sufficient time for the startup procedures to execute.

IF...	THEN...
The startup procedures fail to complete after a period that is normal for your site.	Try to access the procedures from another OpenVMS Cluster computer and make appropriate adjustments. For example, verify that all required devices are configured and available. One cause of such a failure could be the lack of some system resource, such as NPAGEDYN or page file space.
You suspect that the value for the NPAGEDYN parameter is set too low.	Perform a conversational bootstrap operation to increase it. Use SYSBOOT to check the current value, and then double the value.
You suspect a shortage of page file space, and another OpenVMS Cluster computer is available.	Log in on that computer and use the System Generation utility (SYSGEN) to provide adequate page file space for the problem computer. Note: Insufficient page-file space on the booting computer might cause other computers to hang.
The computer still cannot complete the startup procedures.	Contact your VSI support representative.

C.5. Diagnosing LAN Component Failures

Section D.5 provides troubleshooting techniques for LAN component failures (for example, broken LAN bridges). That appendix also describes techniques for using the Local Area OpenVMS Cluster Network Failure Analysis Program.

Intermittent LAN component failures (for example, packet loss) can cause problems in the NISCA transport protocol that delivers System Communications Services (SCS) messages to other nodes in the OpenVMS Cluster. Appendix F describes troubleshooting techniques and requirements for LAN analyzer tools.

C.6. Diagnosing Cluster Hangs

Conditions like the following can cause a OpenVMS Cluster computer to suspendprocess or system activity (that is, to hang):

Condition	Reference
Cluster quorum is lost.	Section C.6.1

Condition	Reference
A shared cluster resource is inaccessible.	Section C.6.2

C.6.1. Cluster Quorum is Lost

The OpenVMS Cluster quorum algorithm coordinates activity among OpenVMS Cluster computers and ensures the integrity of shared cluster resources. (The quorum algorithm is described fully in Chapter 2.) Quorum is checked after any change to the cluster configuration—for example, when a voting computer leaves or joins the cluster. If quorum is lost, process and I/O activity on all computers in the cluster are blocked.

Information about the loss of quorum and about clusterwide events that cause loss of quorum are sent to the OPCOM process, which broadcasts messages to designated operator terminals. The information is also broadcast to each computer's operator console (OPA0), unless broadcast activity is explicitly disabled on that terminal. However, because quorum may be lost before OPCOM has been able to inform the operator terminals, the messages sent to OPA0 are the most reliable source of information about events that cause loss of quorum.

If quorum is lost, you might add or reboot a node with additional votes.

Reference: See also the information about cluster quorum in Section 10.11.

C.6.2. Inaccessible Cluster Resource

Access to shared cluster resources is coordinated by the distributed lock manager. If a particular process is granted a lock on a resource (for example, a shared data file), other processes in the cluster that request incompatible locks on that resource must wait until the original lock is released. If the original process retains its lock for an extended period, other processes waiting for the lock to be released may appear to hang.

Occasionally, a system activity must acquire a restrictive lock on a resource for an extended period. For example, to perform a volume rebuild, system software takes out an exclusive lock on the volume being rebuilt. While this lock is held, no processes can allocate space on the disk volume. If they attempt to do so, they may appear to hang.

Access to files that contain data necessary for the operation of the system itself is coordinated by the distributed lock manager. For this reason, a process that acquires a lock on one of these resources and is then unable to proceed may cause the cluster to appear to hang.

For example, this condition may occur if a process locks a portion of the system authorization file (SYS\$SYSTEM:SYSUAF.DAT) for write access. Any activity that requires access to that portion of the file, such as logging in to an account with the same or similar user name or sending mail to that user name, is blocked until the original lock is released. Normally, this lock is released quickly, and users do not notice the locking operation.

However, if the process holding the lock is unable to proceed, other processes could enter a wait state. Because the authorization file is used during login and for most process creation operations (for example, batch and network jobs), blocked processes could rapidly accumulate in the cluster. Because the distributed lock manager is functioning normally under these conditions, users are not notified by broadcast messages or other means that a problem has occurred.

C.7. Diagnosing CLUEXIT Bugchecks

The operating system performs **bugcheck** operations only when it detects conditions that could compromise normal system activity or endanger data integrity. A **CLUEXIT bugcheck** is a type of bugcheck initiated by the connection manager, the OpenVMS Cluster software component that manages the interaction of cooperating OpenVMS Cluster computers. Most such bugchecks are triggered by conditions resulting from hardware failures (particularly failures in communications paths), configuration errors, or system management errors.

C.7.1. Conditions Causing Bugchecks

The most common conditions that result in CLUEXIT bugchecks are as follows:

Possible Bugcheck Causes	Recommendations
<p>The cluster connection between two computers is broken for longer than RECNXINTERVAL seconds. Thereafter, the connection is declared irrevocably broken. If the connection is later reestablished, one of the computers shut down with a CLUEXIT bugcheck.</p> <p>This condition can occur:</p> <ul style="list-style-type: none"> • Upon recovery with battery backup after a power failure • After the repair of an SCS communication link • After the computer was halted for a period longer than the number of seconds specified for the RECNXINTERVAL parameter and was restarted with a CONTINUE command entered at the operator console 	<p>Determine the cause of the interrupted connection and correct the problem. For example, if recovery from a power failure is longer than RECNXINTERVAL seconds, you may want to increase the value of the RECNXINTERVAL parameter on all computers.</p>
<p>Cluster partitioning occurs. A member of a cluster discovers or establishes connection to a member of another cluster, or a foreign cluster is detected in the quorum file.</p>	<p>Review the setting of EXPECTED_VOTES on all computers.</p>
<p>The value specified for the SCSMAXMSG system parameter on a computer is too small.</p>	<p>Verify that the value of SCSMAXMSG on all OpenVMS Cluster computers is set to a value that is at the least the default value.</p>

C.8. Port Communications

These sections provide detailed information about port communications to assisting diagnosing port communication problems.

C.8.1. LAN Communications

For clusters that include Ethernet or FDDI interconnects, a multicast scheme is used to locate computers on the LAN. Approximately every 3 seconds, the port emulator driver (PEDRIVER) sends a HELLO datagram message through each LAN adapter to a cluster-specific multicast address that is derived from the cluster group number. The driver also enables the reception of these messages from other computers. When the driver receives a HELLO datagram message from a computer with which

it does not currently share an open virtual circuit, it attempts to create a circuit. HELLO datagram messages received from a computer with a currently open virtual circuit indicate that the remote computer is operational.

A standard, three-message exchange handshake is used to create a virtual circuit. The handshake messages contain information about the transmitting computer and its record of the cluster password. These parameters are verified at the receiving computer, which continues the handshake only if its verification is successful. Thus, each computer authenticates the other. After the final message, the virtual circuit is opened for use by both computers.

C.8.2. System Communications Services (SCS) Connections

System services such as the disk class driver, connection manager, and the MSCP and TMSCP servers communicate between computers with a protocol called System Communications Services (SCS). SCS is responsible primarily for forming and breaking inter system process connections and for controlling flow of message traffic over those connections. SCS is implemented in the port driver (for example, PADRIVER, PBDRIVER, PEDRIVER, PIDRIVER), and in a loadable piece of the operating system called SCSLOA.EXE (loaded automatically during system initialization).

When a virtual circuit has been opened, a computer periodically probes are mote computer for system services that the remote computer may be offering. The SCS directory service, which makes known services that a computer is offering, is always present both on computers and HSC subsystems. As system services discover their counterparts on other computers and HSC subsystems, they establish SCS connections to each other. These connections are full duplex and are associated with a particular virtual circuit. Multiple connections are typically associated with a virtual circuit.

C.9. Diagnosing Port Failures

This section describes the hierarchy of communication paths and describes where failures can occur.

C.9.1. Hierarchy of Communication Paths

Taken together, SCS, the port drivers, and the port itself support a hierarchy of communication paths. Starting with the most fundamental level, these are as follows:

- The physical wires. The Ethernet is a single coaxial cable. The port chooses the free path or, if both are free, an arbitrary path(implemented in the cables and managed by the port).
- The virtual circuit (implemented in LAN port emulator driver (PEDRIVER) and partly in SCS software).
- The SCS connections (implemented in system software).

C.9.2. Where Failures Occur

Failures can occur at each communication level and in each component. Failures at one level translate into failures elsewhere, as described in Table C.3.

Table C.3. Port Failures

Communication Level	Failures
Wires	If the LAN fails or is disconnected, LAN traffic stops or is interrupted, depending on the nature of the failure. All traffic is directed over the

Communication Level	Failures
	remaining good path. When the wire is repaired, the repair is detected automatically by port polling, and normal operations resume on all ports.
Virtual circuit	<p>If no path works between a pair of ports, the virtual circuit fails and is closed. A path failure is discovered for the LAN, when no multicast HELLO datagram message or incoming traffic is received from another computer.</p> <p>When a virtual circuit fails, every SCS connection on it is closed. The software automatically reestablishes connections when the virtual circuit is reestablished. Normally, reestablishing a virtual circuit takes several seconds after the problem is corrected.</p>
LAN adapter	If a LAN adapter device fails, attempts are made to restart it. If repeated attempts fail, all channels using that adapter are broken. A channel is a pair of LAN addresses, one local and one remote. If the last open channel for a virtual circuit fails, the virtual circuit is closed and the connections are broken.
SCS connection	When the software protocols fail or, in some instances, when the software detects a hardware malfunction, a connection is terminated. Other connections are usually unaffected, as is the virtual circuit. Breaking of connections is also used under certain conditions as an error recovery mechanism—most commonly when there is insufficient nonpaged pool available on the computer.
Computer	If a computer fails because of operator shutdown, bugcheck, or halt, all other computers in the cluster record the shutdown as failures of their virtual circuits to the port on the shut down computer.

C.9.3. Verifying Virtual Circuits

To diagnose communication problems, you can invoke the Show Cluster utility using the instructions in Table C.4.

Table C.4. How to Verify Virtual Circuit States

Step	Action	What to Look for
1	Tailor the SHOW CLUSTER report by entering the SHOW CLUSTER command ADD CIRCUIT, CABLE_STATUS. This command adds a class of information about all the virtual circuits as seen from the computer on which you are running SHOW CLUSTER. CABLE_STATUS indicates the status of the path for the circuit from the CI interface on the local system to the CI interface on the remote system.	<p>Primarily, you are checking whether there is a virtual circuit in the OPEN state to the failing computer. Common causes of failure to open a virtual circuit and keep it open are the following:</p> <ul style="list-style-type: none"> • Port errors on one side or the other • Cabling errors • A port set off line because of software problems • Insufficient nonpaged pool on both sides

Step	Action	What to Look for	
		<ul style="list-style-type: none"> Failure to set correct values for the SCSNODE, SCSSYSTEMID, PAMAXPORT, PANOPOLL, PASTIMOUT, and PAPOLLINTERVAL system parameters 	
2	Run SHOW CLUSTER from each active computer in the cluster to verify whether each computer's view of the failing computer is consistent with every other computer's view.	<p>If no virtual circuit is open to the failing computer, check the bottom of the SHOW CLUSTER display:</p> <ul style="list-style-type: none"> For information about circuits to the port of the failing computer. Virtual circuits in partially open states are shown at the bottom of the display. If the circuit is shown in a state other than OPEN, communications between the local and remote ports are taking place, and the failure is probably at a higher level than in port or cable hardware. To see whether both path A and path B to the failing port are good. The loss of one path should not prevent a computer from participating in a cluster. 	
	WHEN...		THEN...
	All the active computers have a consistent view of the failing computer		The problem may be in the failing computer.
	Only one of several active computers detects that the newcomer is failing	That particular computer may have a problem.	

C.9.4. Verifying LAN Connections

The Local Area OpenVMS Cluster Network Failure Analysis Program described in Section D.4 uses the HELLO datagram messages to verify continuously the network paths (channels) used by PEDRIVER. This verification process, combined with physical description of the network, can:

- Isolate failing network components
- Group failing channels together and map them onto the physical network description
- Call out the common components related to the channel failures

C.10. Analyzing Error-Log Entries for Port Devices

Monitoring events recorded in the error log can help you anticipate and avoid potential problems. From the total error count (displayed by the DCL command SHOW DEVICES *device-name*), you can determine whether errors are increasing. If so, you should examine the error log.

C.10.1. Examine the Error Log

The DCL command `ANALYZE/ERROR_LOG` invokes the Error Log utility to report the contents of an error-log file.

Reference: For more information about the Error Log utility, see the *VSI OpenVMS System Management Utilities Reference Manual*.

Some error-log entries are informational only while others require action.

Table C.5. Informational and Other Error-Log Entries

Error Type	Action Required?	Purpose
<p><i>Informational</i> error-log entries require no action. For example, if you shut down a computer in the cluster, all other active computers that have open virtual circuits between themselves and the computer that has been shut down make entries in their error logs. Such computers record up to three errors for the event:</p> <ul style="list-style-type: none"> • Path A received no response. • Path B received no response. • The virtual circuit is being closed. 	No	These messages are normal and reflect the change of state in the circuits to the computer that has been shut down.
<p><i>Other</i> error-log entries indicate problems that degrade operation or nonfatal hardware problems. The operating system might continue to run satisfactorily under these conditions.</p>	Yes	Detecting these problems early is important to preventing nonfatal problems (such as loss of a single CI path) from becoming serious problems (such as loss of both paths).

C.10.2. Formats

Errors and other events on LAN cause port drivers to enter information in the system error log in one of two formats:

- Device attention

Device-attention entries for the LAN, device-attention entries typically record errors on a LAN adapter device.

- Logged message

Logged-message entries record the receipt of a message packet that contains erroneous data or that signals an error condition.

Section C.10.4 describe those formats.

C.10.3. LAN Device-Attention Entries

Example C.1 shows device-attention entries for the LAN.

Example C.1. LAN Device-Attention Entry

```

**** V3.4 ***** ENTRY 337
***** ❶

Logging OS                1. OpenVMS
System Architecture       2. Alpha
OS version                XC56-BL2
Event sequence number     96.
Timestamp of occurrence   16-SEP-2009 16:33:03 ❷
Time since reboot         0 Day(s) 0:50:08
Host name                 PERK

System Model              AlphaServer ES45 Model 2 ❸

Entry Type                98. Asynchronous Device Attention

---- Device Profile ----
Unit                      PERK$PEAO ❹
Product Name              NI-SCA Port

---- NISCA Port Data ----
Error Type and SubType    x0700 Device Error, Fatal Error Detected by
                          Datalink ❺
Status                    x0000120100000001 ❻
Datalink Device Name      EIA2: ❼
Remote Node Name ❸
Remote Address            x0000000000000000 ❹
Local Address             x000063B4000400AA ❿
Error Count               1. Error Occurrences This Entry ❶

----- Software Info -----
UCB$x_ERRCNT             2. Errors This Unit

```

The following table describes the LAN device-attention entries in Example C.1.

- ❶ The four lines are the entry heading. These lines contain the number of the entry in this error log file, the architecture, the OS version and the sequence number of this error. Each entry in the log file contains such a heading.
- ❷ This line contains the date and time.
- ❸ The next two lines contain the system model and the entry type.
- ❹ This line shows the name of the subsystem and component that caused the entry.
- ❺ This line shows the reason for the entry. The LAN driver has shut down the data link because of a fatal error. The data link will be restarted automatically, if possible.
- ❻ The first longword shows the I/O completion status returned by the LAN driver. The second longword is the VCI event code delivered to PEDRIVER by the LAN driver.
- ❼ DATALINK NAME is the name of the LAN device on which the error occurred.

- ⑧ REMOTE NODE is the name of the remote node to which the packet was being sent. If zeros are displayed, either no remote node was available or no packet was associated with the error.
- ⑨ REMOTE ADDR is the LAN address of the remote node to which the packet was being sent. If zeros are displayed, no packet was associated with the error.
- ⑩ LOCAL ADDR is the LAN address of the local node.
- ⑪ ERROR CNT. Because some errors can occur at extremely high rates, some error log entries represent more than one occurrence of an error. This field indicates how many. The errors counted occurred in the 3 seconds preceding the timestamp on the entry.

C.10.4. Logged Message Entries

Logged-message entries are made when the LAN port receives a response that contains either data that the port driver cannot interpret or an error code in status field of the response.

C.10.5. Error-Log Entry Descriptions

This section describes error-log entries for the CI and LAN ports. Each entry shown is followed by a brief description of what the associated port driver (for example, PADRIVER, PBDRIVER, PEDRIVER) does, and the suggested action a system manager should take. In cases where you are advised to contact your VSI support representative. and save crash dumps, it is important to capture the crash dumps as soon as possible after the error. For CI entries, note that path A and path 0 are the same path, and that path B and path 1 are the same path.

Table C.6 lists error-log messages.

Table C.6. Port Messages for All Devices

Message	Result	User Action
BIIC FAILURE	The port driver attempts to reinitialize the port; after 50 failed attempts, it marks the device off line.	Contact your VSI support representative.
11/750 CPU MICROCODE NOT ADEQUATE FOR PORT	The port driver sets the port off line with no retries attempted. In addition, if this port is needed because the computer is booted from an HSC subsystem or is participating in a cluster, the computer bugchecks with a UCODEREV code bugcheck.	Read the appropriate section in the current OpenVMS Cluster Software SPD for information on required computer microcode revisions. Contact VSI support representative, if necessary.
PORT MICROCODE REV NOT CURRENT, BUT SUPPORTED	The port driver detected that the microcode is not at the current level, but the port driver will continue normally. This error is logged as a warning only.	Contact your VSI support representative when it is convenient to have the microcode updated.
DATAGRAM FREE QUEUE INSERT FAILURE	The port driver attempts to reinitialize the port; after 50 failed attempts, it marks the device off line.	Contact your VSI support representative. This error is caused by a failure to obtain access to an interlocked queue. Possible sources of the problem

Message	Result	User Action
		are CI hardware failures, or memory, SBI (11/780), CMI (11/750), or BI (8200, 8300, and 8800) contention.
DATAGRAM FREE QUEUE REMOVE FAILURE	The port driver attempts to reinitialize the port; after 50 failed attempts, it marks the device off line.	Contact your VSI support representative. This error is caused by a failure to obtain access to an interlocked queue. Possible sources of the problem are CI hardware failures, or memory, SBI (11/780), CMI (11/750), or BI (8200, 8300, and 8800) contention.
FAILED TO LOCATE PORT MICROCODE IMAGE	The port driver marks the device off line and makes no retries.	Make sure console volume contains the microcode file CI780.BIN (for the CI780, CI750, or CIBCI) or the microcode file CIBCA.BIN for the CIBCA-AA. Then reboot the computer.
HIGH PRIORITY COMMAND QUEUE INSERT FAILURE	The port driver attempts to reinitialize the port; after 50 failed attempts, it marks the device off line.	Contact your VSI support representative. This error is caused by a failure to obtain access to an interlocked queue. Possible sources of the problem are CI hardware failures, or memory, SBI (11/780), CMI (11/750), or BI (8200, 8300, and 8800) contention.
MSCP ERROR LOGGING DATAGRAM RECEIVED	On receipt of an error message from the HSC subsystem, the port driver logs the error and takes no other action. You should disable the sending of HSC informational error-log datagrams with the appropriate HSC console command because such datagrams take considerable space in the error-log data file.	Error-log datagrams are useful to read only if they are not captured on the HSC console for some reason (for example, if the HSC console ran out of paper.) This logged information duplicates messages logged on the HSC console.
INAPPROPRIATE SCA CONTROL MESSAGE	The port driver closes the port-to-port virtual circuit to the remote port.	Contact your VSI support representative. Save the error logs and the crash dumps from the local and remote computers.
INSUFFICIENT NON-PAGED POOL FOR INITIALIZATION	The port driver marks the device off line and makes no retries.	Reboot the computer with a larger value for NPAGEDYN or NPAGEVIR.
LOW PRIORITY CMD QUEUE INSERT FAILURE	The port driver attempts to reinitialize the port; after 50	Contact your VSI support representative. This error is

Message	Result	User Action
	failed attempts, it marks the device off line.	caused by a failure to obtain access to an interlocked queue. Possible sources of the problem are CI hardware failures, or memory, SBI (11/780), CMI (11/750), or BI (8200, 8300, and 8800) contention.
MESSAGE FREE QUEUE INSERT FAILURE	The port driver attempts to reinitialize the port; after 50 failed attempts, it marks the device off line.	Contact your VSI support representative. This error is caused by a failure to obtain access to an interlocked queue. Possible sources of the problem are CI hardware failures, or memory, SBI (11/780), CMI (11/750), or BI (8200, 8300, and 8800) contention.
MESSAGE FREE QUEUE REMOVE FAILURE	The port driver attempts to reinitialize the port; after 50 failed attempts, it marks the device off line.	Contact your VSI support representative. This error is caused by a failure to obtain access to an interlocked queue. Possible sources of the problem are CI hardware failures, or memory, SBI (11/780), CMI (11/750), or BI (8200, 8300, and 8800) contention.
MICRO-CODE VERIFICATION ERROR	The port driver detected an error while reading the microcode that it just loaded into the port. The driver attempts to reinitialize the port; after 50 failed attempts, it marks the device off line.	Contact your VSI support representative.
NO PATH-BLOCK DURING VIRTUAL CIRCUIT CLOSE	The port driver attempts to reinitialize the port; after 50 failed attempts, it marks the device off line.	Contact your VSI support representative. Save the error log and a crash dump from the local computer.
NO TRANSITION FROM UNINITIALIZED TO DISABLED	The port driver attempts to reinitialize the port; after 50 failed attempts, it marks the device off line.	Contact your VSI support representative.
PORT ERROR BIT(S) SET	The port driver attempts to reinitialize the port; after 50 failed attempts, it marks the device off line.	A maintenance timer expiration bit may mean that the PASTIMOUT system parameter is set too low and should be increased, especially if the local computer is running privileged user-written software. For all other bits, call your VSI support representative.

Message	Result	User Action
PORT HAS CLOSED VIRTUAL CIRCUIT	The port driver closed the virtual circuit that the local port opened to the remote port.	<p>Check the PPD\$B_STATUS field of the error-log entry for the reason the virtual circuit was closed. This error is normal if the remote computer failed or was shut down. For PEDRIVER, ignore the PPD\$B_OPC field value; it is an unknown opcode.</p> <p>If PEDRIVER logs a large number of these errors, there may be a problem either with the LAN or with a remote system, or nonpaged pool may be insufficient on the local system.</p>
PORT POWER DOWN	The port driver halts port operations and then waits for power to return to the port hardware.	Restore power to the port hardware.
PORT POWER UP	The port driver reinitializes the port and restarts port operations.	No action needed.
RECEIVED CONNECT WITHOUT PATH-BLOCK	The port driver attempts to reinitialize the port; after 50 failed attempts, it marks the device off line.	Contact your VSI support representative. Save the error log and a crash dump from the local computer.
REMOTE SYSTEM CONFLICTS WITH KNOWN SYSTEM	The configuration poller discovered a remote computer with SCSSYSTEMID and/or SCSNODE equal to that of another computer to which a virtual circuit is already open.	Shut down the new computer as soon as possible. Reboot it with a unique SCSSYSTEMID and SCSNODE. Do not leave the new computer up any longer than necessary. If you are running a cluster, and two computers with conflicting identity are polling when any other virtual circuit failure takes place in the cluster, then computers in the cluster may shut down with a CLUEXIT bugcheck.
RESPONSE QUEUE REMOVE FAILURE	The port driver attempts to reinitialize the port; after 50 failed attempts, it marks the device off line.	Contact your VSI support representative. This error is caused by a failure to obtain access to an interlocked queue. Possible sources of the problem are CI hardware failures, or memory, SBI (11/780), CMI (11/750), or BI (8200, 8300, and 8800) contention.

Message	Result	User Action
SCSSYSTEMID MUST BE SET TO NON-ZERO VALUE	The port driver sets the port off line without attempting any retries.	Reboot the computer with a conversational boot and set the SCSSYSTEMID to the correct value. At the same time, check that SCSNODE has been set to the correct nonblank value.
SOFTWARE IS CLOSING VIRTUAL CIRCUIT	The port driver closes the virtual circuit to the remote port.	Check error-log entries for the cause of the virtual circuit closure. Faulty transmission or reception on both paths, for example, causes this error and may be detected from the one or two previous error-log entries noting bad paths to this remote computer.
SOFTWARE SHUTTING DOWN PORT	The port driver attempts to reinitialize the port; after 50 failed attempts, it marks the device off line.	Check other error-log entries for the possible cause of the port reinitialization failure.
UNEXPECTED INTERRUPT	The port driver attempts to reinitialize the port; after 50 failed attempts, it marks the device off line.	Contact your VSI support representative.
UNRECOGNIZED SCA PACKET	The port driver closes the virtual circuit to the remote port. If the virtual circuit is already closed, the port driver inhibits datagram reception from the remote port.	Contact your VSI support representative. Save the error-log file that contains this entry and the crash dumps from both the local and remote computers.
VIRTUAL CIRCUIT TIMEOUT	The port driver closes the virtual circuit that the local CI port opened to the remote port. This closure occurs if the remote computer is running CI microcode Version 7 or later, and if the remote computer has failed to respond to any messages sent by the local computer.	This error is normal if the remote computer has halted, failed, or was shutdown. This error may mean that the local computer's TIMVCFail system parameter is set too low, especially if the remote computer is running privileged user-written software.
INSUFFICIENT NON-PAGED POOL FOR VIRTUAL CIRCUITS	The port driver closes virtual circuits because of insufficient pool.	Enter the DCL command SHOW MEMORY to determine pool requirements, and then adjust the appropriate system parameter requirements.

The descriptions in Table C.7 apply only to LAN devices.

Table C.7. Port Messages for LAN Devices

Message	Completion Status	Explanation	User Action
FATAL ERROR DETECTED BY DATALINK	First longword SSS_NORMAL (00000001), second longword (00001201)	The LAN driver stopped the local area OpenVMS Cluster protocol on the device. This completion status is returned when the SYS \$LAVC_STOP_BUS routine completes successfully. The SYS\$LAVC_STOP_BUS routine is called either from within the LAVC \$STOP_BUS.MAR program found in SYS\$EXAMPLES or from a user-written program. The local area OpenVMS Cluster protocol remains stopped on the specified device until the SYS\$LAVC_START_BUS routine execute successfully. The SYS \$LAVC_START_BUS routine is called from within the LAVC \$START_BUS.MAR program found in SYS \$EXAMPLES or from a user-written program.	If the protocol on the device was stopped inadvertently, then restart the protocol by assembling and executing the LAVC\$START_BUS program found in SYS \$EXAMPLES. Reference: See Appendix D for an explanation of the local area OpenVMS Cluster sample programs. Otherwise, this error message can be safely ignored.
	First longword is any value other than (00000001), second longword (00001201)	The LAN driver has shut down the device because of a fatal error and is returning all outstanding transmits with SSS_OPINCOMPL. The LAN device is restarted automatically.	Infrequent occurrences of this error are typically not a problem. If the error occurs frequently or is accompanied by loss or reestablishment of connections to remote computers, there may be a hardware problem. Check for the proper LAN adapter revision level or contact your VSI support representative.
	First longword (undefined), second longword (00001200)	The LAN driver has restarted the device successfully after a fatal error. This error-log message is usually preceded by a FATAL ERROR DETECTED BY DATALINK error-log message whose first completion status longword is anything other than	No action needed.

Message	Completion Status	Explanation	User Action
		00000001 and whose second completion status longword is 00001201.	
TRANSMIT ERROR FROM DATALINK	SS \$_OPINCOMPL (000002D4)	The LAN driver is in the process of restarting the data link because an error forced the driver to shut down the controller and all users (see FATALERROR DETECTED BY DATALINK).	
	SS \$_DEVREQERR (00000334)	The LAN controller tried to transmit the packet 16 times and failed because of defers and collisions. This condition indicates that LAN traffic is heavy.	
	SS \$_DISCONNECT (0000204C)	There was a loss of carrier during or after the transmit. This includes transmit attempts when the link is down.	The port emulator automatically recovers from any of these errors, but many such errors indicate either that the LAN controller is faulty or that the LAN is overloaded. If you suspect either of these conditions, contact your VSI support representative.
INVALID CLUSTER PASSWORD RECEIVED		A computer is trying to join the cluster using the correct cluster group number for this cluster but an invalid password. The port emulator discards the message. The probable cause is that another cluster on the LAN is using the same cluster group number.	Provide all clusters on the same LAN with unique cluster group numbers.
NISCS PROTOCOL VERSION MISMATCH RECEIVED		A computer is trying to join the cluster using a version of the cluster LAN protocol that is incompatible with the one in use on this cluster.	Install a version of the operating system that uses a compatible protocol, or change the cluster group number so that the computer joins a different cluster.

C.11. OPA0 Error-Message Logging and Broadcasting

Port drivers detect certain error conditions and attempt to log them. The port driver attempts both OPA0 error broadcasting and standard error logging under any of the following circumstances:

- The system disk has not yet been mounted.
- The system disk is undergoing mount verification.
- During mount verification, the system disk drive contains the wrong volume.
- Mount verification for the system disk has timed out.
- The local computer is participating in a cluster, and quorum has been lost.

Note the implicit assumption that the system and error-logging devices are one and the same.

The following table describes error-logging methods and their reliability.

Method	Reliability	Comments
Standard error logging to an error-logging device.	Under some circumstances, attempts to log errors to the error-logging device can fail. Such failures can occur because the error-logging device is not accessible when attempts are made to log the error condition.	Because of the central role that the port device plays in clusters, the loss of error-logged information in such cases makes it difficult to diagnose and fix problems.
Broadcasting selected information about the error condition to OPA0. (This is in addition to the port driver's attempt to log the error condition to the error-logging device.)	This method of reporting errors is not entirely reliable, because some error conditions may not be reported due to the way OPA0 error broadcasting is performed. This situation occurs whenever a second error condition is detected before the port driver has been able to broadcast the first error condition to OPA0. In such a case, only the first error condition is reported to OPA0, because that condition is deemed to be the more important one.	This second, redundant method of error logging captures at least some of the information about port-device error conditions that would otherwise be lost.

Note: Certain error conditions are always broadcast to OPA0, regardless of whether the error-logging device is accessible. In general, these are errors that cause the port to shut down either permanently or temporarily.

C.11.1. OPA0 Error Messages

One OPA0 error message for each error condition is always logged. The text of each error message is similar to the text in the summary displayed by formatting the corresponding standard error-log entry

using the Error Log utility. (See Section C.10.5 for a list of Error Log utility summary messages and their explanations).

Table C.8 lists the OPA0 error messages. The table is divided into units by error type. Many of the OPA0 error messages contain some optional information, such as the remote port number, CI packet information (flags, port operation code, response status, and port number fields), or specific CI port registers. The codes specify whether the message is always logged on OPA0 or is logged only when the system device is inaccessible.

Table C.8. OPA0 Messages

Error Message	Logged or Inaccessible
Software Errors During Initialization	
%PEA0, Configuration data for IP cluster not found	Logged
%Pxxn, Insufficient Non-Paged Pool for Initialization	Logged
%Pxxn, Failed to Locate Port Micro-code Image	Logged
%Pxxn, SCSSYSTEMID has NOT been set to a Non-Zero Value	Logged
Hardware Errors	
%Pxxn, BIIC failure – BICSR/BER/CNF xxxxxx/xxxxxx/xxxxxx	Logged
%Pxxn, Micro-code Verification Error	Logged
%Pxxn, Port Transition Failure – CNF/PMC/PSRxxxxxx/xxxxxx/xxxxxx	Logged
%Pxxn, Port Error Bit(s) Set – CNF/PMC/PSRxxxxxx/xxxxxx/xxxxxx	Logged
%Pxxn, Port Power Down	Logged
%Pxxn, Port Power Up	Logged
%Pxxn, Unexpected Interrupt – CNF/PMC/PSRxxxxxx/xxxxxx/xxxxxx	Logged
Queue Interlock Failures	
%Pxxn, Message Free Queue Remove Failure	Logged
%Pxxn, Datagram Free Queue Remove Failure	Logged
%Pxxn, Response Queue Remove Failure	Logged
%Pxxn, High Priority Command Queue Insert Failure	Logged
%Pxxn, Low Priority Command Queue Insert Failure	Logged
%Pxxn, Message Free Queue Insert Failure	Logged
%Pxxn, Datagram Free Queue Insert Failure	Logged
Cable Change-of-State Notification	
%Pxxn, Path #0. Has gone from GOOD to BAD – REMOTE PORT ¹ xxx	Inaccessible
%Pxxn, Path #1. Has gone from GOOD to BAD – REMOTE PORT ¹ xxx	Inaccessible
%Pxxn, Path #0. Has gone from BAD to GOOD – REMOTE PORT ¹ xxx	Inaccessible
%Pxxn, Path #1. Has gone from BAD to GOOD – REMOTE PORT ¹ xxx	Inaccessible
Key to CI Port Registers:	
CNF—configuration register	
PMC—port maintenance and control register	
PSR—port status register	
See also the CI hardware documentation for a detailed description of the CI port registers.	

Error Message	Logged or Inaccessible
%Pxxn, Cables have gone from UNCROSSED to CROSSED – REMOTE PORT ¹ xxx	Inaccessible
%Pxxn, Cables have gone from CROSSED to UNCROSSED – REMOTE PORT ¹ xxx	Inaccessible
%Pxxn, Path #0. Loopback has gone from GOOD to BAD – REMOTE PORT ¹ xxx	Logged
%Pxxn, Path #1. Loopback has gone from GOOD to BAD – REMOTE PORT ¹ xxx	Logged
%Pxxn, Path #0. Loopback has gone from BAD to GOOD – REMOTE PORT ¹ xxx	Logged
%Pxxn, Path #1. Loopback has gone from BAD to GOOD – REMOTE PORT ¹ xxx	Logged
%Pxxn, Path #0. Has become working but CROSSED to Path #1. – REMOTE PORT ¹ xxx	Inaccessible
%Pxxn, Path #1. Has become working but CROSSED to Path #0. – REMOTE PORT ¹ xxx	Inaccessible
Key to CI Port Registers:	
CNF—configuration register	
PMC—port maintenance and control register	
PSR—port status register	
See also the CI hardware documentation for a detailed description of the CI port registers.	

¹If the port driver can identify the remote SCS node name of the affected computer, the driver replaces the “REMOTE PORT xxx” text with “REMOTE SYSTEM X...”, where X... is the value of the system parameter SCSNODE on the remote computer. If the remote SCS node name is not available, the port driver uses the existing message format.

C.12. Integrity server Satellite Booting Messages

Table C.9 lists the Integrity server satellite booting messages.

Table C.9. Integrity server Satellite Booting Messages

Booting message	Comments
MAC address Booting over the network Loading.: EIA0 Mac(00-17-a4-51-ce-4a)	This message displays the MAC address of the satellite system that is being used for booting.
BOOTP database Client MAC Address: 00 17 A4 51 CE 4A ./ Client IP Address: 15.146.235.22 Subnet Mask: 255.255.254.0 BOOTP Server IP Address: 15.146.235.23 DHCP Server IP Address: 0.240.0.0 Boot file name: \$2\$DKA0:[SYS10.SYSCOMMON.SYSEXE] VMS_LOADER.EFI	This message displays the BOOTP database of the satellite system. It shows all the information provided on the boot server while configuring the satellite.
Small memory configurations	When booting OpenVMS Integrity server systems over

Booting message	Comments
<pre> ERROR: Unable to allocate aligned memory %VMS_LOADER-I-Cannot allocate 256Meg for memory disk. Falling back to 64Meg. %VMS_LOADER-I-Memorydisk allocated at:0x0000000010000000 </pre>	<p>the network or while booting OpenVMS as a guest OS under Integrity VM, OpenVMS allocates a memory disk from the main memory. For OpenVMS Version 8.4, the size of this memory disk defaults to 256 MB. However, for some older systems with relatively small memory configurations, this size cannot be allocated, and displays the following error message:</p> <pre> Unable to allocate aligned memory. </pre> <p>After this message is displayed, OpenVMS adopts a fallback strategy by allocating only 64 MB and excludes some newer drivers from the initial boot. The fallback message indicates that the action was performed. If the fallback message is displayed with no further error messages, the initial error message can be ignored.</p>
<pre> Boot progress Retrieving File Size. Retrieving File (TFTP). Starting: EIA0 Mac(00-17-a4-51-ce-4a) Loading memory disk from IP 15.146.235.23 Loading file: \$2\$DKA0: [SYS10.SYSCOMMON.SYSEXE]IPB.EXE from IP 15.146.235.23 %IPB-I-SATSYSDIS, Satellite boot from system device \$2\$DKA0: </pre>	<p>The system displays the detailed boot progress in the form of a system message when VMS_LOADER is obtained from the network, followed by one period character written to the console device for every file downloaded to start the boot sequence and last by a message indicating that IPB (the primary bootstrap image) has been loaded.</p>

Caution: Satellite node boot may fail if you register the hardware address of Integrity server satellite node for multiple purposes.

For example, if you attempt a satellite boot of an Integrity server node in a cluster that has an Integrity server node configured and another cluster node configured as an Info server boot node with the same MAC address, Integrity sever satellite node will fail its satellite boot.

This is because the hardware address of the Integrity server satellite node is registered as an Infoserver boot node as well as an Integrity server satellite node.

An output similar to the following is displayed:

```

Loading.: eib0 Mac(00-0e-7f-7e-08-d9)
Running LoadFile()
                    
```

```
CLIENT MAC ADDR: 00 0E 7F 7E 08 D9  
CLIENT IP: 16.116.42.85  MASK: 255.0.0.0  DHCP IP: 0.240.0.0
```

```
TSize.Running LoadFile()
```

```
Starting: eib0 Mac(00-0e-7f-7e-08-d9)
```

```
Loading memory disk from IP 16.116.40.168
```

```
Unable to open SYS$MEMORYDISK.DAT
```

```
FATAL ERROR: Unable to boot using memorydisk method.
```

Where; 16.116.40.168 is the IP address of the Alpha Infoserver node's IP address.

Appendix D. Sample Programs for LAN Control

Sample programs are provided to start and stop the NISCA protocol on a LAN adapter, and to enable LAN network failure analysis. The following programs are located in SYS\$EXAMPLES:

Program	Description
LAVC\$START_BUS.MAR	Starts the NISCA protocol on a specified LAN adapter.
LAVC\$STOP_BUS.MAR	Stops the NISCA protocol on a specified LAN adapter.
LAVC \$FAILURE_ANALYSIS.MAR	Enables LAN network failure analysis.
LAVC\$BUILD.COM	Assembles and links the sample programs.

Reference: The NISCA protocol, responsible for carrying messages across Ethernet LANs to other nodes in the cluster, is described in Appendix F.

D.1. Purpose of Programs

The port emulator driver, PEDRIVER, starts the NISCA protocol on all of the LAN adapters in the cluster. LAVC\$START_BUS.MAR and LAVC\$STOP_BUS.MAR are provided for cluster managers who want to split the network load according to protocol type and therefore do not want the NISCA protocol running on all of the LAN adapters.

Reference: See Section D.5 for information about editing and using the network failure analysis program.

D.2. Starting the NISCA Protocol

The sample program LAVC\$START_BUS.MAR, provided in SYS\$EXAMPLES, starts the NISCA protocol on a specific LAN adapter.

To build the program, perform the following steps:

Step	Action
1	Copy the files LAVC\$START_BUS.MAR and LAVC\$BUILD.COM from SYS\$EXAMPLES to your local directory.
2	Assemble and link the sample program using the following command: \$ @LAVC\$BUILD.COM LAVC\$START_BUS.MAR

D.2.1. Start the Protocol

To start the protocol on a LAN adapter, perform the following steps:

Step	Action
1	Use an account that has the PHY_IO privilege—you need this to run LAVC\$START_BUS.EXE.
2	Define the foreign command (DCL symbol).

Step	Action
3	Execute the foreign command (LAVC\$START_BUS.EXE), followed by the name of the LAN adapter on which you want to start the protocol.

Example: The following example shows how to start the NISCA protocol on LAN adapter ETA0:

```
$ START_BUS:==$SYS$DISK:[ ]LAVC$START_BUS.EXE
$ START_BUS ETA
```

D.3. Stopping the NISCA Protocol

The sample program LAVC\$STOP_BUS.MAR, provided in SYS\$EXAMPLES, stops the NISCA protocol on a specific LAN adapter.

Caution: Stopping the NISCA protocol on all LAN adapters causes satellites to hang and could cause cluster systems to fail with a CLUEXIT bugcheck.

Follow the steps below to build the program:

Step	Action
1	Copy the files LAVC\$STOP_BUS.MAR and LAVC\$BUILD.COM from SYS\$EXAMPLES to your local directory.
2	Assemble and link the sample program using the following command: \$ @LAVC\$BUILD.COM LAVC\$STOP_BUS.MAR

D.3.1. Stop the Protocol

To stop the NISCA protocol on a LAN adapter, perform the following steps:

Step	Action
1	Use an account that has the PHY_IO privilege—you need this to run LAVC\$STOP_BUS.EXE.
2	Define the foreign command (DCL symbol).
3	Execute the foreign command (LAVC\$STOP_BUS.EXE), followed by the name of the LAN adapter on which you want to stop the protocol.

Example: The following example shows how to stop the NISCA protocol on LAN adapter ETA0:

```
$ STOP_BUS:==$SYS$DISK[ ]LAVC$STOP_BUS.EXE
$ STOP_BUS ETA
```

D.3.2. Verify Successful Execution

When the LAVC\$STOP_BUS module executes successfully, the following device-attention entry is written to the system error log:

```
DEVICE ATTENTION...
NI-SCS SUB-SYSTEM...
FATAL ERROR DETECTED BY DATALINK...
```

In addition, the following hexadecimal values are written to the STATUS field of the entry:

First longword (00000001)

Second longword (00001201)

The error-log entry indicates expected behavior and can be ignored. However, if the first longword of the STATUS field contains a value other than hexadecimal value 00000001, an error has occurred and further investigation may be necessary.

D.4. Analyzing Network Failures

LAVC\$FAILURE_ANALYSIS.MAR is a sample program, located in SYS\$EXAMPLES, that you can edit and use to help detect and isolate a failed network component. When the program executes, it provides the physical description of your cluster communications network to the set of routines that perform the failure analysis.

D.4.1. Failure Analysis

Using the network failure analysis program can help reduce the time necessary for detection and isolation of a failing network component and, therefore, significantly increase cluster availability.

D.4.2. How the LAVC\$FAILURE_ANALYSIS Program Works

The following table describes how the LAVC\$FAILURE_ANALYSIS program works.

Step	Program Action
1	The program groups channels that fail and compares them with the physical description of the cluster network.
2	The program then develops a list of nonworking network components related to the failed channels and uses OPCOM messages to display the names of components with a probability of causing one or more channel failures. If the network failure analysis cannot verify that a portion of a path (containing multiple components) works, the program: <ol style="list-style-type: none"> 1. Calls out the first component in the path as the primary suspect (%LAVC-W-PSUSPECT) 2. Lists the other components as secondary or additional suspects (%LAVC-I-ASUSPECT)
3	When the component works again, OPCOM displays the message %LAVC-S-WORKING.

D.5. Using the Network Failure Analysis Program

Table D.1 describes the steps you perform to edit and use the network failure analysis program.

Table D.1. Procedure for Using the LAVC\$FAILURE_ANALYSIS.MAR Program

Step	Action	Reference
1	Collect and record information specific to your cluster communications network.	Section D.5.1

Step	Action	Reference
2	Edit a copy of LAVC \$FAILURE_ANALYSIS.MAR to include the information you collected.	Section D.5.2
3	Assemble, link, and debug the program.	Section D.5.3
4	Modify startup files to run the program only on the node for which you supplied data.	Section D.5.4
5	Execute the program on one or more of the nodes where you plan to perform the network failure analysis.	Section D.5.5
6	Modify MODPARAMS.DAT to increase the values of nonpaged pool parameters.	Section D.5.6
7	Test the Local Area OpenVMS Cluster Network Failure Analysis Program.	Section D.5.7

D.5.1. Create a Network Diagram

Follow the steps in Table D.2 to create a physical description of the network configuration and include it in electronic form in the LAVC\$FAILURE_ANALYSIS.MAR program.

Table D.2. Creating a Physical Description of the Network

Step	Action	Comments
1	Draw a diagram of your OpenVMS Cluster communications network.	<p>When you edit LAVC \$FAILURE_ANALYSIS.MAR, you include this drawing (in electronic form) in the program. Your drawing should show the physical layout of the cluster and include the following components:</p> <ul style="list-style-type: none"> • LAN segments or rings • LAN bridges • Wiring concentrators, DELNI interconnects, or DEMPR repeaters • LAN adapters • Integrity servers and Alpha systems <p>For large clusters, you may need to verify the configuration by tracing the cables.</p>
2	Give each component in the drawing a unique label.	<p>If your OpenVMS Cluster contains a large number of nodes, you may want to replace each node name with a shorter abbreviation. Abbreviating node names can help save space in the electronic form of the drawing when you include it in LAVC \$FAILURE_ANALYSIS.MAR. For example, you can replace the node name ASTRA with A and</p>

Step	Action	Comments
		call node ASTRA's two LAN adapters A1 and A2.
3	<p>List the following information for each component:</p> <ul style="list-style-type: none"> • Unique label • Type [SYSTEM, LAN_ADP, DELNI] • Location (the physical location of the component) • LAN address or addresses (if applicable) 	Devices such as DELNI interconnects, DEMPR repeaters, and cables do not have LAN addresses.
4	<p>Classify each component into one of the following categories:</p> <ul style="list-style-type: none"> • Node: Integrity server or Alpha system in the OpenVMS Cluster configuration. • Adapter: LAN adapter on the system that is normally used for OpenVMS Cluster communications. • <i>Component:</i> Generic component in the network. Components in this category can usually be shown to be working if at least one path through them is working. Wiring concentrators, DELNI interconnects, DEMPR repeaters, LAN bridges, and LAN segments and rings typically fall into this category. • Cloud: Generic component in the network. Components in this category cannot be shown to be working even if one or more paths are shown to be working. 	The cloud component is necessary only when multiple paths exist between two points within the network, such as with redundant bridging between LAN segments. At a high level, multiple paths can exist; however, during operation, this bridge configuration allows only one path to exist at one time. In general, this bridge example is probably better handled by representing the active bridge in the description as a component and ignoring the standby bridge. (You can identify the active bridge with such network monitoring software as RBMS or DECelms.) With the default bridge parameters, failure of the active bridge will be called out.
5	Use the component labels from step 3 to describe each of the connections in the OpenVMS Cluster communications network.	


```

;
;           Describe the network connections.
;
CONNECTION      Sa,      MPR_A
CONNECTION      MPR_A,  A1
CONNECTION      A1,      A
CONNECTION      MPR_A,  B1
CONNECTION      B1,      B

CONNECTION      Sa,      D1
CONNECTION      D1,      D

CONNECTION      Sa,      BRIDGES
CONNECTION      Sb,      BRIDGES

CONNECTION      Sb,      LNI_A
CONNECTION      LNI_A,  A2
CONNECTION      A2,      A
CONNECTION      LNI_A,  B2
CONNECTION      B2,      B

CONNECTION      Sb,      D2
CONNECTION      D2,      D

.PAGE

;           *** End of edits ***

```

In the program, Edit *number* identifies a place where you edit the program to incorporate information about your network. Make the following edits to the program:

Location	Action
Edit 1	<p>Define a category for each component in the configuration. Use the information from step 5 in Section D.5.1. Use the following format:</p> <pre>NEW_COMPONENT component_type category</pre> <p>Example: The following example shows how to define a DEMPR repeater as part of the component category:</p> <pre>NEW_COMPONENT DEMPR COMPONENT</pre>
Edit 2	<p>Incorporate the network map you drew for step 1 of Section D.5.1. Including the map here in LAVC\$FAILURE_ANALYSIS.MAR gives you an electronic record of the map that you can locate and update more easily than a drawing on paper.</p>
Edit 3	<p>List each OpenVMS Cluster node and its LAN adapters. Use one line for each node. Each line should include the following information. Separate the items of information with commas to create a table of the information.</p> <ul style="list-style-type: none"> • Component type, followed by a comma. • Label from the network map, followed by a comma. • Node name (for SYSTEM components only). If there is no node name, enter a comma.

Location	Action
	<ul style="list-style-type: none"> • Descriptive text that the network failure analysis program displays if it detects a failure with this component. Put this text within angle brackets (<>). This text should include the component's physical location. • LAN hardware address (for LAN adapters). • DECnet LAN address for the LAN adapter that DECnet uses.
Edit 4	<p>List each of the other network components. Use one line for each component. Each line should include the following information:</p> <ul style="list-style-type: none"> • Component name and category you defined with NEW_COMPONENT. • Label from the network map. • Descriptive text that the network failure analysis program displays if it detects a failure with this component. Include a description of the physical location of the component. • LAN hardware address (optional). • Alternate LAN address (optional).
Edit 5	<p>Define the connections between the network components. Use the CONNECTION macro and the labels for the two components that are connected. Include the following information:</p> <ul style="list-style-type: none"> • CONNECTION macro name • First component label • Second component label
<p>Reference: You can find more detailed information about this exercise within the source module SYS\$EXAMPLES:LAVC\$FAILURE_ANALYSIS.MAR.</p>	

D.5.3. Assemble and Link the Program

Use the following command procedure to assemble and link the program:

```
$ @LAVC$BUILD.COM LAVC$FAILURE_ANALYSIS.MAR
```

Make the edits necessary to fix the assembly or link errors, such as errors caused by mistyping component labels in the path description. Assemble the program again.

D.5.4. Modify Startup Files

Before you execute the LAVC\$FAILURE_ANALYSIS.EXE procedure, modify the startup files to run the procedure only on the node for which you supplied data.

Example: To execute the program on node OMEGA, you would modify the startup files in SYS\$COMMON:[SYSMGR] to include the following conditional statement:

```
$ If F$GETSYI ("nodename").EQS."OMEGA"
$ THEN
$   RUN SYS$MANAGER:LAVC$FAILURE_ANALYSIS.EXE
```

```
$ ENDIF
```

D.5.5. Execute the Program

To run the LAVC\$FAILURE_ANALYSIS.EXE program, follow these steps:

Step	Action
1	Use an account that has the PHY_IO privilege.
2	Execute the program on each of the nodes that will perform the network failure analysis: \$ RUN SYS\$MANAGER:LAVC\$FAILURE_ANALYSIS.EXE

After it executes, the program displays the approximate amount of nonpaged pool required for the network description. The display is similar to the following:

```
Non-paged Pool Usage: ~ 10004 bytes
```

D.5.6. Modify MODPARAMS.DAT

On each system running the network failure analysis, modify the file SYSS\$SPECIFIC:[SYSEXE]MODPARAMS.DAT to include the following lines, replacing *value* with the value that was displayed for nonpaged pool usage:

```
ADD_NPAGEDYN =
valueADD_NPAGEVIR =
value
```

Run AUTOGEN on each system for which you modified MODPARAMS.DAT.

D.5.7. Test the Program

Test the program by causing a failure. For example, disconnect a transceiver cable or ThinWire segment, or cause a power failure on a bridge, a DELNI interconnect, or a DEMPR repeater. Then check the OPCOM messages to see whether LAVC\$FAILURE_ANALYSIS reports the failed component correctly. If it does not report the failure, check your edits to the network failure analysis program.

D.5.8. Display Suspect Components

When an OpenVMS Cluster network component failure occurs, OPCOM displays a list of suspected components. Displaying the list through OPCOM allows the system manager to enable and disable selectively the display of these messages.

The following are sample displays:

```
%%%%%%%%%%%%% OPCOM 1-JAN-1994 14:16:13.30 %%%%%%%%%%%%%%
(from node BETA at 1-JAN-1994 14:15:55.38)
Message from user SYSTEM on BETA LAVC-W-PSUSPECT, component_name

%%%%%%%%%%%%% OPCOM 1-JAN-1994 14:16:13.41 %%%%%%%%%%%%%%
(from node BETA at 1-JAN-1994 14:15:55.49)
Message from user SYSTEM on BETA %LAVC-W-PSUSPECT, component_name
```

```
%%%%%%%%%% OPCOM 1-JAN-1994 14:16:13.50 %%%%%%%%%%%  
(from node BETA at 1-JAN-1994 14:15:55.58)  
Message from user SYSTEM on BETA %LAVC-I-ASUSPECT, component_name
```

The OPCOM display of suspected failures uses the following prefixes to list suspected failures:

- %LAVC-W-PSUSPECT – Primary suspects
- %LAVC-I-ASUSPECT – Secondary or additional suspects
- %LAVC-S-WORKING – Suspect component is now working

The text following the message prefix is the description of the network component you supplied when you edited LAVC\$FAILURE_ANALYSIS.MAR.

Appendix E. Subroutines for LAN Control

E.1. Introduction

In addition to the sample programs described in Appendix D, a number of subroutines are provided as a way of extending the capabilities of the sample programs. Table E.1 describes the subroutines.

Table E.1. Subroutines for LAN Control

Subroutine	Description
To manage LAN adapters:	
SYSS\$LAVC_START_BUS	Directs PEDRIVER to start the NISCA protocol on a specific LAN adapter.
SYSS\$LAVC_STOP_BUS	Directs PEDRIVER to stop the NISCA protocol on a specific LAN adapter.
To control the network failure analysis system:	
SYSS\$LAVC_DEFINE_NET_COMPONENT	Creates a representation of a physical network component.
SYS\$LAVC_DEFINE_NET_PATH	Creates a directed list of network components between two network nodes.
SYS\$LAVC_ENABLE_ANALYSIS	Enables the network failure analysis, which makes it possible to analyze future channel failures.
SYS\$LAVC_DISABLE_ANALYSIS	Stops the network failure analysis and deallocates the memory used for the physical network description.

E.1.1. Purpose of the Subroutines

The subroutines described in this appendix are used by the LAN control programs, LAVC\$FAILURE_ANALYSIS.MAR, LAVC\$START_BUS.MAR, and LAVC\$STOP_BUS.MAR. Although these programs are sufficient for controlling LAN networks, you may also find it helpful to use the LAN control subroutines to further manage LAN adapters.

E.2. Starting the NISCA Protocol

The SYSS\$LAVC_START_BUS subroutine starts the NISCA protocol on a specified LAN adapter. To use the routine SYSS\$LAVC_START_BUS, specify the following parameter:

Parameter	Description
BUS_NAME	String descriptor representing the LAN adapter name buffer, passed by reference. The LAN adapter name must consist of 15 characters or fewer.

Example: The following Fortran sample program uses SYSS\$LAVC_START_BUS to start the NISCA protocol on the LAN adapter XQA:

```
PROGRAM START_BUS
```

```

EXTERNAL SYS$LAVC_START_BUS
INTEGER*4 SYS$LAVC_START_BUS
INTEGER*4 STATUS

STATUS = SYS$LAVC_START_BUS ( 'XQA0:' )

CALL SYS$EXIT ( %VAL ( STATUS ) )

END

```

E.2.1. Status

The SYS\$LAVC_START_BUS subroutine returns a status value in register R0, as described in Table E.2.

Table E.2. SYS\$LAVC_START_BUS Status

Status	Result
Success	Indicates that PEDRIVER is attempting to start the NISCA protocol on the specified adapter.
Failure	Indicates that PEDRIVER cannot start the protocol on the specified LAN adapter.

E.2.2. Error Messages

SYS\$LAVC_START_BUS can return the error condition codes shown in the following table.

Condition Code	Description
SS\$_ACCVIO	This status is returned for the following conditions: <ul style="list-style-type: none"> No access to the argument list No access to the LAN adapter name buffer descriptor No access to the LAN adapter name buffer
SS\$_DEVACTIONE	Bus already exists. PEDRIVER is already trying to use this LAN adapter for the NISCA protocol.
SS\$_INSFARG	Not enough arguments supplied.
SS\$_INSMEM	Insufficient nonpaged pool to create the bus data structure.
SS\$_INVBUSNAM	Invalid bus name specified. The device specified does not represent a LAN adapter that can be used for the protocol.
SS\$_IVBUFLN	This status value is returned under the following conditions: <ul style="list-style-type: none"> The LAN adapter name contains no characters (length = 0). The LAN adapter name contains more than 15 characters.
SS\$_NOSUCHDEV	This status value is returned under the following conditions: <ul style="list-style-type: none"> The LAN adapter name specified does not correspond to a LAN device available to PEDRIVER on this system.

Condition Code	Description
	<ul style="list-style-type: none"> No LAN drivers are loaded in this system; the value for NET\$AR_LAN_VECTOR is 0. PEDRIVER is not initialized; PEDRIVER's PORT structure is not available. <p>Note: By calling this routine, an error-log message may be generated.</p>
SS\$_NOTNETDEV	PEDRIVER does not support the specified LAN device.
SS\$_SYSVERDIF	The specified LAN device's driver does not support the VCI interface version required by PEDRIVER.

PEDRIVER can return additional errors that indicate it has failed to create the connection to the specified LAN adapter.

E.3. Stopping the NISCA Protocol

The SYSSLAVC_STOP_BUS routine stops the NISCA protocol on a specific LAN adapter.

Caution: Stopping the NISCA protocol on all LAN adapters causes satellites to hang and could cause cluster systems to fail with a CLUEXIT bugcheck.

To use this routine, specify the parameter described in the following table.

Parameter	Description
BUS_NAME	String descriptor representing the LAN adapter name buffer, passed by reference. The LAN adapter name must consist of 15 characters or fewer.

Example: The following Fortran sample program shows how SYSSLAVC_STOP_BUS is used to stop the NISCA protocol on the LAN adapter XQB:

```
PROGRAM STOP_BUS

EXTERNAL SYSSLAVC_STOP_BUS
INTEGER*4 SYSSLAVC_STOP_BUS
INTEGER*4 STATUS

STATUS = SYSSLAVC_STOP_BUS ( 'XQB' )

CALL SYS$EXIT ( %VAL ( STATUS ) )

END
```

E.3.1. Status

The SYSSLAVC_STOP_BUS subroutine returns a status value in register R0, as described in Table E.3.

Table E.3. SYSSLAVC_STOP_BUS Status

Status	Result
Success	Indicates that PEDRIVER is attempting to shut down the NISCA protocol on the specified adapter.

Status	Result
Failure	Indicates that PEDRIVER cannot shut down the protocol on the specified LAN adapter. However, PEDRIVER performs the shutdown asynchronously, and there could be other reasons why PEDRIVER is unable to complete the shutdown.

When the LAVC\$STOP_BUS module executes successfully, the following device-attention entry is written to the system error log:

```
DEVICE ATTENTION...
NI-SCS SUB-SYSTEM...
FATAL ERROR DETECTED BY DATALINK...
```

In addition, the following hexadecimal values are written to the STATUS field of the entry:

```
First longword (00000001)
Second longword (00001201)
```

This error-log entry indicates expected behavior and can be ignored. However, if the first longword of the STATUS field contains a value other than hexadecimal value 00000001, an error has occurred and further investigation may be necessary.

E.3.2. Error Messages

SYSS\$LAVC_STOP_BUS can return the error condition codes shown in the following table.

Condition Code	Description
SS\$_ACCVIO	This status is returned for the following conditions: <ul style="list-style-type: none"> No access to the argument list No access to the LAN adapter name buffer descriptor No access to the LAN adapter name buffer
SS\$_INVBUSNAM	Invalid bus name specified. The device specified does not represent a LAN adapter that can be used for the NISCA protocol.
SS\$_IVBUFLN	This status value is returned under the following conditions: <ul style="list-style-type: none"> The LAN adapter name contains no characters (length = 0). The LAN adapter name has more than 15 characters.
SS\$_NOSUCHDEV	This status value is returned under the following conditions: <ul style="list-style-type: none"> The LAN adapter name specified does not correspond to a LAN device that is available to PEDRIVER on this system. No LAN drivers are loaded in this system. NET\$AR_LAN_VECTOR is zero. PEDRIVER is not initialized. PEDRIVER's PORT structure is not available.

E.4. Creating a Representation of a Network Component

The SYSSLAVC_DEFINE_NET_COMPONENT subroutine creates a representation for a physical network component.

Use the following format to specify the parameters:

```
STATUS = SYSSLAVC_DEFINE_NET_COMPONENT (
component_description,
nodename_length,
component_type,
lan_hardware_addr,
lan_decnet_addr,
component_id_value)
```

Table E.4 describes the SYSSLAVC_DEFINE_NET_COMPONENT parameters.

Table E.4. SYSSLAVC_DEFINE_NET_COMPONENT Parameters

Parameter	Description
component_description	Address of a string descriptor representing network component name buffer. The length of the network component name must be less than or equal to the number of COMP\$C_MAX_NAME_LEN characters.
nodename_length	Address of the length of the node name. This address is located at the beginning of the network component name buffer for COMP\$C_NODE types. You should use zero for other component types.
component_type	Address of the component type. These values are defined by \$PEMCOMPDEF, found in SYSSLIBRARY:LIB.MLB.
lan_hardware_addr	Address of a string descriptor of a buffer containing the component's LAN hardware address (6 bytes). You must specify this value for COMP\$C_ADAPTER types. For other component types, this value is optional.
lan_decnet_addr	String descriptor of a buffer containing the component's LAN DECnet address (6 bytes). This is an optional parameter for all component types.
component_id_value	Address of a longword that is written with the component ID value.

E.4.1. Status

If successful, the SYSSLAVC_DEFINE_NET_COMPONENT subroutine creates a COMP data structure and returns its ID value. This subroutine copies user-specified parameters into the data structure and sets the reference count to zero.

The component ID value is a 32-bit value that has a one-to-one association with a network component. Lists of these component IDs are passed to SYSSLAVC_DEFINE_NET_PATH to specify the components used when a packet travels from one node to another.

E.4.2. Error Messages

SYSSLAVC_DEFINE_NET_COMPONENT can return the error condition codes shown in the following table.

Condition Code	Description
SS\$_ACCVIO	<p>This status is returned for the following conditions:</p> <ul style="list-style-type: none"> • No access to the network component name buffer descriptor • No access to the network component name buffer • No access to the component's LAN hardware address if a nonzero value was specified • No access to the component's LAN DECnet address if a nonzero value was specified • No access to the lan_hardware_addr string descriptor • No access to the lan_decnet_addr string descriptor • No write access to the component_id_value address • No access to the component_type address • No access to the nodename_length address • No access to the argument list
SS\$_DEVACTIVE	Analysis program already running. You must stop the analysis by calling the SYSSLAVC_DISABLE_ANALYSIS before you define the network components and the network component lists.
SS\$_INSFARG	Not enough arguments supplied.
SS\$_INVCOMPTYPE	The component type is either 0 or greater than or equal to COMP\$_C_INVALID.
SS\$_IVBUFLN	<p>This status value is returned under the following conditions:</p> <ul style="list-style-type: none"> • The component name has no characters (length = 0). • Length of the component name is greater than COMP\$_C_MAX_NAME_LEN. • The node name has no characters (length = 0) and the component type is COMP\$_C_NODE. • The node name has more than 8 characters and the component type is COMP\$_C_NODE. • The lan_hardware_addr string descriptor has fewer than 6 characters. • The lan_decnet_addr has fewer than 6 characters.

E.5. Creating a Network Component List

The SYSS\$LAVC_DEFINE_NET_PATH subroutine creates a directed list of network components between two network nodes. A **directed list** is a list of all the components through which a packet passes as it travels from the failure analysis node to other nodes in the cluster network.

Use the following format to specify the parameters:

```
STATUS = SYSS$LAVC_DEFINE_NET_PATH (
network_component_list,
used_for_analysis_status,
bad_component_id)
```

Table E.5 describes the SYSS\$LAVC_DEFINE_NET_PATH parameters.

Table E.5. SYSS\$LAVC_DEFINE_NET_PATH Parameters

Parameter	Description
network_component_list	<p>Address of a string descriptor for a buffer containing the component ID values for each of the components in the path. List the component ID values in the order in which a network message travels through them. Specify components in the following order:</p> <ol style="list-style-type: none"> 1. Local node 2. Local LAN adapter 3. Intermediate network components 4. Remote network LAN adapter 5. Remote node <p>You must list two nodes and two LAN adapters in the network path. The buffer length must be greater than 15 bytes and less than 509 bytes.</p>
used_for_analysis_status	Address of a longword status value that is written. This status indicates whether this network path has any value for the network failure analysis.
bad_component_id	Address of a longword value that contains the erroneous component ID if an error is detected while processing the component list.

E.5.1. Status

This subroutine creates a directed list of network components that describe a specific network path. If SYSS\$LAVC_DEFINE_NET_PATH is successful, it creates a CLST data structure. If one node is the local node, then this data structure is associated with a PEDRIVER channel. In addition, the reference count for each network component in the list is incremented. If neither node is the local node, then the used_for_analysis_status address contains an error status.

The SYSSLAVC_DEFINE_NET_PATH subroutine returns a status value in register R0, as described in Table E.6, indicating whether the network component list has the correct construction.

Table E.6. SYSSLAVC_DEFINE_NET_PATH Status

Status	Result
Success	The used_for_analysis_status value indicates whether the network path is useful for network analysis performed on the local node.
Failure	If a failure status returned in R0 is SSS_INVCOMPID, the bad_component_id address contains the value of the bad_component_id found in the buffer.

E.5.2. Error Messages

SYSSLAVC_DEFINE_NET_PATH can return the error condition codes shown in the following table.

Condition Code	Description
SS\$_ACCVIO	This status value can be returned under the following conditions: <ul style="list-style-type: none"> No access to the descriptor or the network component ID value buffer No access to the argument list No write access to the used_for_analysis_status address No write access to the bad_component_id address
SS\$_DEVACTIVE	Analysis already running. You must stop the analysis by calling the SYSSLAVC_DISABLE_ANALYSIS function before defining the network components and the network component lists.
SS\$_INSFARG	Not enough arguments supplied.
SS\$_INVCOMPID	Invalid network component ID specified in the buffer. The bad_component_id address contains the failed component ID.
SS\$_INVCOMPLIST	This status value can be returned under the following conditions: <ul style="list-style-type: none"> Fewer than two nodes were specified in the node list. More than two nodes were specified in the list. The first network component ID was not a COMP\$C_NODE type. The last network component ID was not a COMP\$C_NODE type. Fewer than two adapters were specified in the list. More than two adapters were specified in the list.
SS\$_IVBUFLN	Length of the network component ID buffer is less than 16, is not a multiple of 4, or is greater than 508.

Condition Code	Description
SS\$ _RMTPATH	Network path is not associated with the local node. This status is returned only to indicate whether this path was needed for network failure analysis on the local node.

E.6. Starting Network Component Failure Analysis

The SYSSLAVC_ENABLE_ANALYSIS subroutine starts the network component failure analysis.

Example: The following is an example of using the SYSSLAVC_ENABLE_ANALYSIS subroutine:

```
STATUS = SYSSLAVC_ENABLE_ANALYSIS ( )
```

E.6.1. Status

This subroutine attempts to enable the network component failure analysis code. The attempt will succeed if at least one component list is defined.

SYSSLAVC_ENABLE_ANALYSIS returns a status in register R0.

E.6.2. Error Messages

SYSSLAVC_ENABLE_ANALYSIS can return the error condition codes shown in the following table.

Condition Code	Description
SS\$ _DEVOFFLINE	PEDRIVER is not properly initialized. ROOT or PORT block is not available.
SS\$ _NOCOMPLSTS	No network connection lists exist. Network analysis is not possible.
SS\$ _WASSET	Network component analysis is already running.

E.7. Stopping Network Component Failure Analysis

The SYSSLAVC_DISABLE_ANALYSIS subroutine stops the network component failure analysis.

Example: The following is an example of using SYSSLAVC_DISABLE_ANALYSIS:

```
STATUS = SYSSLAVC_DISABLE_ANALYSIS ( )
```

This subroutine disables the network component failure analysis code and, if analysis was enabled, deletes all the network component definitions and network component list data structures from nonpaged pool.

E.7.1. Status

SYSSLAVC_DISABLE_ANALYSIS returns a status in register R0.

E.7.2. Error Messages

SYSS\$LAVC_DISABLE_ANALYSIS can return the error condition codes shown in the following table.

Condition Code	Description
SS\$_DEVOFFLINE	PEDRIVER is not properly initialized. ROOT or PORT block is not available.
SS\$_WASCLR	Network component analysis already stopped.

Appendix F. Troubleshooting the NISCA Protocol

NISCA is the transport protocol responsible for carrying messages, such as disk I/Os and lock messages, across Ethernet LANs to other nodes in the cluster. The acronym NISCA refers to the protocol that implements an Ethernet network interconnect (NI) according to the System Communications Architecture (SCA).

Using the NISCA protocol, an OpenVMS software interface emulates the CI port interface, that is, the software interface is identical to that of the CI bus, except that data is transferred over a LAN or IP network. The NISCA protocol allows OpenVMS Cluster communication over the LAN or IP network without the need for any special hardware.

This appendix describes the NISCA transport protocol and provides troubleshooting strategies to help a network manager pinpoint network-related problems. Because troubleshooting hard component failures in the LAN is best accomplished using a LAN analyzer, this appendix also describes the features and setup of a LAN analysis tool.

Note

Additional troubleshooting information specific to the revised PEDRIVER is planned for the next revision of this manual.

F.1. How NISCA Fits into the SCA

The NISCA protocol is an implementation of the Port-to-Port Driver (PPD) protocol of the SCA.

F.1.1. SCA Protocols

As described in Chapter 2, the SCA is a software architecture that provides efficient communication services to low-level distributed applications (for example, device drivers, file services, network managers).

The SCA specifies a number of protocols for OpenVMS Cluster systems, including System Applications (SYSAP), System Communications Services (SCS), the Port-to-Port Driver (PPD), and the Physical Interconnect (PI) of the device driver and LAN adapter. Figure F.1 shows these protocols as interdependent levels that make up the SCA architecture. Figure F.1 shows the NISCA protocol as a particular implementation of the PPD layer of the SCA architecture.

Figure F.1. Protocols in the SCA Architecture

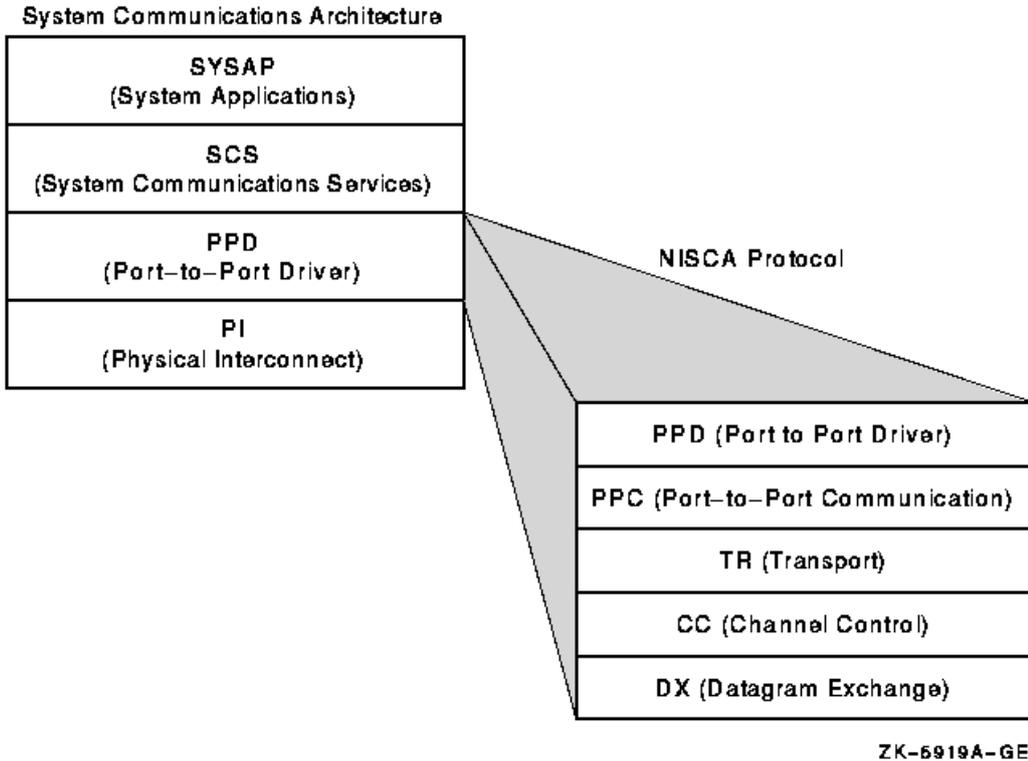


Table F.1 describes the levels of the SCA protocol shown in Figure F.1.

Table F.1. SCA Protocol Layers

Protocol	Description			
SYSAP	Represents clusterwide system applications that execute on each node. These system applications share communication paths in order to send messages between nodes. Examples of system applications are disk class drivers (such as DUDRIVER), the MSCP server, and the connection manager.			
SCS	Manages connections around the OpenVMS Cluster and multiplexes messages between system applications over a common transport called a virtual circuit (see Section F.1.2). The SCS layer also notifies individual system applications when a connection fails so that they can respond appropriately. For example, an SCS notification might trigger DUDRIVER to fail over a disk, trigger a cluster state transition, or notify the connection manager to start timing reconnect (RECNXINTERVAL) intervals.			
PPD	Provides a message delivery service to other nodes in the OpenVMS Cluster system.			
	<table border="1"> <thead> <tr> <th>PPD Level</th> <th>Description</th> </tr> </thead> <tbody> <tr> <td>Port-to-Port Driver (PPD)</td> <td>Establishes virtual circuits and handles errors.</td> </tr> </tbody> </table>	PPD Level	Description	Port-to-Port Driver (PPD)
PPD Level	Description			
Port-to-Port Driver (PPD)	Establishes virtual circuits and handles errors.			

Protocol	Description												
	Port-to-Port Communication (PPC)	Provides port-to-port communication, datagrams, sequenced messages, and block transfers. “Segmentation” also occurs at the PPC level. Segmentation of large blocks of data is done differently on a LAN than on a CI or a DSSI bus. LAN data packets are fragmented according to the size allowed by the particular LAN communications path, as follows:											
		<table border="1"> <thead> <tr> <th data-bbox="908 573 1206 647">Port-to-Port Communications</th> <th data-bbox="1211 573 1450 647">Packet Size Allowed</th> </tr> </thead> <tbody> <tr> <td data-bbox="908 654 1206 692">Ethernet-to-Ethernet</td> <td data-bbox="1211 654 1450 692">1498 bytes</td> </tr> <tr> <td data-bbox="908 698 1206 772">Gb Ethernet-to-Gb Ethernet</td> <td data-bbox="1211 698 1450 772">up to 8192 bytes</td> </tr> <tr> <td data-bbox="908 779 1206 853">Gb Ethernet-to-10Gb Ethernet</td> <td data-bbox="1211 779 1450 853">up to 8192 bytes</td> </tr> <tr> <td data-bbox="908 860 1206 934">10Gb Ethernet-to-10Gb Ethernet</td> <td data-bbox="1211 860 1450 934">up to 8192 bytes</td> </tr> </tbody> </table>	Port-to-Port Communications	Packet Size Allowed	Ethernet-to-Ethernet	1498 bytes	Gb Ethernet-to-Gb Ethernet	up to 8192 bytes	Gb Ethernet-to-10Gb Ethernet	up to 8192 bytes	10Gb Ethernet-to-10Gb Ethernet	up to 8192 bytes	Packet Size Allowed
		Port-to-Port Communications	Packet Size Allowed										
		Ethernet-to-Ethernet	1498 bytes										
		Gb Ethernet-to-Gb Ethernet	up to 8192 bytes										
		Gb Ethernet-to-10Gb Ethernet	up to 8192 bytes										
		10Gb Ethernet-to-10Gb Ethernet	up to 8192 bytes										
Ethernet-to-Ethernet	1498 bytes												
Gb Ethernet-to-Gb Ethernet	up to 8192 bytes												
Gb Ethernet-to-10Gb Ethernet	up to 8192 bytes												
10Gb Ethernet-to-10Gb Ethernet	up to 8192 bytes												
<p>Note: The default value is 1498 bytes for both Ethernet and FDDI.</p>													
Transport (TR)	Provides an error-free path, called a virtual circuit (see Section F.1.2), between nodes. The PPC level uses a virtual circuit for transporting sequenced messages and datagrams between two nodes in the cluster.												
Channel Control (CC)	Manages network paths, called channels, between nodes in an OpenVMS Cluster. The CC level maintains channels by sending HELLO datagram messages between nodes. A node sends a HELLO datagram messages to indicate it is still functioning. The TR level uses channels to carry virtual circuit traffic.												
Datagram Exchange (DX)	Interfaces to the LAN driver.												
PI	Provides connections to LAN devices. PI represents LAN drivers and adapters over which packets are sent and received.												
	PI Component	Description											
	LAN drivers	Multiplex NISCA and many other clients (such as DECnet, TCP/IP, LAT, LAD/LAS) and provide them with datagram services on Ethernet and FDDI network interfaces.											
	LAN adapters	Consist of the LAN network driver and adapter hardware.											

Figure F.2 shows the NISCA protocol as a particular implementation of the TCP/IP layer of the SCA architecture.

Figure F.2. Protocols in the SCA Architecture for Cluster over IP

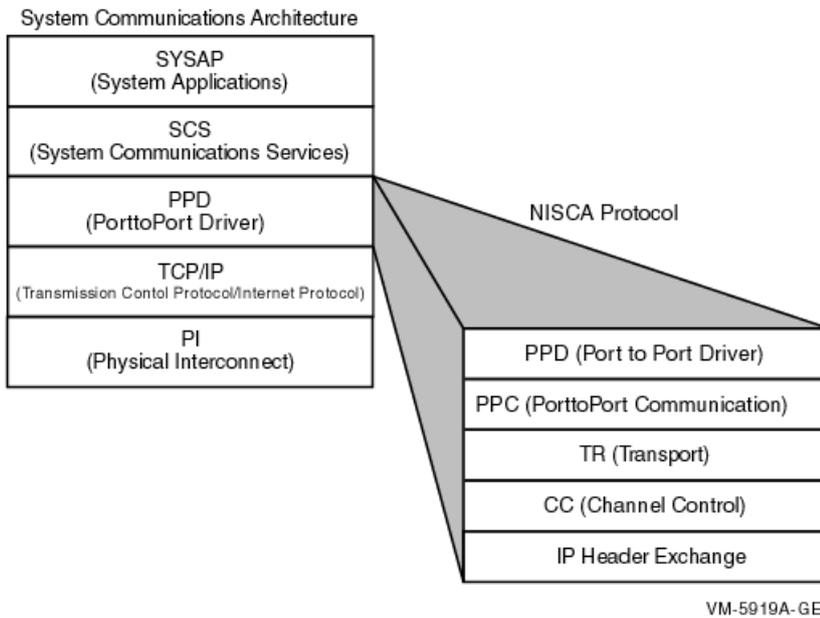


Table F.2 describes the levels of the SCA protocol shown in Figure F.2.

Table F.2. SCA Protocol Layers for Cluster over IP

Protocol	Description						
SYSAP	Represents clusterwide system applications that execute on each node. These system applications share communication paths in order to send messages between nodes. Examples of system applications are disk class drivers (such as DUDRIVER), the MSCP server, and the connection manager.						
SCS	Manages connections around the OpenVMS Cluster and multiplexes messages between system applications over a common transport called a virtual circuit (see Section F.1.2). The SCS layer also notifies individual system applications when a connection fails so that they can respond appropriately. For example, an SCS notification might trigger DUDRIVER to fail over a disk, trigger a cluster state transition, or notify the connection manager to start timing reconnect (RECNXINTERVAL) intervals.						
PPD	Provides a message delivery service to other nodes in the OpenVMS Cluster system.						
	<table border="1"> <thead> <tr> <th>PPD Level</th> <th>Description</th> </tr> </thead> <tbody> <tr> <td>Port-to-Port Driver (PPD)</td> <td>Establishes virtual circuits and handles errors.</td> </tr> <tr> <td>Port-to-Port Communication (PPC)</td> <td>Provides port-to-port communication, datagrams, sequenced messages, and block transfers. “Segmentation” also occurs at the PPC level. Segmentation of large blocks of data is done differently on a LAN than on a CI or a DSSI bus. LAN data packets are fragmented according to the size allowed by the particular LAN communications path, as follows:</td> </tr> </tbody> </table>	PPD Level	Description	Port-to-Port Driver (PPD)	Establishes virtual circuits and handles errors.	Port-to-Port Communication (PPC)	Provides port-to-port communication, datagrams, sequenced messages, and block transfers. “Segmentation” also occurs at the PPC level. Segmentation of large blocks of data is done differently on a LAN than on a CI or a DSSI bus. LAN data packets are fragmented according to the size allowed by the particular LAN communications path, as follows:
	PPD Level	Description					
Port-to-Port Driver (PPD)	Establishes virtual circuits and handles errors.						
Port-to-Port Communication (PPC)	Provides port-to-port communication, datagrams, sequenced messages, and block transfers. “Segmentation” also occurs at the PPC level. Segmentation of large blocks of data is done differently on a LAN than on a CI or a DSSI bus. LAN data packets are fragmented according to the size allowed by the particular LAN communications path, as follows:						

Protocol	Description	
	Port-to-Port Communications	Packet Size Allowed
	Ethernet-to-Ethernet	1498 bytes
	Gb Ethernet-to-Gb Ethernet	up to 8192 bytes
	Gb Ethernet-to-10Gb Ethernet	up to 8192 bytes
	10Gb Ethernet-to-10Gb Ethernet	up to 8192 bytes
	Note: The default value is 1498 bytes for both Ethernet and FDDI.	
Transport (TR)	Provides an error-free path, called a virtual circuit (see Section F.1.2), between nodes. The PPC level uses a virtual circuit for transporting sequenced messages and datagrams between two nodes in the cluster.	
Channel Control (CC)	Manages network paths, called channels, between nodes in an OpenVMS Cluster. The CC level maintains channels by sending HELLO datagram messages between nodes. A node sends a HELLO datagram messages to indicate it is still functioning. The TR level uses channels to carry virtual circuit traffic.	
IP header exchange	Interfaces to the TCP/IP stack.	
TCP/IP	Cluster over IP uses UDP for cluster communication	
PI	Provides connections to LAN devices. PI represents LAN drivers and adapters over which packets are sent and received.	
	PI Component	Description
	LAN drivers	Multiplex NISCA and many other clients (such as DECnet, TCP/IP, LAT, LAD/LAS) and provide them with datagram services on Ethernet and FDDI network interfaces.
LAN adapters	Consist of the LAN network driver and adapter hardware.	

F.1.2. Paths Used for Communication

The NISCA protocol controls communications over the paths described in Table F.3.

Table F.3. Communication Paths

Path	Description
Virtual circuit	A common transport that provides reliable port-to-port communication between OpenVMS Cluster nodes in order to:

Path	Description
	<ul style="list-style-type: none"> • Ensure the delivery of messages without duplication or loss, each port maintains a virtual circuit with every other remote port. • Ensure the sequential ordering of messages, virtual circuit sequence numbers are used on the individual packets. Each transmit message carries a sequence number; duplicates are discarded. <p>The virtual circuit descriptor table in each port indicates the status of its port-to-port circuits. After a virtual circuit is formed between two ports, communication can be established between SYSAPs in the nodes.</p>
Channel	A logical communication path between two LAN adapters located on different nodes. Channels between nodes are determined by the pairs of adapters and the connecting network. For example, two nodes, each having two adapters, could establish four channels. The messages carried by a particular virtual circuit can be sent over any of the channels connecting the two nodes.

Note: The difference between a channel and a virtual circuit is that channels provide a path for datagram service. Virtual circuits, layered on channels, provide an error-free path between nodes. Multiple channels can exist between nodes in an OpenVMS Cluster but only one virtual circuit can exist between any two nodes at a time.

F.1.3. PEDRIVER

The port emulator driver, PEDRIVER, implements the NISCA protocol and establishes and controls channels for communication between local and remote LAN ports.

PEDRIVER implements a packet delivery service (at the TR level of the NISCA protocol) that guarantees the sequential delivery of messages. The messages carried by a particular virtual circuit can be sent over any of the channels connecting two nodes. The choice of channel is determined by the sender (PEDRIVER) of the message. Because a node sending a message can choose any channel, PEDRIVER, as a receiver, must be prepared to receive messages over any channel.

At any point in time, the TR level uses single “preferred channel” to carry the traffic for a particular virtual circuit.

Starting with OpenVMS Version 8.3, the PEDRIVER also supports the following features:

- Data compression
- Multi-gigabit line speed and long distance performance scaling

Data compression can be used to reduce the time to transfer data between two OpenVMS nodes when the LAN speed between them is limiting the data transfer rate, and there is idle CPU capacity available. For example, it may be used to reduce shadow copy times. PEDriver data compression can be enabled by using SCACP, Availability Manager, or the NISCS_PORT_SERV sysgen parameter.

The number of packets in flight between nodes needs to increase proportionally to both the speed of LAN links and the inter-node distance. Historically, PEDriver had fixed transmit and receive windows (buffering capacity) of 31 outstanding packets. Beginning with OpenVMS Version 8.3, PEDriver now automatically selects transmit and receive window sizes (sometimes called pipe quota by other network protocols) based on the speed of the current set of local and remote LAN adapters being

used for cluster communications between nodes. Additionally, SCACP and Availability Manager now provide management override of the automatically-selected window sizes.

For more information, see the SCACP utility chapter, and NISCS_PORT_SERV in the *VSI OpenVMS System Management Utilities Reference Manual* and the *HP OpenVMS Availability Manager User's Guide*.

Reference: See Appendix G for more information about how transmit channels are selected.

F.2. Addressing LAN Communication Problems

This section describes LAN Communication Problems and how to address them.

F.2.1. Symptoms

Communication trouble in OpenVMS Cluster systems may be indicated by symptoms such as the following:

- Poor performance
- Console messages
 - “Virtual circuit closed” messages from PEA0 (PEDRIVER) on the console
 - “Connection loss” OPCOM messages on the console
 - CLUEXIT bugchecks
 - “Excessive packet losses on LAN Path” messages on the console
- Repeated loss of a virtual circuit or multiple virtual circuits over a short period of time (fewer than 10 minutes)

Before you initiate complex diagnostic procedures, do not overlook the obvious. Always make sure the hardware is configured and connected properly and that the network is started. Also, make sure system parameters are set correctly on all nodes in the OpenVMS Cluster.

F.2.2. Traffic Control

Keep in mind that an OpenVMS Cluster system generates substantially heavier traffic than other LAN protocols. In many cases, cluster behavior problems that appear to be related to the network might actually be related to software, hardware, or user errors. For example, a large amount of traffic does not necessarily indicate a problem with the OpenVMS Cluster network. The amount of traffic generated depends on how the users utilize the system and the way that the OpenVMS Cluster is configured with additional interconnects (such as DSSI and CI).

If the amount of traffic generated by the OpenVMS Cluster exceeds the expected or desired levels, then you might be able to reduce the level of traffic by:

- Shifting the user load between machines
- Adding LAN segments and reconfiguring the LAN connections across the OpenVMS Cluster system

F.2.3. Excessive Packet Losses on LAN Paths

Prior to OpenVMS Version 7.3, an SCS virtual circuit closure was the first indication that a LAN path had become unusable. In OpenVMS Version 7.3, whenever the last usable LAN path is losing packets at an excessive rate, PEDRIVER displays the following console message:

```
%PEA0, Excessive packet losses on LAN path from local-device-name
to device-name on REMOTE NODE node-name
```

This message is displayed when PEDRIVER recently had to perform an excessively high rate of packet retransmissions on the LAN path consisting of the local device, the intervening network, and the device on the remote node. The message indicates that the LAN path has degraded and is approaching, or has reached, the point where reliable communications with the remote node are no longer possible. It is likely that the virtual circuit to the remote node will close if the losses continue. Furthermore, continued operation with high LAN packet losses can result in significant loss in performance because of the communication delays resulting from the packet loss detection timeouts and packet retransmission.

The corrective steps to take are:

1. Check the local and remote LAN device error counts to see whether a problem exists on the devices. Issue the following commands on each node:

```
$ SHOW DEVICE local-device-name
$ MC SCACPSCACP> SHOW LAN device-name
$ MC LANCLANCP> SHOW DEVICE device-name/COUNTERS
```

2. If device error counts on the local devices are within normal bounds, contact your network administrators to request that they diagnose the LAN path between the devices.

F.2.4. Preliminary Network Diagnosis

If the symptoms and preliminary diagnosis indicate that you might have a network problem, troubleshooting LAN communication failures should start with the step-by-step procedures described in Appendix C. Appendix C helps you diagnose and solve common Ethernet and FDDILAN communication failures during the following stages of OpenVMS Cluster activity:

- When a computer or a satellite fails to boot
- When a computer fails to join the OpenVMS Cluster
- During run time when startup procedures fail to complete
- When a OpenVMS Cluster hangs

The procedures in Appendix C require that you verify a number of parameters during the diagnostic process. Because system parameter settings play a key role in effective OpenVMS Cluster communications, Section F.2.6 describes several system parameters that are especially important to the timing of LAN bridges, disk failover, and channel availability.

F.2.5. Tracing Intermittent Errors

Because PEDRIVER communication is based on channels, LAN network problems typically fall into these areas:

- Channel formation and maintenance

Channels are formed when HELLO datagram messages are received from a remote system. A failure can occur when the HELLO datagram messages are not received or when the channel control message contains the wrong data.

- Retransmission

A well-configured OpenVMS Cluster system should not perform excessive retransmissions between nodes. Retransmissions between any nodes that occur more frequently than once every few seconds deserve network investigation.

Diagnosing failures at this level becomes more complex because the errors are usually intermittent. Moreover, even though PEDRIVER is aware when a channel is unavailable and performs error recovery based on this information, it does not provide notification when a channel failure occurs; PEDRIVER provides notification only for virtual circuit failures.

However, the Local Area OpenVMS Cluster Network Failure Analysis Program (LAVC \$FAILURE_ANALYSIS), available in SYS\$EXAMPLES, can help you use PEDRIVER information about channel status. The LAVC\$FAILURE_ANALYSIS program (documented in Appendix D) analyzes long-term channel outages, such as hard failures in LAN network components that occur during run time.

This program uses tables in which you describe your LAN hardware configuration. During a channel failure, PEDRIVER uses the hardware configuration represented in the table to isolate which component might be causing the failure. PEDRIVER reports the suspected component through an OPCOM display. You can then isolate the LAN component for repair or replacement.

Reference: Section F.8 addresses the kinds of problems you might find in the NISCA protocol and provides methods for diagnosing and solving them.

F.2.6. Checking System Parameters

Table F.4 describes several system parameters relevant to the recovery and failover time limits for LANs in an OpenVMS Cluster.

Table F.4. System Parameters for Timing

Parameter	Use
RECNXINTERVAL	
Defines the amount of time to wait before removing a node from the OpenVMS Cluster after detection of a virtual circuit failure, which could result from a LAN bridge failure.	If your network uses multiple paths and you want the OpenVMS Cluster to survive failover between LAN bridges, make sure the value of RECNXINTERVAL is greater than the time it takes to fail over those paths. Reference: The formula for calculating this parameter is discussed in Section 3.2.10.
MVTIMEOUT	
Defines the amount of time the OpenVMS operating system tries to recover a path to a disk before returning failure messages to the application.	Relevant when an OpenVMS Cluster configuration is set up to serve disks over either the Ethernet or FDDI. MVTIMEOUT is similar to RECNXINTERVAL except that RECNXINTERVAL is CPU to CPU, and MVTIMEOUT is CPU to disk.
SHADOW_MBR_TIMEOUT	

Parameter	Use
Defines the amount of time that the Volume Shadowing for OpenVMS tries to recover from a transient disk error on a single member of a multiple-member shadow set.	SHADOW_MBR_TIMEOUT differs from MVTIMEOUT because it removes a failing shadow set member quickly. The remaining shadow set members can recover more rapidly once the failing member is removed.

Note: The TIMVCFAIL system parameter, which optimizes the amount of time needed to detect a communication failure, is not recommended for use with LAN communications. This parameter is intended for CI and DSSI connections. PEDRIVER (which is for Ethernet and FDDI) usually surpasses the detection provided by TIMVCFAIL with the listen timeout of 8 to 9 seconds.

F.2.7. Channel Timeouts

Channel timeouts are detected by PEDRIVER as described in Table F.5.

Table F.5. Channel Timeout Detection

PEDRIVER Actions	Comments
Listens for HELLO datagram messages, which are sent over channels at least once every 3 seconds	Every node in the OpenVMS Cluster multicasts HELLO datagram messages on each LAN adapter to notify other nodes that it is still functioning. Receiving nodes know that the network connection is still good.
Closes a channel when HELLO datagrams or sequenced messages have not been received for a period of 8 to 9 seconds	Because HELLO datagram messages are transmitted at least once every 3 seconds, PEDRIVER times out a channel only if at least two HELLO datagram messages are lost and there is no sequenced message traffic.
Closes a virtual circuit when: <ul style="list-style-type: none"> No channels are available. The packet size of the only available channels is insufficient. 	The virtual circuit is not closed if any other channels to the node are available except when the packet sizes of available channels are smaller than the channel being used for the virtual circuit. For example, if a channel fails over from FDDI to Ethernet, PEDRIVER may close the virtual circuit and then reopen it after negotiating the smaller packet size that is necessary for Ethernet segmentation.
Does not report errors when a channel is closed	OPCOM “Connection loss” errors or SYSAP messages are not sent to users or other system applications until after the virtual circuit shuts down. This fact is significant, especially if there are multiple paths to a node and a LAN hardware failure or IP network issue occurs. In this case, you might not receive an error message; PEDRIVER continues to use the virtual circuit over another available channel.
Reestablishes a virtual circuit when a channel becomes available again	PEDRIVER reopens a channel when HELLO datagram messages are received again.

F.3. Using SDA to Monitor LAN or IP Communications

This section describes how to use SDA to monitor LAN or IP communications.

F.3.1. Isolating Problem Areas

If your system shows symptoms of intermittent failures during runtime, you need to determine whether there is a network problem or whether the symptoms are caused by some other activity in the system.

Generally, you can diagnose problems in the NISCA protocol or the network using the OpenVMS System Dump Analyzer utility (SDA). SDA is an effective tool for isolating problems on specific nodes running in the OpenVMS Cluster system.

Reference: The following sections describe the use of some SDA commands and qualifiers. You should also refer to the *VSI OpenVMS System Analysis Tools Manual* or the *OpenVMS VAX System Dump Analyzer Utility Manual* for complete information about SDA for your system.

F.3.2. SDA Command SHOW PORT

The SDA command SHOW PORT provides relevant information that is useful in troubleshooting PEDRIVER and LAN adapters in particular. Begin by entering the SHOW PORT command, which causes SDA to define cluster symbols. Example F.1 illustrates how the SHOWPORT command provides a summary of OpenVMS Cluster data structures.

Example F.1. SDA Command SHOW PORT Display

```
$ ANALYZE/SYSTEM
SDA> SHOW PORT
VAXcluster data structures
-----

                --- PDT Summary Page ---

PDT Address          Type          Device          Driver Name
-----
80C3DBA0             pa           PAA0            PADRIVER
80C6F7A0             pe           PEA0            PEDRIVER
```

F.3.3. Monitoring Virtual Circuits

To examine information about the virtual circuit (VC) that carries messages between the local node (where you are running SDA) and another remote node, enter the SDA command SHOW PORT/VC=VC_remote-node-name. Example F.2 shows how to examine information about the virtual channel running between a local node and the remote node, NODE11.

Example F.2. SDA Command SHOW PORT/VC Display

```
SDA> SHOW PORT/VC=VC_NODE11
VAXcluster data structures
-----

                --- Virtual Circuit (VC) 98625380 ---
Remote System Name:  NODE11  (0:VAX)      Remote SCSSYSTEMID:  19583
Local System ID:    217 (D9)              Status: 0005 open,path
----- Transmit -----  ----- VC Closures -----  ●----- Congestion Control
-----
```

```

Msg Xmt②      46193196  SeqMsg TMO          0  Pipe Quota/Slo/Max③ 31/
7/31
  Unsequence      3  CC DFQ Empty        0  Pipe Quota Reached④
213481
  Sequence      41973703  Topology Change⑤    0  Xmt C/T⑥
0/1984
  ReXmt⑦         128/106  NPAGEDYN Low⑧       0  RndTrp uS⑨
18540+7764
  Lone ACK       4219362                UnAcked Msgs
0
Bytes Xmt     137312089                CMD Queue Len/Max
0/21
----- Receive ----- - Messages Discarded - ----- Channel Selection
-----
Msg Rcv⑩      47612604  No Xmt Chan         0  Preferred Channel
9867F400
  Unsequence      3  Rcv Short Msg       0  Delay Time
FAAD63E0
  Sequence      37877271  Illegal Seq Msg     0  Buffer Size
1424
  ReRcv⑪         13987  Bad Checksum        0  Channel Count
18
  Lone ACK       9721030  TR DFQ Empty        0  Channel Selections
32138
  Cache          314  TR MFQ Empty        0  Protocol
1.3.0
  Ill ACK        0  CC MFQ Empty        0  Open⑫ 8-FEB-1994
17:00:05.12
Bytes Rcv    3821742649  Cache Miss          0  Cls⑬ 17-NOV-1858
00:00:00.00

```

The SHOW PORT/VC=VC_remote-node-name command displays a number of performance statistics about the virtual circuit for the target node. The display groups the statistics into general categories that summarize such things as packet transmissions to the remote node, packets received from the remote node, and congestion control behavior. The statistics most useful for problem isolation are called out in Example F.2 and described below.

Note: The counters shown in Example F.2 are stored in fixed-size fields and are automatically reset to 0 when a field reaches its maximum value (or when the system is rebooted). Because fields have different maximum sizes and growth rates, the field counters are likely to reset at different times. Thus, for a system that has been running for a long time, some field values may seem illogical and appear to contradict others.

- ② Shows the total number of packets transmitted over the virtual circuit to the remote node, including both sequenced and unsequenced (channel control) messages, and lone acknowledgments. (All application data is carried in sequenced messages). The counters for sequenced messages and lone acknowledgments grow more quickly than most other fields.
- ⑦ Indicates the number of retransmissions and retransmit related timeouts for the virtual circuit.
 - The rightmost number (106) in the ReXmt field indicates the number of times a timeout occurred. A timeout indicates one of the following problems:
 - The remote system NODE11 did not receive the sequenced message sent by UPNVMS.
 - The sequenced message arrived but was delayed in transit to NODE11.

- The local system UPNVMS did not receive the acknowledgment to the message sent to remote node NODE11.
- The acknowledgment arrived but was delayed in transit from NODE11.

Congestion either in the network or at one of the nodes can cause the following problems:

- Congestion in the network can result in delayed or lost packets. Network hardware problems can also result in lost packets.
- Congestion in UPNVMS or NODE11 can result either in packet delay because of queuing in the adapter or in packet discard because of insufficient buffer space.
- The leftmost number (128) indicates the number of packets actually retransmitted. For example, if the network loses two packets at the same time, one timeout is counted but two packets are retransmitted. A retransmission occurs when the local node does not receive an acknowledgment for a transmitted packet within a predetermined timeout interval.

Although you should expect to see a certain number of retransmissions especially in heavily loaded networks, an excessive number of retransmissions wastes network bandwidth and indicates excessive load or intermittent hardware failure. If the leftmost value in the ReXmt field is greater than about 0.01% to 0.05% of the total number of the transmitted messages shown in the Msg Xmt field, the OpenVMS Cluster system probably is experiencing excessive network problems or local loss from congestion.

- ⑩ Indicates the total number of messages received by local node UPNVMS over this virtual circuit. The values for sequenced messages and lone acknowledgments usually increase at a rapid rate.
- ⑪ Displays the number of packets received redundantly by this system. A remote system may retransmit packets even though the local node has already successfully received them. This happens when the cumulative delay of the packet and its acknowledgment is longer than the estimated round-trip time being used as a timeout value by the remote node. Therefore, the remote node retransmits the packet even though it is unnecessary.

Underestimation of the round-trip delay by the remote node is not directly harmful, but the retransmission and subsequent congestion-control behavior on the remote node have a detrimental effect on data throughput. Large numbers indicate frequent bursts of congestion in the network or adapters leading to excessive delays. If the value in the ReRcv field is greater than approximately 0.01% to 0.05% of the total messages received, there maybe a problem with congestion or network delays.

- ⑤ Indicates the number of times PEDRIVER has performed a failover from FDDI to Ethernet, which necessitated closing and reopening the virtual circuit. In Example F.2, there have been no failovers. However, if the field indicates a number of failovers, a problem may exist on the FDDI ring.
- ⑥ Displays the number of times the virtual circuit was closed because of a pool allocation failure on the local node. If this value is nonzero, you probably need to increase the value of the NPAGEDYN system parameter on the local node.
- ① Displays information about the virtual circuit to control the pipe quota (the number of messages that can be sent to the remote node [put into the “pipe”] before receiving an acknowledgment and the retransmission timeout). PEDRIVER varies the pipe quota and the timeout value to control the amount of network congestion.

- ③ Indicates the current thresholds governing the pipe quota.
- The leftmost number (31) is the current value of the pipe quota (transmit window). After a timeout, the pipe quota is reset to 1 to decrease congestion and is allowed to increase quickly as acknowledgments are received.
 - The middle number (7) is the slow-growth threshold (the size at which the rate of increase is slowed) to avoid congestion on the network again.
 - The rightmost number (31) is the maximum value currently allowed for the VC based on channel limitations.

Reference: See Appendix G for PEDRIVER congestion control and channel selection information.

- ④ Indicates the number of times the entire transmit window was full. If this number is small as compared with the number of sequenced messages transmitted, it indicates that the local node is not sending large bursts of data to the remote node.
- ⑥ Shows both the number of successful transmissions since the last time the pipe quota was increased and the target value at which the pipe quota is allowed to increase. In the example, the count is 0 because the pipe quota is already at its maximum value (31), so successful transmissions are not being counted.
- ⑨ Displays values that are used to calculate the retransmission timeout in microseconds. The leftmost number (18540) is the average round-trip time, and the rightmost number (7764) is the average variation in round-trip time. In the example, the values indicate that the round trip is about 19 milliseconds plus or minus about 8 milliseconds.

VC round trip time values are dependent on the delayed ACK or the ACKholdoff delay, that is, 100 ms. The VC trip time is also dependent on the network traffic.

If there is sufficient cluster traffic, the receive window at the remote node gets filled and the ACK is delivered sooner.

If the cluster is idle with no traffic, there may be a delay of 100 ms to send the ACK. Hence, in an idle cluster with less traffic, the VC round trip delay value is normally high. As the traffic increases, the VC round trip time delay value drops.

Deviation/Variance: Whenever a new ACK delay is measured, it is compared with the current estimate of the ACK delay. The difference is a measure of the error in the delay estimate (delayError). This delayError is used as a correction to update the current estimate of ACK delay.

To prevent a "bad" measurement from estimate, the correction due to a single measurement is limited to a fraction.

The average of the absolute value of the delayError from the mean is used as estimation for the delay's variance.

- ⑫ Displays open (Open) and closed (CIs) timestamps for the last significant changes in the virtual circuit. The repeated loss of one or more virtual circuits over a short period of time (fewer than 10 minutes) indicates network problems.
- ⑬ If you are analyzing a crash dump, you should check whether the crash-dump time corresponds to the timestamp for channel closures (CIs).

F.3.4. Monitoring PEDRIVER for LAN devices

The SDA command PE LAN_DEVICE is useful for displaying PEDRIVER LAN device data. Each LAN device is a local LAN device on the system being used for NISCACP communications.

```
SDA> PE LAN_DEVICE
```

In the following example PE LAN_DEVICE displays the LAN device summary of I64MOZ

Example F.3. SDA Command PE LAN_DEVICE

```
SDA> PE LAN_DEVICE
```

```
PE$SDA Extension on I64MOZ (HP rx4640 (1.50GHz/6.0MB)) at 21-NOV-2008
15:43:12.53
```

```
-----
I64MOZ Device Summary 21-NOV-2008 15:43:12.53:
```

Total Device Bytes	Device Type	Line Speed	Buffer Size	MgtBuf SizeCap	Load Class	Mgt Priority	Current LAN Address
	Errors & Events	Status					
31126556	LCL	0	1426	0	0	0	00-00-00-00-00-00
5086238	EIA	100	1426	0	1000	0	00-30-6E-5D-97-AE
0	EIB	1000	1426	0	1000	0	00-30-6E-5D-97-AF
229120		Run	Online	Restart			

F.3.5. Monitoring PEDRIVER Buses for LAN Devices

The SDA command SHOW PORT/BUS=BUS_<LAN-device> command is useful for displaying the PEDRIVER representation of a LAN adapter. To PEDRIVER, a bus is the logical representation of the LAN adapter. (To list the names and addresses of buses, enter the SDA command SHOW PORT/ADDR=PE_PDT and then press the **Return** key twice.) Example F.4 shows a display for the LAN adapter named EXA.

Example F.4. SDA Command SHOW PORT/BUS Display

```
SDA> SHOW PORT/BUS=BUS_EXA
```

```
VAXcluster data structures
```

```
-----
--- BUS: 817E02C0 (EXA) Device: EX_DEMNA LAN Address: AA-00-04-00-64-4F
---
                                  LAN Hardware Address: 08-00-2B-2C-20-B5
Status: 00000803 run,online ①,restart
----- Transmit ----- Receive ----- Structure Addresses
-----
Msg Xmt          20290620  Msg Rcv          67321527  PORT Address
817E1140
Mcast Msgs      1318437    Mcast Msgs      39773666  VCIB Addr
817E0478
Mcast Bytes 168759936    Mcast Bytes 159660184  HELLO Message Addr
817E0508
```

```

Bytes Xmt      2821823510  Bytes Rcv      3313602089  BYE Message Addr
817E0698
Outstand I/Os      0  Buffer Size      1424  Delete BUS Rtn Adr
80C6DA46
Xmt Errors②      15896  Rcv Ring Size      31
Last Xmt Error 0000005C      Time of Last Xmt Error③ 21-JAN-1994
15:33:38.96
--- Receive Errors ----  ----- BUS Timer -----  ----- Datalink Events
-----
TR Mcast Rcv      0  Handshake TMO  80C6F070  Last  7-DEC-1992
17:15:42.18
Rcv Bad SCSID      0  Listen TMO      80C6F074  Last Event
00001202
Rcv Short Msg      0  HELLO timer      3  Port Usable
1
Fail CH Alloc      0  HELLO Xmt err ④  1623  Port Unusable
0
Fail VC Alloc      0  Address Change
1
Wrong PORT      0  Port Restart Fail
0

```

① Status

The Status line should always display a status of “online” to indicate that PEDRIVER can access its LAN adapter.

② Xmt Errors (transmission errors)

Indicates the number of times PEDRIVER has been unable to transmit a packet using this LAN adapter.

③ Time of Last Xmt Error

You can compare the time shown in this field with the Open and Cls times shown in the VC display in Example F.2 to determine whether the time of the LAN adapter failure is close to the time of a virtual circuit failure.

Note: Transmission errors at the LAN adapter bus level cause a virtual circuit breakage.

④ HELLO Xmt err (HELLO transmission error)

Indicates how many times a message transmission failure has “dropped” a PEDRIVER HELLO datagram message. (The Channel Control [CC] level description in Section F.1 briefly describes the purpose of HELLO datagram messages). If many HELLO transmission errors occur, PEDRIVER on other nodes probably is timing out a channel, which could eventually result in closure of the virtual circuit.

The 1623 HELLO transmission failures shown in Example F.4 contributed to the high number of transmission errors (15896). Note that it is impossible to have a low number of transmission errors and a high number of HELLO transmission errors.

F.3.6. Monitoring LAN Adapters

Use the SDA command SHOW LAN/COUNT to display information about the LAN adapters as maintained by the LAN device driver (the command shows counters for all protocols, not just

PEDRIVER [SCA] related counters). Example F.5 shows a sample display from the SHOW LAN/COUNTERS command.

Example F.5. SDA Command SHOW LAN/COUNTERS Display

```
$ ANALYZE/SYSTEM
SDA> SHOW LAN/COUNTERS
```

LAN Data Structures

```
-----
-- EXA Counters Information 22-JAN-1994 11:21:19 --

Seconds since zeroed      3953329      Station failures
0
Octets received          13962888501    Octets sent
11978817384
PDUs received            121899287      PDUs sent
76872280
Mcast octets received    7494809802      Mcast octets sent
183142023
Mcast PDUs received      58046934        Mcast PDUs sent
1658028
Unrec indiv dest PDUs    0              PDUs sent, deferred
4608431
Unrec mcast dest PDUs   0              PDUs sent, one coll
3099649
Data overruns           2              PDUs sent, mul coll
2439257
Unavail station buffs ①  0              Excessive collisions ②
5059
Unavail user buffers     0              Carrier check failure
0
Frame check errors       483           Short circuit failure
0
Alignment errors         10215         Open circuit failure
0
Frames too long          142           Transmits too long
0
Rcv data length error    0              Late collisions
14931
802E PDUs received       28546         Coll detect chk fail
0
802 PDUs received        0              Send data length err
0
Eth PDUs received        122691742     Frame size errors
0
```

LAN Data Structures

```
-----
-- EXA Internal Counters Information 22-JAN-1994 11:22:28 --

Internal counters address 80C58257      Internal counters size
24
Number of ports           0              Global page transmits
0
No work transmits         3303771       SVAPTE/BOFF transmits
0
```

```

Bad PTE transmits          0   Buffer_Adr transmits
0

Fatal error count          0   RDL errors
0

Transmit timeouts          0   Last fatal error
None

Restart failures          0   Prev fatal error
None

Power failures             0   Last error CSR
00000000

Hardware errors            0   Fatal error code
None

Control timeouts          0   Prev fatal error
None

Loopback sent              0   Loopback failures
0

System ID sent             0   System ID failures
0

ReqCounters sent          0   ReqCounters failures
0

```

-- EXA1 60-07 (SCA) Counters Information 22-JAN-1994 11:22:31 --

```

Last receive ③             22-JAN 11:22:31   Last transmit(3)      22-JAN
11:22:31

Octets received           7616615830   Octets sent
2828248622

PDUs received             67375315     PDUs sent

Mcast octets received     0       Mcast octets sent
0

Mcast PDUs received       0       Mcast PDUs sent
0

Unavail user buffer       0       Last start attempt
None

Last start done           7-DEC 17:12:29   Last start failed
None

.
.
.

```

The SHOW LAN/COUNTERS display usually includes device counter information about several LAN adapters. However, for purposes of example, only one device is shown in Example F.5.

❶ Unavail station buffs (unavailable station buffers)

Records the number of times that fixed station buffers in the LAN driver were unavailable for incoming packets. The node receiving a message can lose packets when the node does not have enough LAN station buffers. (LAN buffers are used by a number of consumers other than PEDRIVER, such as DECnet, TCP/IP, and LAT). Packet loss because of insufficient LAN station buffers is a symptom of either LAN adapter congestion or the system's inability to reuse the existing buffers fast enough.

❷ Excessive collisions

Indicates the number of unsuccessful attempts to transmit messages on the adapter. This problem is often caused by:

- A LAN loading problem resulting from heavy traffic (70% to 80% utilization) on the specific LAN segment.
- A component called a screamer. A **screamer** is an adapter whose protocol does not adhere to Ethernet or FDDI hardware protocols. A screamer does not wait for permission to transmit packets on the adapter, thereby causing collision errors to register in this field.

If a significant number of transmissions with multiple collisions have occurred, then OpenVMS Cluster performance is degraded. You might be able to improve performance either by removing some nodes from the LAN segment or by adding another LAN segment to the cluster. The overall goal is to reduce traffic on the existing LAN segment, thereby making more band width available to the OpenVMS Cluster system.

- ④ The difference in the times shown in the Last receive and Last transmit message fields should not be large. Minimally, the timestamps in these fields should reflect that HELLO datagram messages are being sent across channels every 3 seconds. Large time differences might indicate:
 - A hardware failure
 - Whether or not the LAN driver sees the NISCA protocol as being active on a specific LAN adapter

F.3.7. Monitoring PEDRIVER Buses for IP interfaces

The SDA command SHOW PORT/BUS=BUS_IP_interface command is useful for displaying the PEDRIVER representation of an IP interface. To PEDRIVER, a bus is the logical representation of the IP interface. (To list the names and addresses of buses, enter the SDA command SHOW PORT/ADDR=PE_PDT and then press the **Return** key twice). The following example shows a display for the IP interface named IE0. command.

Example F.6. SDA Command SHOW PORT/BUS=BUS_IP_interface

```
$ ANALYZE/SYSTEM
SDA> SHOW PORT/BUS=886C0010

VMScluster data structures
-----
--- BUS: 886C0010 (IE0) Device: IP IP Address: 16.138.182.6 ①
Status: 00004203 run,online,xmt_chaining_disabled ②
----- Transmit ----- Receive ----- Structure Addresses
---
Msg Xmt      2345987277 ③ Msg Rcv      2452130165 ④ PORT Address
8850B9B8
Mcast Msgs      0 Mcast Msgs      0 VCIB Addr
886C02A0
Mcast Bytes      0 Mcast Bytes      0 HELLO Message Addr
886C02A0
Bytes Xmt      3055474713 Bytes Rcv      3545255112 BYE Message Addr
886C05CC
Outstand I/Os      0 Buffer Size      1394 Delete BUS Rtn Adr
90AA2EC8
Xmt Errors ⑤      0 Rcv Ring Size      0
```

```

--- Receive Errors ----  ----- BUS Timer -----  ----- Datalink Events
-----
TR Mcast Rcv           0  Handshake TMO   00000000  Last 22-SEP-2008
 12:20:50.06
Rcv Bad SCSID         0  Listen TMO      00000000  Last Event
 00004002
Rcv Short Msgs       0  HELLO timer           6  Port Usable
 1
Fail CH Alloc         0  HELLO Xmt err       0  Port Unusable
 0
Fail VC Alloc         0                               Address Change
 0
Wrong PORT           0                               Port Restart Fail
 0

```

- ❶ Displays the IP address of the interface.
- ❷ The Status line should always display a status of "online" to indicate that PEDRIVER can access its IP interface.
- ❸ Shows the total number of packets transmitted over the virtual circuit to the remote node. It provides the Multicast (mcast) and Multicast bytes transmitted.
- ❹ Shows the total number of packets received over the virtual circuit from the remote node. It provides the Multicast (mcast) and Multicast bytes transmitted.
- ❺ Indicates the number of times PEDRIVER has been unable to transmit a packet using this IP interface.

F.3.8. Monitoring PEDRIVER Channels for IP Interfaces

The SDA command `SHOW PORT/Channel=Channel_IP_interface` command is useful for displaying the PEDRIVER representation of an IP interface. To the PEDRIVER, a channel is the logical communication path between two IP interfaces located on different nodes. (To list the names and addresses of channels created, enter the SDA command `SHOW SYMBOL CH_*` and then press the **Return** key.) The following example shows a display for the IP interface named IE0.

Example F.7. SDA Command SHOW PORT/CHANNEL Display

```

$ ANALYZE/SYSTEM
SDA> show port/channel=CH_OOTY_IE0_WE0
VMScluster data structures
-----
-- PEDRIVER Channel (CH:886C5A40) for Virtual Circuit (VC:88161A80) OOTY
--
State: 0004 open           Status: 6F
  path,open,xchndis,rmhwavld,tight,fast
                               ECS Status: Tight,Fast
BUS: 886BC010 (IE0)  Lcl Device: IP    Lcl IP Address: 16.138.182.6 1 ❶
Rmt BUS Name:  WE0  Rmt Device: IP    Rmt IP Address: 15.146.235.10 2 ❷
Rmt Seq #: 0004  Open:  4-OCT-2008 00:18:58.94  Close:  4-OCT-2008
 00:18:24.53

- Transmit Counters --- - Receive Counters ---- - Channel Characteristics
--

```

```

Bytes Xmt      745486312  Bytes Rcv      2638847244  Protocol Version
1.6.0
Msg Xmt        63803681  Msg Rcv        126279729  Supported Services
00000000
  Ctrl Msgs      569    Ctrl Msgs      565  Local CH Sequence #
0003
  Ctrl Bytes    63220    Ctrl Bytes    62804  Average RTT (usec)
5780.8

                                Mcast Msgs    106871  Buffer Size:
                                Mcast Bytes  11114584  Current
1394
- Errors ----- Remote
1394
Listen TMO      2  Short CC Msgs    0  Local
1394
TR ReXmt       605  Incompat Chan    0  Negotiated
1394
DL Xmt Errors   0  No MSCP Srvr    0  Priority
0
CC HS TMO      0  Disk Not Srvd   0  Hops
2
Bad Authorize   0  Old Rmt Seq#    0  Load Class
100
Bad ECO        0
                                Rmt TR Rcv Cache Size
64
Bad Multicast   0
                                Rmt DL Rcv Buffers
8
                                Losses
0
- Miscellaneous ----- - Buf Size Probing----- - Delay Probing
-----
Prv Lstn Timer   5  SP Schd Timeout  6  DP Schd Timeouts
0
Next ECS Chan  886C5A40  SP Starts        1  DP Starts
0
                                SP Complete      1  DP Complete
0
- Management ----- SP HS TMO          0  DP HS TMO
1
Mgt Priority     0  HS Remaining Retries  4
Mgt Hops        0  Last Probe Size   1395
Mgt Max Buf Siz 8110

```

- ❶ Displays the IP address of the local interface.
- ❷ Displays the IP address of the remote interface.

F.4. Using SCACP to Monitor Cluster Communications

The SCA Control Program (SCACP) utility is designed to monitor and manage cluster communications. It is derived from the Systems Communications Architecture (SCA), which defines the communications mechanisms that allow nodes in an OpenVMS Cluster system to cooperate.

SCA does the following:

- Governs the sharing of data between resources at the nodes.
- Binds together System Applications (SYSAPs) that run on different OpenVMS Alpha and Integrity server systems.

To invoke SCACP, enter the following command at the DCL prompt:

```
$ RUN SYS$SYSTEM:SCACP
```

SCACP displays the following prompt, at which you can enter SCACP commands using the standard rules of DCL syntax:

```
SCACP>
```

For more information about SCACP, see *VSI OpenVMS System Management Utilities Reference Manual*.

F.5. Troubleshooting NISCA Communications

F.5.1. Areas of Trouble

Sections F.6 and F.7 describe two likely areas of trouble for LAN networks: channel formation and retransmission. The discussions of these two problems often include references to the use of a LAN analyzer tool to isolate information in the NISCA protocol.

Reference: As you read about how to diagnose NISCA problems, you may also find it helpful to refer to Section F.8, which describes the NISCA protocol packet, and Section F.9, which describes how to choose and use a LAN network failure analyzer.

F.6. Channel Formation

Channel-formation problems occur when two nodes cannot communicate properly between LAN adapters.

F.6.1. How Channels Are Formed

Table F.6 provides a step-by-step description of channel formation.

Table F.6. Channel Formation

Step	Action
1	Channels are formed when a node sends a HELLO datagram from its LAN adapter to a LAN adapter on another cluster node. If this is a new remote LAN adapter address, or if the corresponding channel is closed, the remote node receiving the HELLO datagram sends a CCSTART datagram to the originating node after a delay of up to 2 seconds.
2	Upon receiving a CCSTART datagram, the originating node verifies the cluster password and, if the password is correct, the node responds with a VERF datagram and waits for up to 5 seconds for the remote node to send a VACK datagram. (VERF, VACK, CCSTART, and HELLO datagrams are described in Section F.8.5).

Step	Action
	<ul style="list-style-type: none"> • Is the remote node powered on? • Is the remote node booted? • Are the required network connections connected? • Do the cluster multicast datagrams pass through all of the required bridges in both directions? • Are the cluster group code and password values the same on all nodes?
2	<p>Check for dead channels by using SDA. The SDA command <code>SHOW PORT/CHANNEL/VC=VC_remote_node</code> can help you determine whether a channel ever existed; the command displays the channel's state.</p> <p>Reference: Refer to Section F.3 for examples of the <code>SHOW PORT</code> command. Section F.11.1 describes how to use a LAN analyzer to troubleshoot channel formation problems.</p>
3	<p>See also Appendix D for information about using the <code>LAVC\$FAILURE_ANALYSIS</code> program to troubleshoot channel problems.</p>

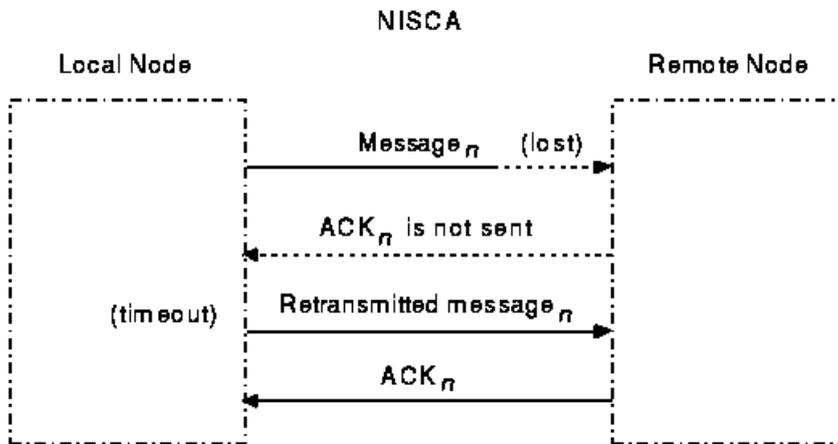
F.7. Retransmission Problems

Retransmissions occur when the local node does not receive acknowledgment of a message in a timely manner.

F.7.1. Why Retransmissions Occur

The first time the sending node transmits the datagram containing the sequenced message data, PEDRIVER sets the value of the REXMT flag bit in the TR header to 0. If the datagram requires retransmission, PEDRIVER sets the REXMT flag bit to 1 and resends the datagram. PEDRIVER retransmits the datagram until either the datagram is received or the virtual circuit is closed. If multiple channels are available, PEDRIVER attempts to retransmit the message on a different channel in an attempt to avoid the problem that caused the retransmission.

Retransmission typically occurs when a node runs out of a critical resource, such as large request packets (LRPs) or nonpaged pool, and a message is lost after it reaches the remote node. Other potential causes of retransmissions include overloaded LAN bridges, slow LAN adapters (such as the DELQA), and heavily loaded systems, which delay packet transmission or reception. Figure F.4 shows an unsuccessful transmission followed by a successful retransmission.

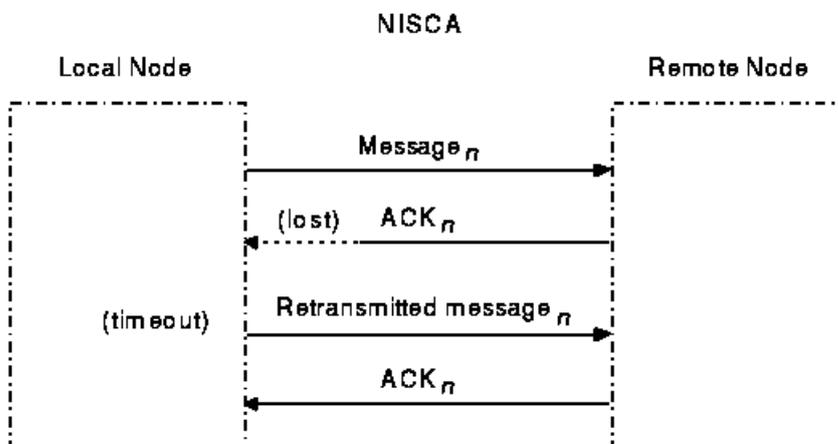
Figure F.4. Lost Messages Cause Retransmissions

Note: n represents a number that identifies each sequenced message.

ZK-5941A-GE

Because the first message was lost, the local node does not receive acknowledgment (ACK) from the remote node. The remote node acknowledged the second (successful) transmission of the message.

Retransmission can also occur if the cables are seated improperly, if the network is too busy and the datagram cannot be sent, or if the datagram is corrupted or lost during transmission either by the originating LAN adapter or by any bridges or repeaters. Figure F.5 illustrates another type of retransmission.

Figure F.5. Lost ACKs Cause Retransmissions

Note: n represents a number that identifies each sequenced message.

ZK-5942A-GE

In Figure F.5, the remote node receives the message and transmits an acknowledgment (ACK) to the sending node. However, because the ACK from the receiving node is lost, the sending node retransmits the message.

F.7.2. Techniques for Troubleshooting

You can troubleshoot cluster retransmissions using a LAN protocol analyzer for each LAN segment. If multiple segments are used for cluster communications, then the LAN analyzers need to support a distributed enable and trigger mechanism (see Section F.9).

Reference: Techniques for isolating the retransmitted datagram using a LAN analyzer are discussed in Section F.11.2. See also Appendix G for more information about congestion control and PEDRIVER message retransmission.

F.8. Understanding NISCA Datagrams

Troubleshooting NISCA protocol communication problems requires an understanding of the NISCA protocol packet that is exchanged across the OpenVMS Cluster system.

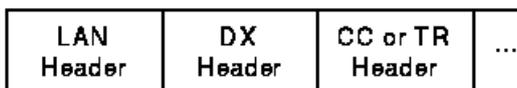
F.8.1. Packet Format

The format of packets on the NISCA protocol is defined by the \$NISCADEF macro, which is located in [DRIVER.LIS] on VAX systems and in [LIB.LIS] for Alpha systems on your CD listing disk.

Figure F.6 shows the general form of NISCA datagrams. A NISCA datagram consists of the following headers, which are usually followed by user data:

- LAN headers, including an Ethernet or an FDDI header
- Datagram exchange (DX) header
- Channel control (CC) or transport (TR) header

Figure F.6. NISCA Headers



ZK-5920A-GE

Caution: The NISCA protocol is subject to change without notice.

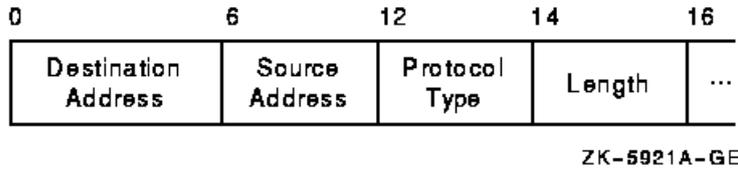
F.8.2. LAN Headers

The NISCA protocol is supported on LANs consisting of Ethernet, described in Section F.8.3. These headers contain information that is useful for diagnosing problems that occur between LAN adapters.

Reference: See Section F.10.4 for methods of isolating information in LAN headers.

F.8.3. Ethernet Header

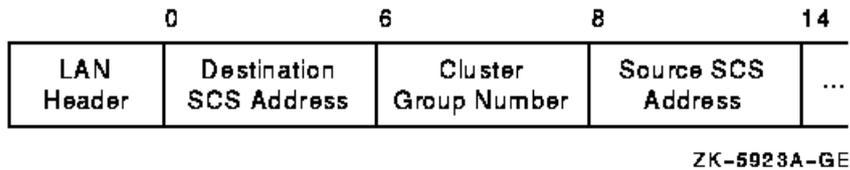
Each datagram that is transmitted or received on the Ethernet is prefixed with an Ethernet header. The Ethernet header, shown in Figure F.7 and described in Table F.7, is 16 bytes long.

Figure F.7. Ethernet Header**Table F.7. Fields in the Ethernet Header**

Field	Description
Destination address	LAN address of the adapter that should receive the datagram
Source address	LAN address of the adapter sending the datagram
Protocol type	NISCA protocol (60–07) hexadecimal
Length	Number of data bytes in the datagram following the length field

F.8.4. Datagram Exchange (DX) Header

The datagram exchange (DX) header for the OpenVMS Cluster protocol is used to address the data to the correct OpenVMS Cluster node. The DX header, shown in Figure F.8 and described in Table F.8, is 14 bytes long. It contains information that describes the OpenVMS Cluster connection between two nodes. See Section F.10.3 about methods of isolating data for the DX header.

Figure F.8. DX Header**Table F.8. Fields in the DX Header**

Field	Description
Destination SCS address	Manufactured using the address AA-00-04-00-- remote-node-SCSSYSTEMID. Append the remote node's SCSSYSTEMID system parameter value for the low-order 16 bits. This address represents the destination SCS transport address or the OpenVMS Cluster multicast address.
Cluster group number	The cluster group number specified by the system manager. See Chapter 8 for more information about cluster group numbers.
Source SCS address	Represents the source SCS transport address and is manufactured using the address AA-00-04-00-- local-node-SCSSYSTEMID. Append the local node's SCSSYSTEMID system parameter value as the low-order 16 bits.

F.8.5. Channel Control (CC) Header

The channel control (CC) message is used to form and maintain working network paths between nodes in the OpenVMS Cluster system. The important fields for network troubleshooting are the datagram flags/type and the cluster password. Note that because the CC and TR headers occupy the same space, there is a TR/CC flag that identifies the type of message being transmitted over the channel. Figure F.9 shows the portions of the CC header needed for network troubleshooting, and Table F.9 describes these fields.

Figure F.9. CC Header

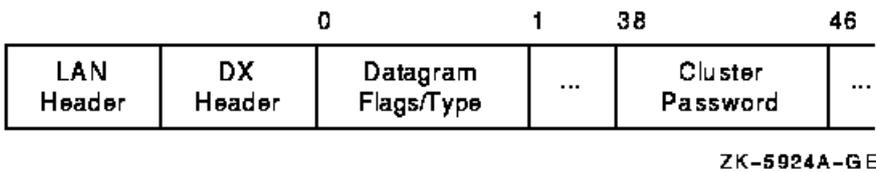


Table F.9. Fields in the CC Header

Field	Description			
Datagram type (bits <3:0>)	Identifies the type of message on the Channel Control level. The following table shows the datagrams and their functions.			
	Value	Abbreviated Datagram Type	Expanded Datagram Type	Function
	0	HELLO	HELLO datagram message	Multicast datagram that initiates the formation of a channel between cluster nodes and tests and maintains the existing channels. This datagram does not contain a valid cluster password.
	1	BYE	Node-stop notification	Datagram that signals the departure of a cluster node.
	2	CCSTART	Channel start	Datagram that starts the channel-formation handshake between two cluster nodes. This datagram is sent in response to receiving a HELLO datagram from an unknown LAN adapter address.
3	VERF	Verify	Datagram that acknowledges the CCSTART datagram and continues the	

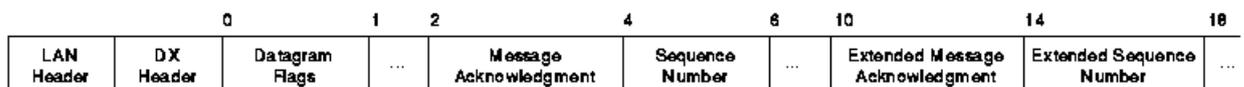
Field	Description			
				channel formation handshake. The datagram is sent in response to receiving a CCSTART or SOLICIT_SRV datagram.
	4	VACK	Verify acknowledge	Datagram that completes the channel-formation handshake. The datagram is sent in response to receiving a VERF datagram.
	5	Reserved		
	6	SOLICIT_SERVICE	Solicit	Datagram sent by a booting node to form a channel to its disk server. The server responds by sending a VERF, which forms the channel.
	7–15	Reserved		
Datagram flags (bits <7 : 4>)	Provide additional information about the control datagram. The following bits are defined:			
	<ul style="list-style-type: none"> • Bit <4> (AUTHORIZE)—Set to 1 if the cluster password field is valid. • Bit <5> (Reserved)—Set to 1. • Bit <6> (Reserved)—Set to 0. • Bit <7> (TR/CC flag)—Set to 1 to indicate the CC datagram. 			
Cluster password	Contains the cluster password.			

F.8.6. Transport (TR) Header

The transport (TR) header is used to pass SCS datagrams and sequenced messages between cluster nodes. The important fields for network troubleshooting are the TR datagram flags, message acknowledgment, and sequence numbers. Note that because the CC and TR headers occupy the same space, a TR/CC flag identifies the type of message being transmitted over the channel.

Figure F.10 shows the portions of the TR header that are needed for network troubleshooting, and Table F.10 describes these fields.

Figure F.10. TR Header



ZK-5825A-G E

Note: The TR header shown in Figure F.10 is used when both nodes are running Version 1.4 or later of the NISCA protocol. If one or both nodes are running Version 1.3 or an earlier version of the protocol, then both nodes will use the message acknowledgment and sequence number fields in place of the extended message acknowledgment and extended sequence number fields, respectively.

Table F.10. Fields in the TR Header

Field	Description			
Datagram flags (bits <7:0>)	Provide additional information about the transport datagram.			
	Value	Abbreviated Datagram Type	Expanded Datagram Type	Function
	0	DATA	Packet data	Contains data to be delivered to the upper levels of software.
	1	SEQ	Sequence flag	Set to 1 if this is a sequenced message and the sequence number is valid.
	2	Reserved		Set to 0.
	3	ACK	Acknowledgment	Acknowledges the field is valid.
	4	RSVP	Reply flag	Set when an ACK datagram is needed immediately.
	5	REXMT	Retransmission	Set for all retransmissions of a sequenced message.
	6	Reserved		Set to 0.
7	TR/CC flag	Transport flag	Set to 0; indicates a TR datagram.	
Message acknowledgment	An increasing value that specifies the last sequenced message segment received by the local node. All messages prior to this value are also acknowledged. This field is used when one or both nodes are running Version 1.3 or earlier of the NISCA protocol.			
Extended message acknowledgment	An increasing value that specifies the last sequenced message segment received by the local node. All messages prior to this value are also acknowledged. This field is used when both nodes are running Version 1.4 or later of the NISCA protocol.			
Sequence number	An increasing value that specifies the order of datagram transmission from the local node. This number is used to provide guaranteed delivery of this sequenced message segment to the remote node. This field is used when one or both nodes are running Version 1.3 or earlier of the NISCA protocol.			
Extended sequence number	An increasing value that specifies the order of datagram transmission from the local node. This number is used to provide guaranteed delivery of this sequenced message segment to the remote node. This field is used when both nodes are running Version 1.4 or later of the NISCA protocol.			

F.9. Using a LAN Protocol Analysis Program

Some failures, such as packet loss resulting from congestion, intermittent network interruptions of less than 20 seconds, problems with backup bridges, and intermittent performance problems, can be difficult to diagnose. Intermittent failures may require the use of a LAN analysis tool to isolate and troubleshoot the NISCA protocol levels described in Section F.1.

As you evaluate the various network analysis tools currently available, you should look for certain capabilities when comparing LAN analyzers. The following sections describe the required capabilities.

F.9.1. Single or Multiple LAN Segments

Whether you need to troubleshoot problems on a single LAN segment or on multiple LAN segments, a LAN analyzer should help you isolate specific patterns of data. Choose a LAN analyzer that can isolate data matching unique patterns that you define. You should be able to define data patterns located in the data regions following the LAN header (described in Section F.8.2). In order to troubleshoot the NISCA protocol properly, a LAN analyzer should be able to match multiple data patterns simultaneously.

To troubleshoot single or multiple LAN segments, you must minimally define and isolate transmitted and retransmitted data in the TR header (see Section F.8.6). Additionally, for effective network troubleshooting across multiple LAN segments, a LAN analysis tool should include the following functions:

- A **distributed enable** function that allows you to synchronize multiple LAN analyzers that are set up at different locations so that they can capture information about the same event as it travels through the LAN configuration
- A **distributed combination trigger** function that automatically triggers multiple LAN analyzers at different locations so that they can capture information about the same event

The purpose of distributed enable and distributed combination trigger functions is to capture packets as they travel across multiple LAN segments. The implementation of these functions discussed in the following sections use multicast messages to reach all LAN segments of the extended LAN in the system configuration. By providing the ability to synchronize several LAN analyzers at different locations across multiple LAN segments, the distributed enable and combination trigger functions allow you to troubleshoot LAN configurations that span multiple sites over several miles.

F.9.2. Multiple LAN Segments

To troubleshoot multiple LAN segments, LAN analyzers must be able to capture the multicast packets and dynamically enable the trigger function of the LAN analyzer, as follows:

Step	Action
1	Start capturing the data according to the rules specific to your LAN analyzer. VSI recommends that only one LAN analyzer transmit a distributed enable multicast packet on the LAN. The packet must be transmitted according to the media access-control rules.
2	Wait for the distributed enable multicast packet. When the packet is received, enable the distributed combination trigger function. Prior to receiving the distributed enable packet, all LAN analyzers must be able to ignore the trigger condition. This feature is required in order to set up multiple LAN analyzers capable of capturing the same

Step	Action
	event. Note that the LAN analyzer transmitting the distributed enable should not wait to receive it.
3	<p>Wait for an explicit (user-defined) trigger event or a distributed trigger packet. When the LAN analyzer receives either of these triggers, the LAN analyzer should stop the data capture.</p> <p>Prior to receiving either trigger, the LAN analyzer should continue to capture the requested data. This feature is required in order to allow multiple LAN analyzers to capture the same event.</p>
4	Once triggered, the LAN analyzer completes the distributed trigger function to stop the other LAN analyzers from capturing data related to the event that has already occurred.

The HP 4972A LAN Protocol Analyzer is one example of a network failure analysis tool that provides the required functions described in this section.

Reference: Section F.11 provides examples that use the HP 4972A LAN Protocol Analyzer.

F.10. Data Isolation Techniques

The following sections describe the types of data you should isolate when you use a LAN analysis tool to capture OpenVMS Cluster data between nodes and LAN adapters.

F.10.1. All OpenVMS Cluster Traffic

To isolate all OpenVMS Cluster traffic on a specific LAN segment, capture all the packets whose LAN header contains the protocol type 60–07.

Reference: See also Section F.8.2 for a description of the LAN headers.

F.10.2. Specific OpenVMS Cluster Traffic

To isolate OpenVMS Cluster traffic for a specific cluster on a specific LAN segment, capture packets in which:

- The LAN header contains the protocol type 60–07.
- The DX header contains the cluster group number specific to that OpenVMS Cluster.

Reference: See Sections F.8.2 and F.8.4 for descriptions of the LAN and DX headers.

F.10.3. Virtual Circuit (Node-to-Node) Traffic

To isolate virtual circuit traffic between a specific pair of nodes, capture packets in which the LAN header contains:

- The protocol type 60–07
- The destination SCS address
- The source SCS address

You can further isolate virtual circuit traffic between a specific pair of nodes to a specific LAN segment by capturing the following additional information from the DX header:

- The cluster group code specific to that OpenVMS Cluster
- The destination SCS transport address
- The source SCS transport address

Reference: See Sections F.8.2 and F.8.4 for LAN and DX header information.

F.10.4. Channel (LAN Adapter-to-LAN Adapter) Traffic

To isolate channel information, capture all packet information on every channel between LAN adapters. The DX header contains information useful for diagnosing heavy communication traffic between a pair of LAN adapters. Capture packets in which the LAN header contains:

- The destination LAN adapter address
- The source LAN adapter address

Because nodes can use multiple LAN adapters, specifying the source and destination LAN addresses may not capture all of the traffic for the node. Therefore, you must specify a channel as the source LAN address and the destination LAN address in order to isolate traffic on a specific channel.

Reference: See Section F.8.2 for information about the LAN header.

F.10.5. Channel Control Traffic

To isolate channel control traffic, capture packets in which:

- The LAN header contains the protocol type 60–07.
- The CC header datagram flags byte (the TR/CC flag, bit <7>) is set to 1.

Reference: See Sections F.8.2 and F.8.5 for a description of the LAN and CC headers.

F.10.6. Transport Data

To isolate transport data, capture packets in which:

- The LAN header contains the protocol type 60–07.
- The TR header datagram flags byte (the TR/CC flag, bit <7>) is set to 0.

Reference: See Sections F.8.2 and F.8.6 for a description of the LAN and TR headers.

F.11. Setting Up an HP 4972A LAN Protocol Analyzer

The HP 4972A LAN Protocol Analyzer is highlighted here because it meets all of the requirements listed in Section F.9. However, the HP 4972A LAN Protocol Analyzer is merely representative of the type of product useful for LAN network troubleshooting.

Note: Use of this particular product as an example here should not be construed as a specific purchase requirement or endorsement.

This section provides some examples of how to set up the HP 4972A LAN Protocol Analyzer to troubleshoot the local area OpenVMS Cluster system protocol for channel formation and retransmission problems.

F.11.1. Analyzing Channel Formation Problems

If you have a LAN protocol analyzer, you can set up filters to capture data related to the channel control header (described in Section F.8.5).

You can trigger the LAN analyzer by using the following datagram fields:

- Protocol type set to 60–07 hexadecimal
- Correct cluster group number
- TR/CC flag set to 1

Then look for the HELLO, CCSTART, VERF, and VACK datagrams in the captured data. The CCSTART, VERF, VACK, and SOLICIT_SRV datagrams should have the AUTHORIZE bit (bit <4>) set in the CC flags byte. Additionally, these messages should contain the scrambled cluster password (nonzero authorization field). You can find the scrambled cluster password and the cluster group number in the first four longwords of SYS\$SYSTEM:CLUSTER_AUTHORIZE.DAT file.

Reference: See Sections F.10.3 through F.10.5 for additional data isolation techniques.

F.11.2. Analyzing Retransmission Problems

Using a LAN analyzer, you can trace datagrams as they travel across an OpenVMS Cluster system, as described in Table F.11.

Table F.11. Tracing Datagrams

Step	Action
1	Trigger the analyzer using the following datagram fields: <ul style="list-style-type: none"> • Protocol type set to 60–07 • Correct cluster group number • TR/CC flag set to 0 • REXMT flag set to 1
2	Use the distributed enable function to allow the same event to be captured by several LAN analyzers at different locations. The LAN analyzers should start the data capture, wait for the distributed enable message, and then wait for the explicit trigger event or the distributed trigger message. Once triggered, the analyzer should complete the distributed trigger function to stop the other LAN analyzers capturing data.
3	Once all the data is captured, locate the sequence number (for nodes running the NISCA protocol Version 1.3 or earlier) or the extended sequence number (for nodes running the NISCA protocol Version 1.4 or later) for the datagram being retransmitted (the datagram with the REXMT flag set). Then, search through the previously captured data for another datagram between the same two nodes (not necessarily the same LAN adapters) with the following characteristics: <ul style="list-style-type: none"> • Protocol type set to 60–07 • Same DX header as the datagram with the REXMT flag set • TR/CC flag set to 0

Step	Action	
	<ul style="list-style-type: none"> • REXMT flag set to 0 • Same sequence number or extended sequence number as the datagram with the REXMT flag set 	
4	The following techniques provide a way of searching for the problem's origin.	
	IF...	THEN...
	The datagram appears to be corrupt	Use the LAN analyzer to search in the direction of the source node for the corruption cause.
	The datagram appears to be correct	Search in the direction of the destination node to ensure that the datagram gets to its destination.
	The datagram arrives successfully at its LAN segment destination	<p>Look for a TR packet from the destination node containing the sequence number (for nodes running the NISCA protocol Version 1.3 or earlier) or the extended sequence number (for nodes running the NISCA protocol Version 1.4 or later) in the message acknowledgment or extended message acknowledgement field. ACK datagram shave the following fields set:</p> <ul style="list-style-type: none"> • Protocol type set to 60–07 • Same DX header as the datagram with the REXMT flag set • TR/CC flag set to 0 • ACK flag set to 1
	The acknowledgment was not sent, or if a significant delay occurred between the reception of the message and the transmission of the acknowledgment	Look for a problem with the destination node and LAN adapter. Then follow the ACK packet through the network.
	The ACK arrives back at the node that sent the retransmission packet	<p>Either of the following conditions may exist:</p> <ul style="list-style-type: none"> • The retransmitting node is having trouble receiving LAN data. • The round-trip delay of the original datagram exceeded the estimated timeout value. <p>You can verify the second possibility by using SDA and looking at the ReRcv field of the virtual circuit display of the system receiving the retransmitted datagram.</p> <p>Reference: See Example F.2 for an example of this type of SDA display.</p>

Reference: See Appendix G for more information about congestion control and PEDRIVER message retransmission.

F.12. Filters

This section describes:

- How to use the HP 4972A LAN Protocol Analyzer filters to isolate packets that have been retransmitted or that are specific to a particular OpenVMS Cluster.
- How to enable the distributed enable and trigger functions.

F.12.1. Capturing All LAN Retransmissions for a Specific OpenVMS Cluster

Use the values shown in Table F.12 to set up a filter, named LAVc_TR_ReXMT, for all of the LAN retransmissions for a specific cluster. Fill in the value for the local area OpenVMS Cluster group code (nn-nn) to isolate a specific OpenVMS Cluster on the LAN.

Table F.12. Capturing Retransmissions on the LAN

Byte Number	Field	Value
1	DESTINATION	xx-xx-xx-xx-xx-xx
7	SOURCE	xx-xx-xx-xx-xx-xx
13	TYPE	60-07
23	LAVC_GROUP_CODE	nn-nn
31	TR FLAGS	0x1xxxxxx 2
33	ACKING MESSAGE	xx-xx
35	SENDING MESSAGE	xx-xx

F.12.2. Capturing All LAN Packets for a Specific OpenVMS Cluster

Use the values shown in Table F.13 to filter all of the LAN packets for a specific cluster. Fill in the value for OpenVMS Cluster group code (nn-nn) to isolate a specific OpenVMS Cluster on the LAN. The filter is named LAVc_all.

Table F.13. Capturing All LAN Packets (LAVc_all)

Byte Number	Field	Value
1	DESTINATION	xx-xx-xx-xx-xx-xx
7	SOURCE	xx-xx-xx-xx-xx-xx
13	TYPE	60-07
23	LAVC_GROUP_CODE	nn-nn
33	ACKING MESSAGE	xx-xx
35	SENDING MESSAGE	xx-xx

F.12.3. Setting Up the Distributed Enable Filter

Use the values shown in Table F.14 to set up a filter, named `Distrib_Enable`, for the distributed enable packet received event. Use this filter to troubleshoot multiple LAN segments.

Table F.14. Setting Up a Distributed Enable Filter (`Distrib_Enable`)

Byte Number	Field	Value	ASCII
1	DESTINATION	01-4C-41-56-63-45	.LAVcE
7	SOURCE	xx-xx-xx-xx-xx-xx	
13	TYPE	60-07	`.
15	TEXT	xx	

F.12.4. Setting Up the Distributed Trigger Filter

Use the values shown in Table F.15 to set up a filter, named `Distrib_Trigger`, for the distributed trigger packet received event. Use this filter to troubleshoot multiple LAN segments.

Table F.15. Setting Up the Distributed Trigger Filter (`Distrib_Trigger`)

Byte Number	Field	Value	ASCII
1	DESTINATION	01-4C-41-56-63-54	.LAVcT
7	SOURCE	xx-xx-xx-xx-xx-xx	
13	TYPE	60-07	`.
15	TEXT	xx	

F.13. Messages

This section describes how to set up the distributed enable and distributed trigger messages.

F.13.1. Distributed Enable Message

Table F.16 shows how to define the distributed enable message (`Distrib_Enable`) by creating a new message. You must replace the source address (`nn nn nn nn nn nn`) with the LAN address of the LAN analyzer.

Table F.16. Setting Up the Distributed Enable Message (`Distrib_Enable`)

Field	Byte Number	Value	ASCII
Destination	1	01 4C 41 56 63 45	.LAVcE
Source		nn nn nn nn nn nn	
Protocol	13	60 07	`.
Text	15	44 69 73 74 72 69 62 75 74 65	Distribute
	25	64 20 65 6E 61 62 6C 65 20 66	d enable f

Field	Byte Number	Value	ASCII
	35	6F 72 20 74 72 6F 75 62 6C 65	or trouble
	45	73 68 6F 6F 74 69 6E 67 20 74	shooting t
	55	68 65 20 4C 6F 63 61 6C 20 41	he Local A
	65	72 65 61 20 56 4D 53 63 6C 75	rea VMScLu
	75	73 74 65 72 20 50 72 6F 74 6F	ster Proto
	85	63 6F 6C 3A 20 4E 49 53 43 41	col: NISCA

F.13.2. Distributed Trigger Message

Table F.17 shows how to define the distributed trigger message (Distrib_Trigger) by creating a new message. You must replace the source address (nn nn nn nn nn nn) with the LAN address of the LAN analyzer.

Table F.17. Setting Up the Distributed Trigger Message (Distrib_Trigger)

Field	Byte Number	Value	ASCII
Destination	1	01 4C 41 56 63 54	.LAVcT
Source	7	nn nn nn nn nn nn	
Protocol	13	60 07	`.
Text	15	44 69 73 74 72 69 62 75 74 65	Distribute
	25	64 20 74 72 69 67 67 65 72 20	d trigger
	35	66 6F 72 20 74 72 6F 75 62 6C	for troubl
	45	65 73 68 6F 6F 74 69 6E 67 20	eshooting
	55	74 68 65 20 4C 6F 63 61 6C 20	the Local
	65	41 72 65 61 20 56 4D 53 63 6C	Area VMScL
	75	75 73 74 65 72 20 50 72 6F 74	uster Prot
	85	6F 63 6F 6C 3A 20 4E 49 53 43	ocol: NISC
	95	41	A

F.14. Programs That Capture Retransmission Errors

You can program the HP 4972 LAN Protocol Analyzer, as shown in the following source code, to capture retransmission errors. The starter program initiates the capture across all of the LAN analyzers. Only one LAN analyzer should run a copy of the starter program. Other LAN analyzers should run either the partner program or the scribe program. The partner program is used when the initial location of the error is unknown and when all analyzers should cooperate in the detection of the error. Use the scribe program to trigger on a specific LAN segment as well as to capture data from other LAN segments.

F.14.1. Starter Program

The starter program initially sends the distributed enable signal to the other LAN analyzers. Next, this program captures all of the LAN traffic, and terminates as a result of either a retransmitted packet detected by this LAN analyzer or after receiving the distributed trigger sent from another LAN analyzer running the partner program.

The starter program shown in the following example is used to initiate data capture on multiple LAN segments using multiple LAN analyzers. The goal is to capture the data during the same time interval on all of the LAN segments so that the reason for the retransmission can be located.

```
Store: frames matching LAVc_all
      or Distrib_Enable
      or Distrib_Trigger
      ending with LAVc_TR_ReXMT
      or Distrib_Trigger

Log file: not used

Block 1:  Enable_the_other_analyzers
          Send message Distrib_Enable
          and then
          Go to block 2

Block 2:  Wait_for_the_event
          When frame matches LAVc_TR_ReXMT then go to block 3

Block 3:  Send the distributed trigger
          Mark frame
          and then
          Send message Distrib_Trigger
```

F.14.2. Partner Program

The partner program waits for the distributed enable; then it captures all of the LAN traffic and terminates as a result of either a retransmission or the distributed trigger. Upon termination, this program transmits the distributed trigger to make sure that other LAN analyzers also capture the data at about the same time as when the retransmitted packet was detected on this segment or another segment. After the data capture completes, the data from multiple LAN segments can be reviewed to locate the initial copy of the data that was retransmitted. The partner program is shown in the following example:

```
Store: frames matching LAVc_all
      or Distrib_Enable
      or Distrib_Trigger
      ending with Distrib_Trigger

Log file: not used

Block 1:  Wait_for_distributed_enable
          When frame matches Distrib_Enable then go to block 2

Block 2:  Wait_for_the_event
          When frame matches LAVc_TR_ReXMT then go to block 3

Block 3:  Send the distributed trigger
          Mark frame
          and then
```

Send message Distrib_Trigger

F.14.3. Scribe Program

The scribe program waits for the distributed enable and then captures all of the LAN traffic and terminates as a result of the distributed trigger. The scribe program allows a network manager to capture data at about the same time as when the retransmitted packet was detected on another segment. After the data capture has completed, the data from multiple LAN segments can be reviewed to locate the initial copy of the data that was retransmitted. The scribe program is shown in the following example:

```
Store: frames matching LAVc_all
      or Distrib_Enable
      or Distrib_Trigger
      ending with Distrib_Trigger
```

Log file: not used

```
Block 1:  Wait_for_distributed_enable
          When frame matches Distrib_Enable then go to block 2
```

```
Block 2:  Wait_for_the_event
          When frame matches LAVc_TR_ReXMT then go to block 3
```

```
Block 3:  Mark_the_frames
          Mark frame
          and then
          Go to block 2
```

Appendix G. NISCA Transport Protocol Congestion Control

G.1. NISCA Congestion Control

Network congestion occurs as the result of complex interactions of workload distribution and network topology, including the speed and buffer capacity of individual hardware components.

Network congestion can have a negative impact on cluster performance in several ways:

- Moderate levels of congestion can lead to increased queue lengths in network components (such as adapters and bridges) that in turn can lead to increased latency and slower response.
- Higher levels of congestion can result in the discarding of packets because of queue overflow.
- Packet loss can lead to packet retransmissions and, potentially, even more congestion. In extreme cases, packet loss can result in the loss of OpenVMS Cluster connections.

At the cluster level, these congestion effects will appear as delays in cluster communications (e.g. delays of lock transactions, served I/Os, ICC messages, etc.). The user visible effects of network congestion can be application response sluggishness, or loss of throughput.

Thus, although a particular network component or protocol cannot guarantee the absence of congestion, the NISCA transport protocol implemented in PEDRIVER incorporates several mechanisms to mitigate the effects of congestion on OpenVMS Cluster traffic and to avoid having cluster traffic exacerbate congestion when it occurs. These mechanisms affect the retransmission of packets carrying user data and the multicast HELLO datagrams used to maintain connectivity.

G.1.1. Congestion Caused by Retransmission

Associated with each virtual circuit from a given node is a transmission window size, which indicates the number of packets that can be outstanding to the remote node (for example, the number of packets that can be sent to the node at the other end of the virtual circuit before receiving an acknowledgment [ACK]).

If the window size is 8 for a particular virtual circuit, then the sender can transmit up to 8 packets in a row but, before sending the ninth, must wait until receiving an ACK indicating that at least the first of the 8 has arrived.

If an ACK is not received, a timeout occurs, the packet is assumed lost, and must be retransmitted. If another timeout occurs for a retransmitted packet, the timeout interval is significantly increased and the packet is retransmitted again. After a large number of consecutive retransmissions of the same packet has occurred, the virtual circuit will be closed.

G.1.1.1. OpenVMS VAX Version 6.0 or OpenVMS AXP Version 1.5, or Later

This section pertains to PEDRIVER running on OpenVMS VAX Version 6.0 or OpenVMS AXP Version 1.5, or later.

The retransmission mechanism is an adaptation of the algorithms developed for the Internet TCP protocol by Van Jacobson and improves on the old mechanism by making both the window size and the retransmission timeout interval adapt to network conditions.

- When a timeout occurs because of a lost packet, the window size is decreased immediately to reduce the load on the network. The window size is allowed to grow only after congestion subsides. More specifically, when a packet loss occurs, the window size is decreased to 1 and remains there, allowing the transmitter to send only one packet at a time until all the original outstanding packets have been acknowledged.

After this occurs, the window is allowed to grow quickly until it reaches half its previous size. Once reaching the halfway point, the window size is allowed to increase relatively slowly to take advantage of available network capacity until it reaches a maximum value determined by the configuration variables (for example, a minimum of the number of adapter buffers and the remote node's resequencing cache).

- The retransmission timeout interval is set based on measurements of actual round-trip times, and the average variance from this average, for packets that are transmitted over the virtual circuit. This allows PEDRIVER to be more responsive to packet loss in most networks but avoids premature timeouts for networks in which the actual round-trip delay is consistently long. The algorithm can accommodate average delays of up to a few seconds.

G.1.2. HELLO Multicast Datagrams

PEDRIVER periodically multicasts a HELLO datagram over each network adapter attached to the node. The HELLO datagram serves two purposes:

- It informs other nodes of the existence of the sender so that they can form channels and virtual circuits.
- It helps to keep communications open once they are established.

HELLO datagram congestion and loss of HELLO datagrams can prevent connections from forming or cause connections to be lost. Table G.1 describes conditions causing HELLO datagram congestion and how PEDRIVER helps avoid the problems. The result is a substantial decrease in the probability of HELLO datagram synchronization and thus a decrease in HELLO datagram congestion.

Table G.1. Conditions that Create HELLO Datagram Congestion

Conditions that cause congestion	How PEDRIVER avoids congestion
If all nodes receiving a HELLO datagram from a new node responded immediately, the receiving network adapter on the new node could be overrun with HELLO datagrams and be forced to drop some, resulting in connections not being formed. This is especially likely in large clusters.	To avoid this problem on nodes running: <ul style="list-style-type: none"> • On VMS Version 5.5–2 or earlier, nodes that receive HELLO datagrams delay for a random time interval of up to 1 second before responding. • On OpenVMS VAX Version 6.0 or later, or OpenVMS AXP Version 1.5 or later, this random delay is a maximum of 2 seconds to support large OpenVMS Cluster systems.
If a large number of nodes in a network became synchronized and transmitted their HELLO	On nodes running VMS Version 5.5–2 or earlier, PEDRIVER multicasts HELLO datagrams over each adapter every 3 seconds, making HELLO datagram congestion more likely.

Conditions that cause congestion	How PEDRIVER avoids congestion
datagrams at or near the same time, receiving nodes could drop some datagrams and time out channels.	<p data-bbox="670 280 1441 589">On nodes running OpenVMS VAX Version 6.0 or later, or OpenVMS AXP Version 1.5 or later, PEDRIVER prevents this form of HELLO datagram congestion by distributing its HELLO datagram multicasts randomly over time. A HELLO datagram is still multicast over each adapter approximately every 3 seconds but not over all adapters at once. Instead, if a node has multiple network adapters, PEDRIVER attempts to distribute its HELLO datagram multicasts so that it sends a HELLO datagram over some of its adapters during each second of the 3-second interval.</p> <p data-bbox="670 622 1441 757">In addition, rather than multicasting precisely every 3 seconds, PEDRIVER varies the time between HELLO datagram multicasts between approximately 1.6 to 3 seconds, changing the average from 3 seconds to approximately 2.3 seconds.</p>

G.1.3. HELLO IP Unicast and IP Multicast Datagrams

PEDRIVER periodically transmits one IP multicast and one IP unicast for each of the IP multicast address. These unicast and multicast messages must be updated in the PE\$IP_CONFIG.DAT file. The HELLO datagrams serves two purposes:

- It informs other nodes of the existence of the sender so that they can form channels and virtual circuits.
- It helps to keep communications open once they are established.

HELLO datagram congestion and loss of HELLO datagrams can prevent connections from forming or causing connections to be lost.

