Microsoft Exchange on Digital Intel and Alpha Systems: Performance Report

This document characterizes the performance of Microsoft Exchange running the Microsoft defined medium user workload on Intel and Alpha systems from Digital Equipment Corporation.

Application Systems Engineering Performance Group. Digital Equipment Corporation
This group may be contacted online by sending mail to: asepg@digital.com

Revision/Update Information: Revision 1.2
This is a public document and contains updated
Performance information on DIGITAL's AlphaServer
and Prioris (Intel-based) Systems using Version 5 of
Microsoft Exchange

04-AUG-1997

TM

d i g i t a l

Digital Equipment Corporation
Maynard, Massachusetts

Microsoft Exchange V5 Software

Table of Contents

TABLES:

FIGURES:

1 OVERVIEW

This document contains information to help configure and size systems using versions 4 and 5 of Microsoft Exchange. This document presents information on configuring:

- Standalone systems

- Multi-server systems located at one site

- Sites connected by site connectors

- Sites connected by Internet Mail Connectors

- Sites connected by X.400

Software engineers from Digital Equipment Corporation's Application Systems Engineering Group (ASEPG) in Nashua, New Hampshire and DECwest Engineering in Bellevue, Washington performed the tests. The focus of testing was to find the number of Microsoft Exchange users supported on DIGITAL's Prioris (Intel-based) and AlphaServer 64-bit systems. The parameters defining the tests for a medium user load were set by Microsoft Corporation. All data reported here met Microsoft's criteria for a successful Load Simulator (LOADSIM) test.

There are important factors involved in setting up and configuring a system to get the best Exchange performance. Those factors will be presented and clarified in this document. You can then use the information in this document to set up a system configuration to meet all of your own unique needs.

The settings recommended in this document are meant to be used as guidelines only and are based on experimental results. Guidelines attempt to address specific circumstances with generalized information and should not override the realities of your own production environment. Consult with your DIGITAL sales or support representatives for help in setting up the best possible configuration to meet your needs.

2 EXECUTIVE SUMMARY

Performance testing for Microsoft Exchange Server Software Versions 4 and 5 was done using a range of DIGITAL Prioris (Intel-based) and AlphaServer systems. Key findings of the testing were:

- Microsoft Exchange memory requirements ranged from 32 to 256 megabytes. See Section 4.2 for more information on memory sizing.

- Cost effective configurations can be built using either Prioris or AlphaServer systems. Table 1 provides information on sizing systems with either platform based on user population.

- In version 4 or 5 of Microsoft Exchange, the Private Store mail capacity has a maximum capacity of 16 gigabytes per system on the server. This may limit the number of users that can be registered on a server.

- Systems using connectors should be configured with additional CPU power. Information on sizing when using connectors is provided in Section 3.2.3, *Connector Positioning*.

- Exchange Version 5 sees an increase in disk I/O over Version 4 and requires an increase in the number of disks for some user populations. Information on disk configurations for Version 4 and Version 5 are discussed in Section 4.3, *Disk I/O*.

- Given an adequate number of disks, Exchange Version 5 can support 3-8% more users than Exchange Version 4.

- The minimum disk configuration for small systems supporting up to 500 users is two disks. More disks may, however, be configured for higher levels of availability.

3 SYSTEM SIZING

Microsoft Exchange performance depends on the number of users who are actively working on the system. Performance will also be affected by any of the following operations:

- Amount and size of mail sent

- Amount of mail sent local or remote

- Number of accesses to the Exchange Private and Public databases

- Multi-server configuration

- Site or other type of connectors

3.1 Sizing by User Population

Use the information provided in this section to select a system solution based on your own needs.

Table 1 is intended to help size a system based on anticipated user population. This generalized data is based on test results using Microsoft specified test criteria. All systems should allow room for growth. For larger systems, the configuration should be designed individually to meet the unique needs of that environment.

Table 1: Sizing Table

# of Users	Single Server Organization	Multiple Server Organization
1-500	Prioris XL 6200 1 CPU (14%)* 64 Megabytes memory 2 Disks	Prioris MX 6200 1 CPU (15%) 96 Megabytes memory 2 Disks
200-1000	Prioris HX 6200 1 CPU (30%) 96 Megabytes memory 2-5 Disks or alternately an, AlphaServer 800 5/400 (40%) 96 Megabytes Memory 2-5 Disks	Prioris HX 6200 1 CPU (40%) 128 Megabytes memory 2-6 Disks or alternately an, AlphaServer 800 5/400 (50%) 128 Megabytes Memory 2-6 Disks
500-1500	Prioris HX 6200 1 CPU (50%) 128 Megabytes memory 3-8 Disks or alternately an, AlphaServer 800 5/500 (60%) 128 Megabytes Memory 3-8 Disks	Prioris HX 6200 2 CPU (45%) 160 Megabytes memory 3-10 Disks or alternately an, AlphaServer 800 5/500 (70%) 160 Megabytes Memory 3-10 Disks
1000-3100	Prioris ZX6200 2 CPU (56%) 96-256 Megabytes Memory 5-14 Disks or alternately an, AlphaServer 4000 5/466 1 CPU (55%) 96-256 Megabytes Memory 5-14 Disks	AlphaServer 4000 5/466 2 CPU (55%) 96-256 Megabytes Memory 5-12 Disks
* Figures in parentheses ( ) indicates CPU utilization for the top end of the range of users.		

3.2 Maximum Capacity of Specific Systems

This section contains a summary of the test results. Figures used here are based on the maximum number of Microsoft Exchange users (medium workload) attained during testing on the Prioris and AlphaServer systems.

The Microsoft specified criteria used to determine the number of users the servers could support was based on:
- The LOADSIM response time had to be under one second*.
- The send queue had to be empty by the end of the test.

The number of users supported by each server met both of the test criteria.

*Does not necessarily mean that all real users will experience sub-second response times.

Microsoft Exchange V5 Software 7

The following sections will help size a system based on the type of system, server and connectors used.

3.2.1 Single Server System

Figure 1 can be used as a guideline to determine the maximum number of users a particular system will support. These numbers were obtained from Exchange Version 4 testing. You can expect Exchange Version 5 numbers to be 3-8% higher. Remember, the number of users supported by system will vary based on your actual workload and could be higher or lower than the numbers presented here.

Figure 1: Maximum Number of Microsoft Exchange V4 Users — Single Server Systems

Maximum Number of Users
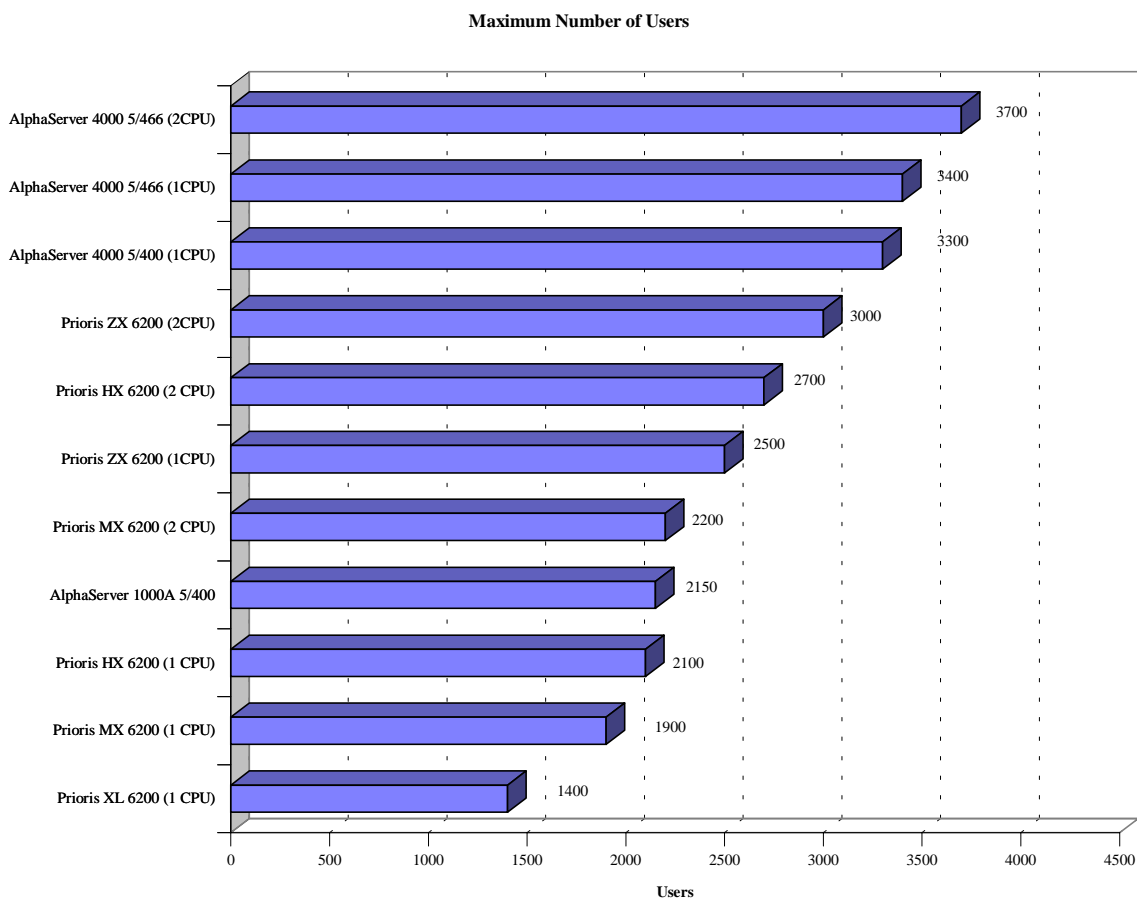
3.2.2 Multi-Server System

A single-server system is an isolated server supporting a number of clients. A multi-server system involves two or more Exchange Server computers in a single site with a replicated directory structure. Each server hosts a number of users. There is no sharing of databases, or resources. Microsoft Exchange automatically routes mail to the correct server within the site.

In these tests, the workload was configured with the default set to 50% of mail being sent to another server. This is considered to be a worst case setting, but is realistic if users are randomly being added to either individual server. Keeping departments, and/or work groups together on the same server may actually decrease the amount of mail being sent to another server and thereby increase the number of users supported.

Both Digital and Microsoft have observed significant performance differences between single-server and multi-server test runs. A drop in the number of users supported of as much as 40% can be common. This is caused by the additional work a server must do when sending mail to users on other servers (See highlighted note below). However, in a system with a workload mix where 80% of the mail recipients were on the same node and 20% were off-node, the drop in the number of users was only 15%.

See Table 2 for more information on the number of users supported with multi-server systems.

3.2.3 Connector Positioning

Configuring Site, X.400, or SMTP connectors will not add any appreciable load on the system. However, an increase in mail processing will directly increase the load on the system and decrease the number of users supported.

For example, in an environment where 80% of the mail recipients are on the same node, and 20% are remote, the number of messages being processed overall may increase by as much as 60% (see highlighted note below). As is the case with multi-server systems, this will result in an overall decrease in the number of users supported. Using Site Connectors with that same mail node mix would result in a decrease of 30% in the number of users supported on a single server. In the case of an X.400 Connector, the result is a decrease of 40% in the number of users supported.

Table 2 shows the number of users supported on a single server when multi-server, site, or X.400 connectors are used. The decrease in number of users is not cumulative. It is caused mainly by the increase in the amount of mail that is processed. In an environment where multi-servers are used with site connectors, a decrease in the number of users supported of up to 50% of the single server numbers.

Note: When all the recipients of a mail message are on your local node, MS Exchange will only create a single copy of that message and all recipients will receive a pointer to the message only. When all recipients have removed their pointers to the message, the message is finally deleted by the system. This is very efficient, especially with large distribution lists. If recipients are on another system, however, a copy will also be delivered to that system, thus increasing the load on the delivery mechanism of both nodes.

Deliveries made off-node can have a surprisingly large effect on the capacity of the system. For example, consider a system where all messages have exactly five local recipients. There will be exactly one copy of each message maintained by the system. But, if a random 20% of recipients are on another node, approximately 67% of messages will have both local and remote recipients and will have to be delivered to both systems, 33% will only have local recipients and a very small number will only have remote recipients. In this situation it is reasonable to assume that the other node is doing the same and 67% of the messages created there are also delivered to your node, imposing yet a further load on the system.

The work done in delivering the mail in this way is only a part of the total work done by the mail server, but it is a significant part. In versions 4 and 5 of Exchange, the send mechanism can be a limiting factor in the total capacity of the system.

Table 2: Number of Users by Server and Connectors

System	Server		Connector			Multi-Server & Connector
	Single	Multi- (15%)	Site (15%)	X.400 (20%)	IMC (25%)	
Prioris XL 6200	1400	1190	1190	1120	1050	980
Prioris MX 6200 (2 CPU)	2200	1870	1870	1760	1650	1540
Prioris ZX 6200 (1 CPU)	2500	2125	2125	2000	1875	1750
AlphaServer 1000A 5/400	2150	1827	1827	1720	1612	1505
Prioris ZX 6200 (2 CPU)	3000	2550	2550	2400	2250	2100
AlphaServer 4000 5/400 (1 CPU)	3300	2805	2805	2640	2475	2310
AlphaServer 4000 5/466 (2 CPU)	3700	2945	2945	2960	2775	2590
The percentage decreases (noted in parentheses) in the Server and Connector columns are based on Microsoft Exchange V5. Use these same percentage decreases when configuring other systems.						

4 SIZING COMPONENTS

4.1 CPU

CPU usage was not high, nor was CPU capacity ever exceeded for any of the systems (single or dual) being tested. In all cases, either the send queue or response times were outside of established test criteria. Using more than two CPUs on multiprocessor systems is not recommended at this time. Table 3 shows CPU usage by system tested.

Table 3: System Positioning for Microsoft Exchange Version 4

Users	System	CPU % Utilization
1400	Prioris XL 6200 (1 CPU)	32
1900	Prioris MX 6200 (1 CPU)	84
2100	Prioris HX 6200 (1 CPU)	74
2200	Prioris MX 6200 (2 CPU)	79
2500	Prioris ZX 6200 (1 CPU)	59
2700	Prioris HX 6200 (2 CPU)	77
3000	Prioris ZX 6200 (2 CPU)	52
2150	AlphaServer 1000A 5/400	72
3400	AlphaServer 4000 5/466 (1 CPU)	94
3700	AlphaServer 4000 5/466 (2 CPU)	77

4.2 Memory

Generally, Microsoft Exchange devotes memory to the working set sizes of the four processes, STORE, DSAMAIN, EMSMTA, and MAD. As messaging demands increase, working sets will grow. However, even at the highest loads tested, Microsoft Exchange did not use more than 256 megabytes of memory.

During installation and when running, the Microsoft Exchange Optimizer will indicate when there are not enough resources or if the system has less than 32 megabytes of memory. However, for 25 or fewer users, testing was done successfully using 16 megabytes of memory on an Intel-based Prioris system. For more information on configuring memory refer to Section 5.4, *Optimizer Utility*.

NT is a virtual memory management system. Those portions of memory that are not being used are put out to disk as long as they are not needed by a program. Paging on NT occurs primarily when demands for memory exceed available capacity. Even on servers with minimal memory configurations, and where committed bytes reported by the NT Performance Monitor was higher than the physical RAM amount, paging was minimal.

NOTE: When configuring a system with more than 256 megabytes of memory, set the Optimizer Utility to limit Exchange to 256 megabytes of memory or less. Microsoft Exchange creates Directory and Store buffers automatically based on the amount of memory the system is configured with. Creating more directory and store buffers than the Exchange system can use will not result in better system performance. Refer to Table 4 when sizing memory. More information on configuring memory and limiting the number of buffers can be found in Section 5.4, *Optimizer Utility*.

Table 4: Sizing Memory

Users	Memory (Megabytes)
25	16
100	32
101- 500	64
501 - 1000	96
1001 - 1500	128
1501 - 2000	160
2001 - 2500	192
2501 - 3000	224
3000+	256
Rule of thumb: After 500 users, add 32 megabytes for each additional 500 users	

4.3 Disk I/O

Disk I/O bandwidth is the single most important consideration for supporting the maximum number of Exchange users on a system. Testing has determined that Exchange Version 5 requires more database disks for user populations over 1200, than did Exchange Version 4.

The types of disks used by Exchange that need to be configured are:

1. System Disk

    This disk should hold only the Windows NT, PAGEFILE, and Exchange executable files. The system disk used for testing was a two gigabyte RZ28 disk.

2. Log File Disk

    Exchange uses a number of different log files. For heavily used systems, or systems that have large numbers of users, the log files should be placed on their own disk. In some test configurations, the log files were placed on the system disk. Reference Table 5 for detailed information on log file placement.

3. Exchange Database Disks

    Microsoft Exchange has two databases — public and private. The file names for both databases are PRIV.EDB and PUB.EDB. The private database (PRIV.EDB) is the private mail database and cannot be shared. More disks will improve database I/O and result in better performance of your system. Additional disks may decrease CPU utilization and enhance the performance of other applications residing on your system.

Table 5 shows how disks were configured for each of the systems tested. These disk configurations provided the best I/O capabilities for each of the systems and the number of users tested.

Table 5: Disk Configuration

# Users	IS Disks V4	IS Disks V5	Log Disks	System Disk	Total Disks V4	Total Disks V5
500	1	1	0	1	2	2
700	2	2	0	1	3	3
1000	3	3	1	1	5	5
1100	4	4	1	1	6	6
1200	4	5	1	1	6	7
1300	5	6	1	1	7	8
1400	5	7	1	1	7	9
1500	5	8	1	1	7	10
1700	6	9	1	1	8	11
1800	7	10	1	1	9	12
2000	7	11	1	1	9	13
2200	8	12	1	1	10	14
2500	9	12	1	1	11	14
2600	10	12	1	1	12	14
2700	10	13	1	1	12	15
3000	11	13	1	1	13	15
3100	custom	14	1	1	--	16

4.3.1 RAID

A rule of thumb to follow for sizing an Exchange Server disk configuration is for more than 1,000 users, a hardware RAID (Redundant Array of Independent Disks) controller should be used for the Information Store. Testing used multi-channel RAID controllers at the RAID-0 (striping) level. RAID controllers may have a limitation on the number of logical disks and the total number of disks that can be used as a stripe set. Refer to your RAID controller documentation for current information on limitations. Multiple controllers may also be used to maximize I/O bandwidth.

DIGITAL recommends the following RAID implementations for Microsoft Exchange:

- Implement RAID-1 (mirroring) for the log file disk.

- Implement RAID-5 for those disks containing database information for the highest levels of data availability and error recovery capability.

For more information on RAID configurations see Appendix A of this document.

4.4 Network

The network configuration used for Exchange testing was made up of a private LAN segment using either Ethernet ThinWire (10 megabit), or 10BaseT (Twisted Pair). The LAN connected the server and load simulator clients. While there are a number of events you can observe with the Performance Monitor Utility, the one primarily used during testing was the Network Interface bytes per second. Plan on 30 to 40 bytes per second per active user. With multi-server, and/or site connectors, plan for an additional 10 bytes per second per user.

Network usage will be higher if users commonly mail large documents (i.e. greater than 10 kilobytes in size). Therefore, monitor network traffic utilization if the mail workload is principally made up of large documents. As a general rule of thumb, a network can be saturated when it is at 50% or higher utilization. If network saturation occurs, adding another network card and LAN segment will improve performance.

5 INSTALLATION CONSIDERATIONS

This section is intended to ensure that when Microsoft Exchange is being installed it will function optimally. Check each of these items and, if appropriate, make changes to these settings.

5.1 Server System Setup

For better performance on systems with multiple PCI buses, place the disk controllers on one bus and the network controller on a second separate bus. CPU secondary cache should be set to ENABLED with WRITE BACK capability selected.

5.2 Operating System

Microsoft NT Version 3.51 with Service Pack 4 or NT Version 4.0 is required when using Microsoft Exchange. No screen savers should ever be used for a server. Screen savers use significant CPU resources and will adversely impact performance of the system.

The Server should be set to run continuously with minimum interaction. Therefore, set the event viewer to *"Overwrite Events as Needed"*. To change the setting, from CTRL-PNL-SYSTEM-TASKING, activate the *"Foregrnd & Backgrnd Events are Equally Responsive"* setting. Having events set to be equally responsive allows the system administrator to query the system if there is a problem.

5.3 Performance Monitor

The NT Performance Monitor should be run to log Exchange events and counters. Items that should be monitored are:

- Information store (private) send queue

- Average time for local delivery

- Number of users

- Any other events that are necessary for system operation.

Performance Monitor should be run at various times of the day, week, and so on. This will help give useful statistical data on system usage.

5.4 Optimizer Utility

Microsoft Exchange comes with an optimizer utility. This utility should be run during installation or when a change to the system occurs (adding memory, disks, and so on). The optimizer utility will set up Exchange for the proper number of buffers, threads, and so on, depending on the number of users specified. It will also analyze disk performance and make suggestions for file placement. The optimizer should always be used when the number of users changes or Microsoft Exchange components are to be moved to a different disk (the Optimizer Utility will move the components).

Microsoft Exchange creates directory and store buffers automatically based on the amount of memory the system is configured with, regardless of the number of users. If you have more than 256 megabytes of memory on your system, set the optimizer utility to limit Exchange to that amount or less. If your system is configured with more memory than specified in Table 4, set the optimizer values only to the values in that table.

NOTE: The Optimizer Utility sets to multi-server by default. If the installation is a single server, be sure to clear this option, otherwise performance will be adversely affected.
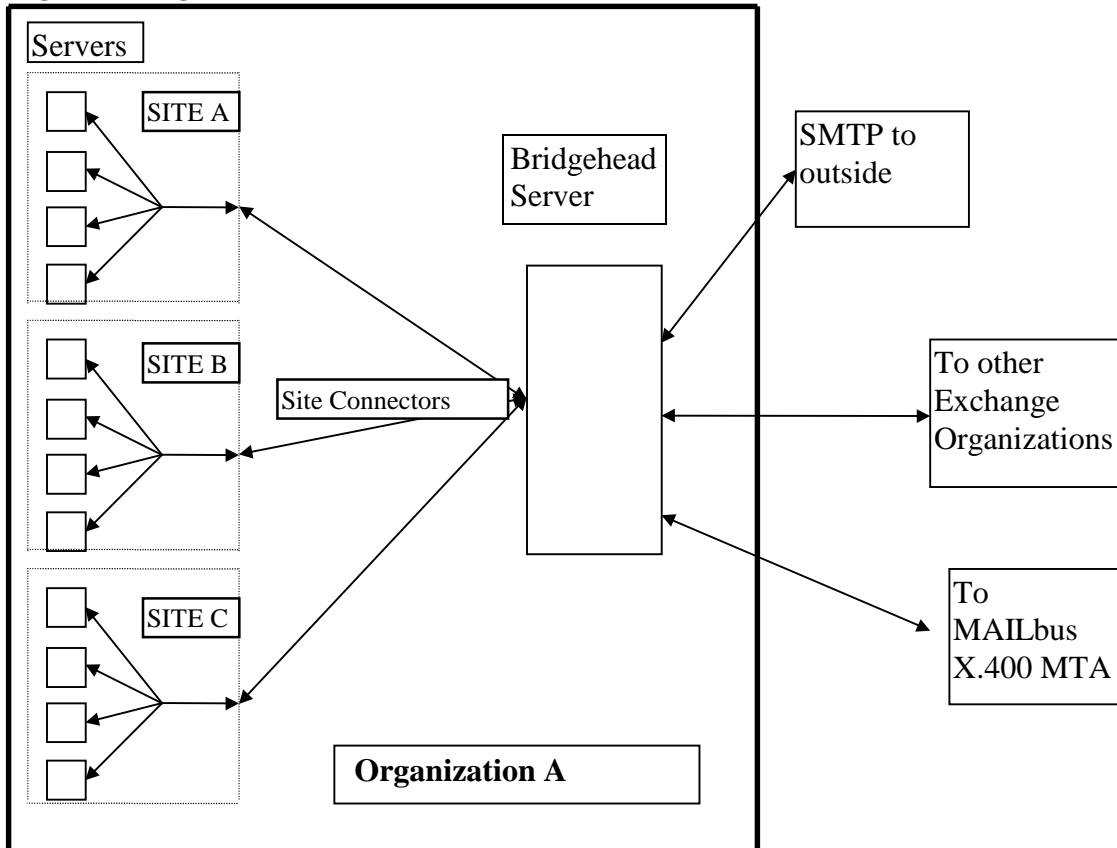
6 BRIDGEHEAD SERVER

Microsoft Exchange can be set up as a Bridgehead server. A Bridgehead server is a system set up to route messages from many servers between sites, organizations, and even companies. Use a bridgehead server if you have more than one connector on each server. Within a site, only one server should be set up to communicate off site.

6.1 System Setup

As a bridgehead server, the system has no users except for the administrator. The Microsoft Exchange optimizer is configured with minimal users, and site connector is selected.

Figure 2 is a representation of how Bridgehead servers can be utilized:

Figure 2: Bridgehead Server Utilization

The following picture shows why a bridgehead should be used. The figures represent five systems connected together using site connectors. Each system has an X.400 or IMC connector. If a bridgehead server was not used (left side), then each system would have at least four site connectors and directory replication connectors plus an X.400 or IMC connector.   By using a bridgehead, there is only one connector on any system and one directory replicator. Management of connectors can be done on one system instead of all the systems.

Figure 3: Systems With and Without a Bridgehead Server
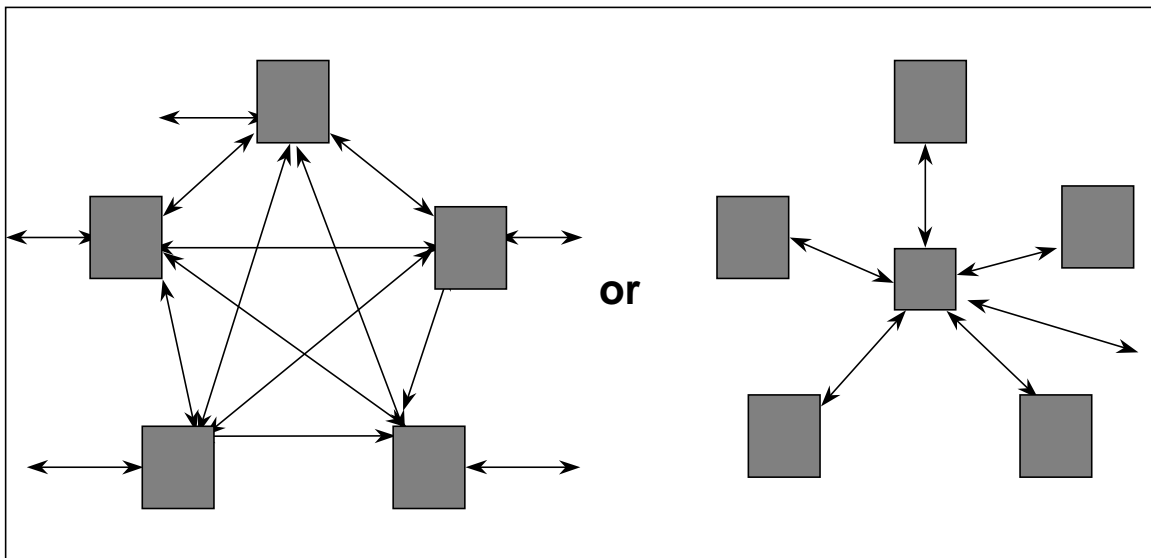
Note, in an actual implementation of a bridgehead server, each system would have a second connector connected to a backup bridgehead server. This could be another separate system or one of the existing systems. The secondary connector would be set up at a higher cost, and would only be used if something happened to the primary bridgehead server.

6.2 Results

Testing was done using a Prioris MX6200 single and dual processor for a low-end server and an AlphaServer 800 for a high-end server. Using the LOADSIM Medium workload with the 80/20 rule, where 80% of all message recipients are local and 20% is remote, 1,000 users would create remote message traffic equal to one message per second.

The size of messages used in these tests were the Load Simulator medium user message mix of from 1 to 60 kilobyte message sizes. Additional tests were done with 10 kilobyte, 40 kilobyte and 100 kilobyte documents. An increase (or decrease) in message size will result in higher (or lower) utilization rates.

6.3 Sizing

The single most important consideration in sizing a Bridgehead server is the network traffic expected on it. CPU, disk and memory resources are also considerations. This section presents details on resource utilization observed during testing.

6.3.1 CPU

Utilizing the Microsoft Load Simulator Medium workload with the 80/20 rule. One thousand users will generate one message per second through the Bridgehead server. Using Table 6 you can determine the type of processor needed for your workload by using the CPU percentage for messages per second (1,000 users = 1 message per second).

The following table shows the comparison between processors:

Table 6: CPU Percentage Usage for Messages per Second

	Prioris MX 6200 (1 CPU)	Prioris MX 6200 (2 CPU)	AlphaServer 800 5/400
Site Connector Users			
Normal	19,000	28,000	40,000
100K Msgs	11,300	16,500	23,300
CPU% per Msg/Sec			
Normal	5.1%	3.5%	2.47%
100K Msgs	8.82%	6.05%	4.28%
X.400 or IMC Users			
Normal	2,800	4,100	7,900
100K Msgs	2,180	3,180	6,000
CPU% per Msg/Sec			
Normal	35%	24%	12.65%
100K Msgs	45.8%	31.4%	16.62%

6.3.2 Memory

When configuring memory for a Bridgehead Server use the following table for sizing:

Table 7: Memory Selection for Bridgehead Servers

Low End Server*	High End Server
64 Megabytes	128 Megabytes
(3+ Disks)	(4+ Disks)
*Low End is defined as under 20,000 Site Connector users.	

6.3.3 Disks

All traffic through the Bridgehead is transient, so a lot of storage space is not required. Additional disks may be needed if connectors are configured based on a time delivery schedule.

6.3.4 Network

Depending on message size and distribution, the network can easily become saturated. However, multiple network adapters would typically be utilized in a Bridgehead server configuration.

7 WORKLOAD METHODOLOGY

Microsoft's Load Simulator was configured for a medium user workload using the settings recommended by Microsoft. These are the same settings that Microsoft uses for testing and characterization of system performance for multiple vendors systems.

When multiple disks were used, they were hardware striped on a RAID controller. RAID disk configurations allow for combining disks for performance improvement or disk failure resistance. RAID configurations can be implemented using I/O controllers or a disk management program residing on the host. A RAID-0 stripe set configuration was used for all testing.

RAID-5 was also tested. No performance degradation was observed using RAID-5. Note that additional disks may be required for a RAID-5 disk configuration.

For detailed information on the workload used for testing, refer to the *Microsoft Exchange Server White Paper* written by Microsoft. This workload sizing document is available from the Microsoft Web site (www.microsoft.com). The document provides a good description of the Load Simulator Utility.

Multiple Alpha and Intel Load Simulator client systems were used to drive the system under test to saturation levels. The tests were run for eight hours to verify that the response times and all measured events and counters were within acceptable limits.

The major selection criteria for establishing the maximum number of users per server was the Information Store send queue and response time. For a test to pass successfully, the rule was that the send queue had to be empty by the end of the test and response times to users had to be under one second.

APPENDIX A

The following is a brief explanation of the most common RAID implementations:

RAID-0 — Known as "striping". Two or more disks are logically combined into a stripe set. Data is written to all disks in the set in a sequential fashion; i.e. five megabytes is written to the first disk, then five megabytes is written to the next block on the next disk, and so on. Many of the reads and writes can be done in parallel. This technique results in faster reads and writes.

RAID-1 — Known as "mirroring". A disk or set of disks have an identical set of disks designated as their mirror image. This design improves read times. It also provides a redundant set of data should a disk failure occur. This also takes twice the amount of disk storage.

RAID-5 — A set of disks are created with one extra disk being added for parity purposes. Data is written redundantly across all disks. Any single disk failure can be tolerated. An additional 'hot' swap disk can be designated for one or more RAID-5 sets in the event of a disk failure. Performance is slower on writes and normal on reads.

Frequent failure resistant configurations and their tradeoffs are shown in Table 8:

Table 8: RAID Combinations

RAID Combination	General Performance Relative to Single Disk	Storage Requirements	Notes
RAID-0 + RAID-1	High on Reads, high on Writes	N* 2 (where N is the amount of storage required)	RAID-1 can split a heavy I/O load in half. In the event of failure, response is still adequate, but recovery is I/O intensive
RAID-5	High on Reads, slower on Writes	N + 1 disk. Add 1 more for 'hot' swap	Lowest cost. Failure results in slower performance, recovery is intensive.
RAID-1 + RAID-5	Average on Reads, slower on Writes	(N + 1 disk) * 2 Add 1 more disk for 'hot' swap	Dependable and splits I/O load. Double redundancy, failures are barely noticeable.