CLIENT/SERVER

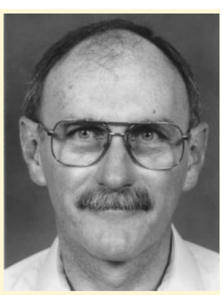
DCE Cell Performance: High Water Marks

By Bob Russell

Will DCE Security and Cell Directory Services handle the needs of a 10,000-user enterprise? This article discusses the tests and environments designed to address this performance and capacity question. This study has yielded some hardware and configuration high water marks that can be helpful in planning large-scale cell topologies.

Many large customers are preparing to consolidate their corporate networks into a single Distributed Computing Environment (DCE) cell. Others plan to do this under the administration of the IBM Warp Server (WS) using DCE Directory and Security Services (DSS). DSS will be based on the Open Software Foundation® (OSF®) DCE Version 1.1. Both DSS and the DCE-enabled client are in beta test and scheduled to ship in late 1995.

Recently, a large IBM customer, which we will call BigCo, decided to consolidate their enterprise under the IBM WS product. BigCo has a 10,000-user LAN environment. The capacity and performance limitations of the current LAN Server 4.0 (LS4) domain control and administration subsystems preclude expansion to a 10,000-user domain.



Bob Russell

In WS, DCE Security and Directory services will replace the LAN Server domain security and directory functions. BigCo's LAN Server 4.0 Domain Control Database (DCDB) and `NET.ACC` will be migrated into the DCE CDS namespace and the DCE Security registry. The existing clients do not need to be upgraded because the replacement of BigCo's current LS4 subsystems will be completely transparent to existing LAN Requesters.

This article evaluates the capacity and performance of the underlying DCE services used by WS relative to BigCo's planned 10,000-user environment.

Customer Requirements

BigCo, like many enterprises, is preparing to consolidate a large corporate network that consists of several LAN Server domains into a single DCE cell.

The core of BigCo's network is a 16 Mbits-per-second Token-Ring backbone with several local departmental LAN segments. A larger fiber-optic network ties the metropolitan area's LAN segments and several remote offices through 56 KB and T1 Wide Area Networks (WANs).

The heart of BigCo's network is a LAN segment containing their LAN Server domain controllers, database, print, file, and Lotus Notes® servers. BigCo was able to provide transactional and network utilization data about their current server LAN segment, which helped to structure a meaningful workload and lab environment.

BigCo Statistics

The following BigCo statistics are relevant to this study:

- BigCo provided graphs of the logon activity throughout the business day. The aggregation of the peak-hour activity for each of BigCo's five major LAN Server 4.0 domains is 1,534 logons per hour.
- The Sniffer analysis of BigCo's 16 Mbits-per-second server LAN segment shows their network is 38% to 43% utilized, with 1,332 to 1,453 frames per second and 567 to 587

bytes per frame. The network has 918 to 922 active network addresses.

Statistics from other LS4 customers were used to populate the DCE Security registry and CDS namespace. These statistics show an average of 50 users per domain in the LS4 security database (`NET.ACC`). The average population of the Domain Control Database (DCDB) is eight shared resources per domain and one subdirectory per user.

Considering these populations and BigCo's requirements, the initial population of the DCE test cell for each 1,000 users will be 1,000 accounts. DCE Security registry will have 20 groups of principals. The DCE Cell Directory namespace will contain 900 directories and 100 objects.

In reality, a pure WS and DSS environment will have a much different mixture of DCE directories and objects. Each WS domain controller will create only nine CDS directories. Since all aliases, names, and resources will be stored in these directories as DCE objects, these tests were conducted under the worst-case assumptions for memory, disk, and performance.

The DCE configuration program for DCE clients and DSS-enabled clients creates one directory and four objects on behalf of each configured client. Some hints for managing the resources and performance of the `/.:/hosts` directory tree will be discussed in the "DCE Cell Directory Services Performance" section.

Testing Approach

The system was measured in a fully loaded condition, running a modified version of our Point of Sale (POS) benchmark application. The average POS Remote Procedure Call (RPC) data size of 3,250 bytes was reduced so it did not reach the 16 Mbits-per-second Token-Ring capacity at throughput levels consistent with 10,000 users. The size was reduced by eliminating the POS Catalog RPC, which transfers 16,384 bytes of data. Eliminating the catalog also reduced the average number of RPCs per customer sale from 5.5 to 4.5. The data size was reduced to 329 bytes, which is more consistent with BigCo's 567 bytes per frame on the network. The network utilization in our tests ranged between 25% and 40%, which is consistent with the utilization reported by BigCo.

An optional function of the POS benchmark is to periodically perform a DCE Login. For these tests, the frequency was set at one DCE Login for every 15 customer sales distributed randomly. Each time a DCE Login was performed, a different principal name was used from the full population of the DCE Security registry. In a one-hour test of the 90 MHz Pentium DCE Security server, 25,920 DCE Logins were performed with a population of 10,000 principals, thus providing full coverage of the DCE Security registry.

The POS benchmark performs one CDS namespace lookup of a CDS object for every customer sale. Since POS does not perform CDS directory lookups, a script that randomly lists directories and subdirectories from the full CDS population was created to avoid artificially efficient caching by the CDS primary and secondary servers.

Cell Population

In a DCE cell of 10,000 users, the memory and disk sizes of DCE Security and CDS entities become significant. DCE Security accounts and principals require 1 KB each, or 10 MB for 10,000. For these tests, a 32 MB DCE Security server is sufficient—16 MB for OS/2 and DCE, and 10 MB for the registry.

DCE Cell Directory Services (CDS) has two types of entities: objects and directories. Objects—exported programs, aliases, names, and resource definitions—require 1 KB each, or 10 MB for 10,000. Directories require 14.2 KB each, or 142 MB for 10,000.

When CDS is distributed across several CDS replicas[1], the memory requirement for each replica would only need to be enough to contain the directories assigned to that CDS replica. For example, with one CDS primary server and nine CDS secondary servers with 10,000 subdirectories distributed evenly on the 10 replicas, only 14.2 MB would be required for the CDS namespace on each replica.

Since the cell root directory is replicated on all the servers, memory adequate for the root contents would also need to be considered. In these tests, 1,000 CDS objects were added to the root directory, which placed a 1 MB additional requirement on all 10 CDS replicas. For example, 16 MB for OS/2 and DCE, 1 MB for the root directory, and 142 MB for 10,000 directories

[1]A DCE cell has one primary CDS server and can also have one or many secondary servers. CDS primary and secondary servers are called *replicas*. Each replica contains the cell root directory, and may contain all or some of the directories that comprise the total CDS namespace.

equals 159 MB. If the namespace is distributed across 10 replicas, then the 142 MB can be divided by 10; therefore, 14.2 MB plus 17 MB equals a 31.2 MB memory requirement for each replica.

The `/.:/hosts` directory tree created by DCE configuration was moved to a dedicated CDS replica. Because of the available memory on our test hardware, some tests were run with a full population of 10,000, while others were run with a reduced population of 1,000. Sufficient testing was performed to quantify the performance difference between the two population levels and to provide a simple algorithm to bridge between them.

In the tests with a CDS primary and some number of CDS secondary servers, half the number of directories and subdirectories were created on each server, and each server's top directory was replicated on one other CDS server. By having the top directory replicated on another server, DCE automatically maintains its contents on the second replica.

Customer Sales per Minute	RPC Calls per Second	CDS Lookups per Second	DCE Logins per Second	Percent 16 Mbits per Second Network Utilized
8,000	933	139	8.9	80%
7,000	817	117	7.8	70%
6,000	700	100	6.7	60%
5,000	583	83	5.6	50%
4,000	467	67	4.4	40%
3,000	350	50	3.3	30%
2,000	233	33	2.2	20%
1,000	117	17	1.1	10%

Figure 1. Workload conversion table

For example, in the test with one CDS primary and eight secondaries (described later), each CDS replica contains one top directory, `/.:/TestDir0` to `/.:/TestDir8`. Each of these directories contains 500 subdirectories, `/.:/TestDir0/D1000` to `D1499`. Also, each CDS replica contained a read-only copy (replica) of the tree from one other server. Finally, each replica has a copy of the 1,000 objects in the root directory `/.:/TestObject000` to `TestObject999`.

The total CDS population was 1,000 objects and 9,000 directories. Since IBM DCE 1.2 for OS/2 and AIX was used for these tests, replication of the DCE Security server is not yet available. A single DCE Security server was used in all configurations tested. The first release of WS and DSS will support security replication; it is currently available on DCE/6000 for AIX Version 1.3. The first release of WS and DSS will require at least one OS/2 DCE Security replica to support the function required by LS4 legacy servers and clients.[2]

A variety of OS/2, AIX, and Windows DCE clients were used to drive the workload for these tests. A total of 56 physical clients were used, with each client running one or more client processes. There were 200 logical sessions. The combined horsepower of these clients was sufficient to drive about 8,500 POS customer sales per minute. This was enough to fully utilize the lower-horsepower DCE server configurations, but not some of the high-end RISC System/6000 configurations.

Measurement Methodology

The methodology for measuring the performance in the following tests was to measure the system in a busy state. In addition to the metrics shown in Figures 2, 5, 6, and 7, there are other performance factors to consider, which occur while the tests are running. Figure 1 is a conversion table to determine the other workload factors from the metric of interest in Figures 2, 5, 6, and 7.

DCE Security Performance

Three different OS/2-based DCE Security servers were measured: 486-33 MHz, 486-66 MHz, and a Pentium 90. No RISC System/6000 models were used because of insufficient client hardware to drive an interesting workload.

Figure 2 shows the DCE Security server performance, expressed as DCE Logins per second. The measurements were taken in a steady state, while the DCE Security server was nearly 100% utilized. The DCE Logins per second in Figure 2 are based on a registry population of 1,000 accounts and principals. When the population

[2]Initial tests have been conducted of DCE Security replication on the OSF 1.1 base under development. These tests indicate that the scale-up characteristics of DCE Security replication will be similar to the CDS replication characteristics presented in this article.

was increased to 10,000, the throughput decreased by 18%. This suggested a 2% reduction in maximum performance for each additional 1,000 accounts and principals above the base 1,000.

DCE Login and other functions contacting the DCE Security server have two options for obtaining the network address of the DCE Security server. (Figure 3 shows the performance of both options for a 386-25 MHz client.)

The first option (the default) is to look up the address of the DCE Security server from CDS. The DCE Security server location is stored in the CDS namespace as an object in the CDS root directory. Using CDS to locate the DCE Security server every time it is needed has two negative performance impacts:

- Figure 3 shows that the response time for the client is more than two times longer than when CDS is not used (see option 2).
- The load on the DCE CDS server is high. For example, using a RISC System/6000 Model 570 CDS server with a DCE Login rate of 8,000 DCE Logins per hour, the CDS server is 40% utilized with no other CDS activity.

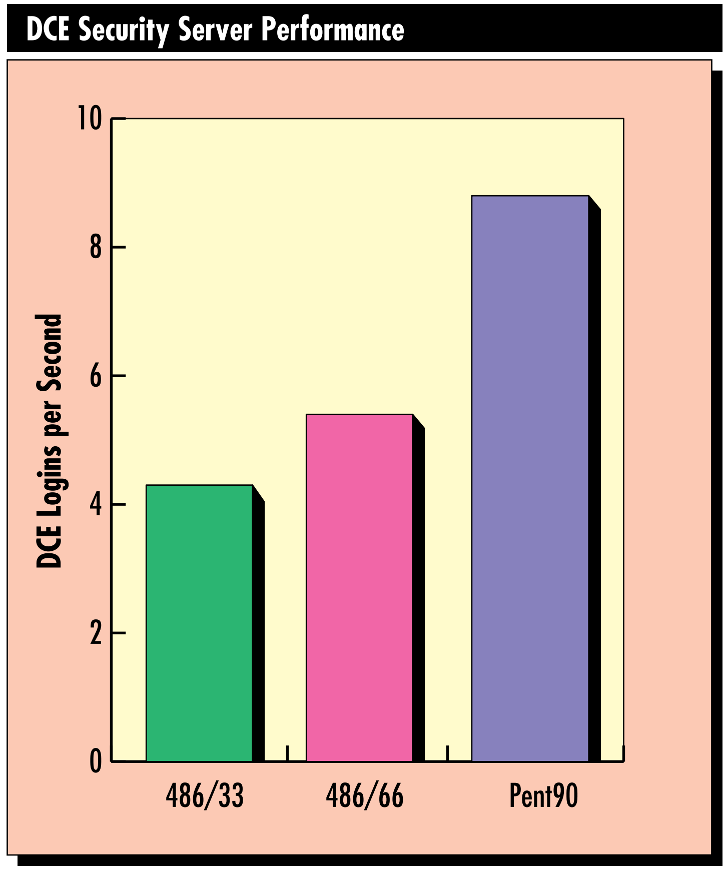


Figure 2. DCE Security Server performance

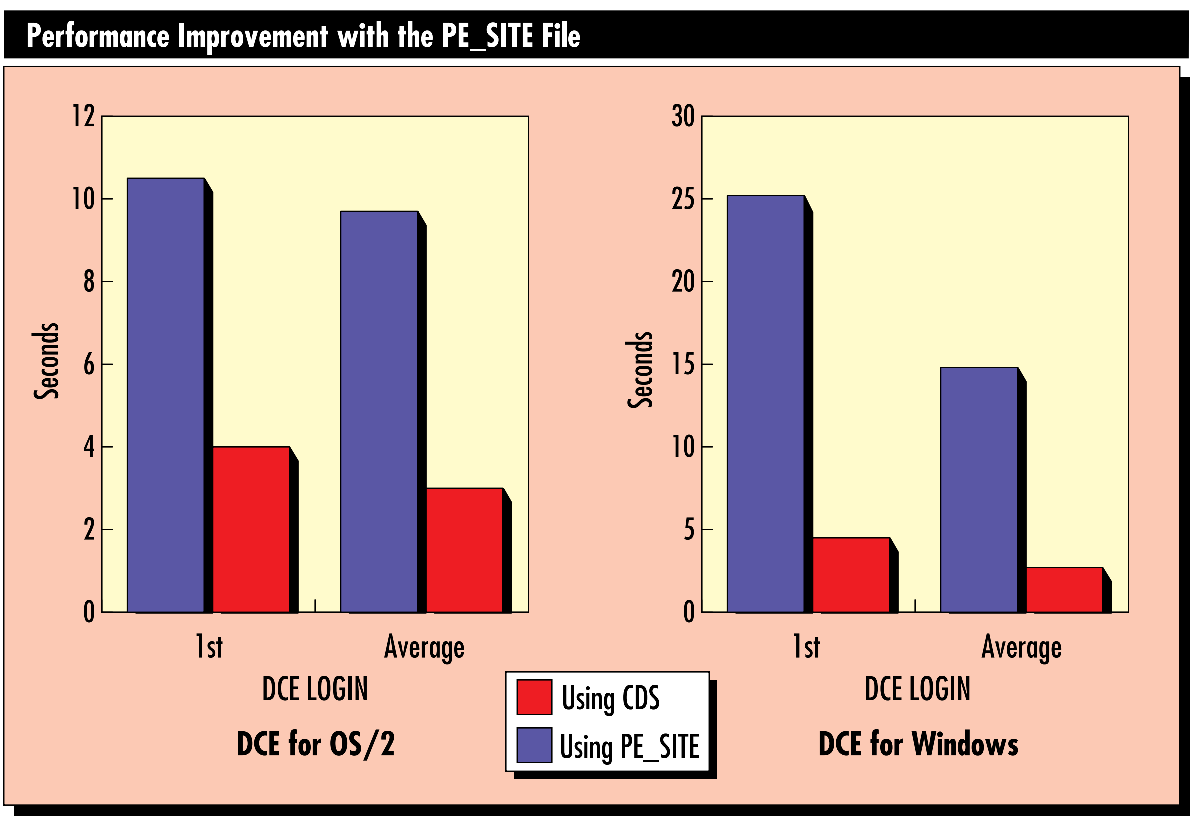


Figure 3. Using the PE_SITE file has improved performance

The second option is to allow the DCE client to store the DCE Security server's address in a file called PE_SITE on the client's hard disk. The PE_SITE option is activated in the client's CONFIG.SYS file with a SET BIND_PE_SITE=1 statement for IBM DCE for OS/2 1.2. In Warp Server using OSF DCE Version 1.1, the statement will be SET TRY_PE_SITE=1; the installation default will be 1.

The Warp Server domain controller acts as the DCE client on behalf of any legacy (LAN Server 4.0 and below) LAN requesters and additional servers.

In the next section, "DCE Cell Directory Service Performance," the results reflect option 2, using PE_SITE. If option 1 is used, 15% must be subtracted from the maximum CDS Lookups per second reported in Figures 5, 6, and 7.

The peak DCE Login activity reported by BigCo was 1,534 LAN logons per hour, or 0.426 logons per second. Since the three OS/2-based DCE Security servers tested can support 3.7 to 7.2 DCE Logins per second, DCE meets BigCo's DCE Login requirements.

The user response time of DCE Login depends on the processor speed of the DCE client. Figure 4 shows the response times for a range of DCE for OS/2 hardware platforms. These measurements were made without using the PE_SITE file. If PE_SITE is used, the improvement would be consistent with Figure 3.

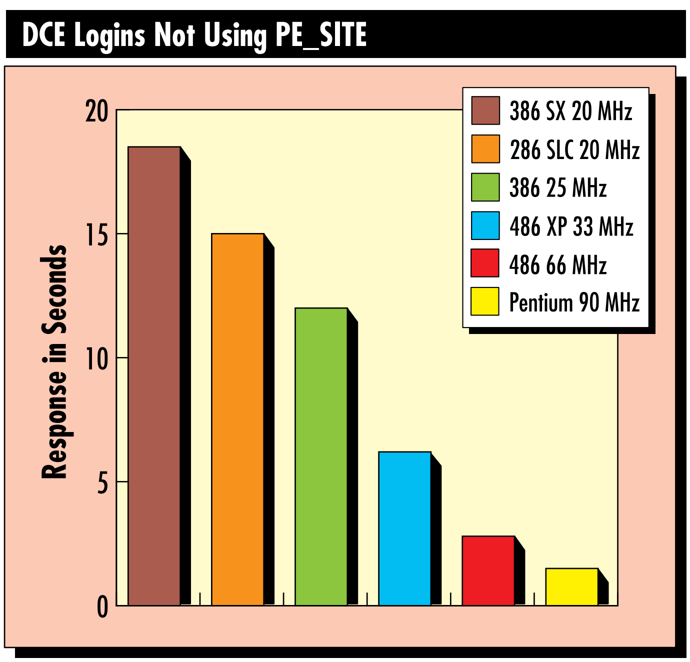


Figure 4. DCE logins that do not use PE_SITE

DCE Cell Directory Service Performance

The performance of the following DCE CDS configurations were tested:

- A stand-alone, single, primary CDS server
- A primary CDS server with three 486-66 MHz secondary servers
- A primary CDS server with one RISC System/6000 Model 990 secondary server
- A primary CDS server with six 486-33 MHz secondary servers
- A primary CDS server with eight 486-33 MHz secondary servers

The primary CDS servers studied were 486-66 MHz, Pentium 90 MHz, and RISC System/6000 Models 570, 580, and 990.

Each client has a local CDS cache, containing some part of the CDS namespace on the client's hard disk. DCE clients look up objects and directories in the CDS namespace in two ways.

The first way is to resolve the lookup in its local cache. If the client fails to find the object/directory in its local cache, the client would then call the CDS server to request a new copy of some portion of the CDS namespace.

In these tests, the client's local CDS cache satisfied all requests. This raises the question of why the CDS server is busy, since the client is satisfying the requests locally. Even though the client has the information in its cache, it must get permission from the CDS server to use the cached information.

The second way is to force the client to request a new cache from the CDS server. This occurs when either the local cache does not contain the requested information, or the cache becomes too old and a refresh is forced. The forced CDS refresh in these tests was accomplished by setting the client's cache expiration age to 0 (zero) seconds.

The main difference between the methods is that no actual data is retrieved from the server's namespace using the first method.

The CDS performance data is presented in both ways—with CDS Refresh and without CDS Refresh.

Performance With CDS Client Cache Refresh

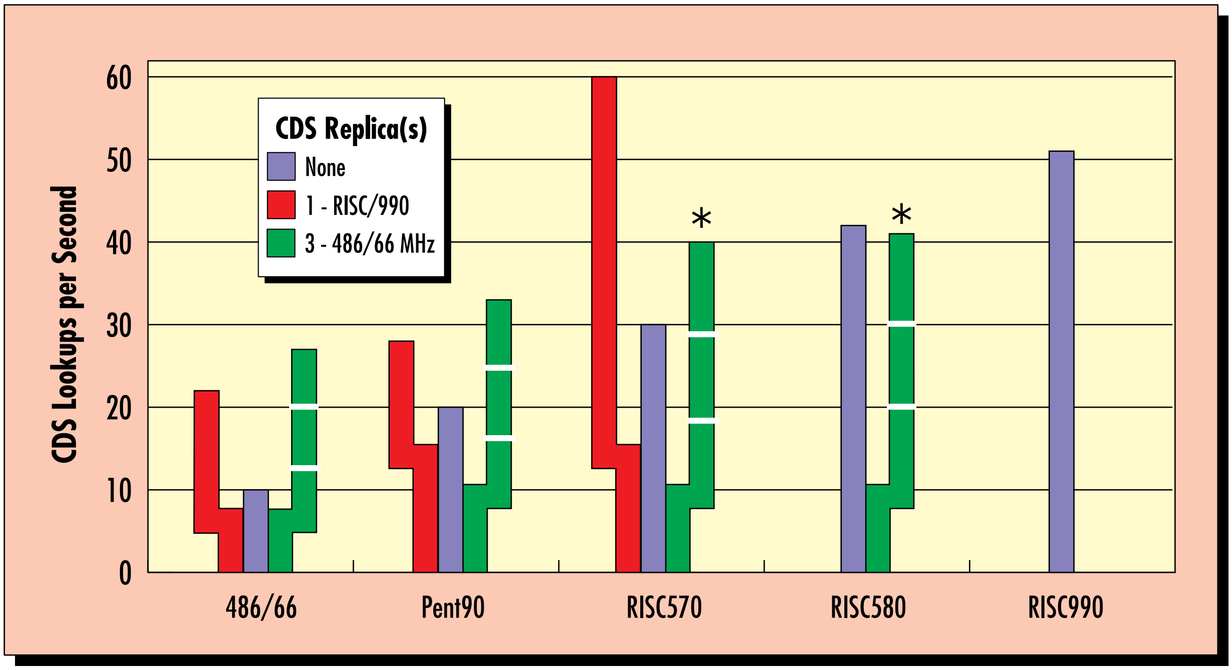


Figure 5. Performance with CDS client cache refresh

Observations on Performance Results

There are several observations of performance results:

- The throughput shown reflects nearly 100% CPU utilization of the primary CDS server.
- In each configuration, the Primary CDS server was the first to reach 100% CPU utilization, except as noted by an asterisk (*).
- If the DCE Security `PE_SITE` (described above) is not used, subtract an additional 15% from the CDS Lookups per second.

CDS Performance with CDS Client Cache Refresh

Figure 5 reflects a CDS population of 1,000 CDS objects and directories. For each additional 1,000 entries, subtract 2% from the CDS Lookups per second.

A stand-alone primary CDS server: Figure 5 shows the maximum CDS Lookups per second for the five CDS server machine types tested. (The total height of the center bar of the "cactus" is the throughput for this configuration.)

A primary CDS server with three 486-66 MHz secondary servers: The right branch of the "cactus" in Figure 5 shows the increase in performance when three 486-66 MHz secondary servers are added to the primary CDS server.

Three secondary CDS servers constituted the bottleneck in the tests for RISC System/6000 Models 570 and 580. This bottleneck can be relieved either by adding more replicas or by changing to faster replicas. The CDS Lookup requests are randomly distributed among all CDS primary and secondary servers. Therefore, if there is a great difference in horsepower, this type of bottleneck might occur. A test described later (in Figure 7) with eight CDS secondary servers demonstrates how adding more replicas can distribute the load across more low-horsepower servers.

A primary CDS server with one RISC System/6000 Model 990 secondary server: The left branch of the "cactus" in Figure 5 shows the increase in performance when one RISC System/6000 Model 990 secondary server was added to the primary CDS server. This configuration demonstrates that having a single high-horsepower CDS secondary server can be less effective than having more replicas to distribute the randomly assigned workload.

For the 486-66 MHz and the Pentium 90 primary CDS servers, the single RISC System/6000

Model 990 secondary server did not do as well as the three 486-66 MHz secondaries. This is due to the random distribution—half the requests caused the primary CDS server to reach 100%, while the Model 990 was underutilized.

In the RISC System/6000 Model 570 test, the single Model 990 looks much better, at nearly twice the stand-alone, as expected. Remember that the three 486-66 MHz secondaries were the bottleneck in these two configurations.

CDS Performance Without CDS Client Cache Refresh

For each additional 1,000 CDS entries, subtract only 1% from the CDS Lookups per second. There is less degradation for additional entries than in the "with CDS refresh" case.

Figure 6 shows the performance of the same configurations shown in Figure 5. The difference is that the clients are allowed to use their local CDS client cache.

The comments above about Figure 5 also apply to the configurations shown in Figure 6.

Since there is less interaction with the CDS servers when the CDS client cache is not refreshed, the performance for each configuration is higher in each case. In real life, there will be a mixture of lookups with and without refreshing the CDS client cache; therefore, the correct answer lies somewhere in-between.

A Primary CDS Server with Eight 486-33 MHz Secondary Servers

After completing the first two sets of tests, a question of quantity versus horsepower of the secondary CDS servers remained. The 486-33 MHz machine was chosen for this test to demonstrate that more might be better than bigger. Two tests were conducted using the Pentium 90 MHz for the primary CDS server.

The first test had six 486-33 MHz secondaries. The performance was much better than either of the earlier configurations since the work was spread across more CPUs. Figure 7 shows a greater benefit with cache refresh than without cache refresh.

The second test used eight 486-33 MHz secondary (replica) servers. The performance of this configuration is equivalent to the stand-alone RISC System/6000 Model 990, also shown in Figure 7.

Results from the six and eight CDS secondary server tests indicate that more is better. Adding

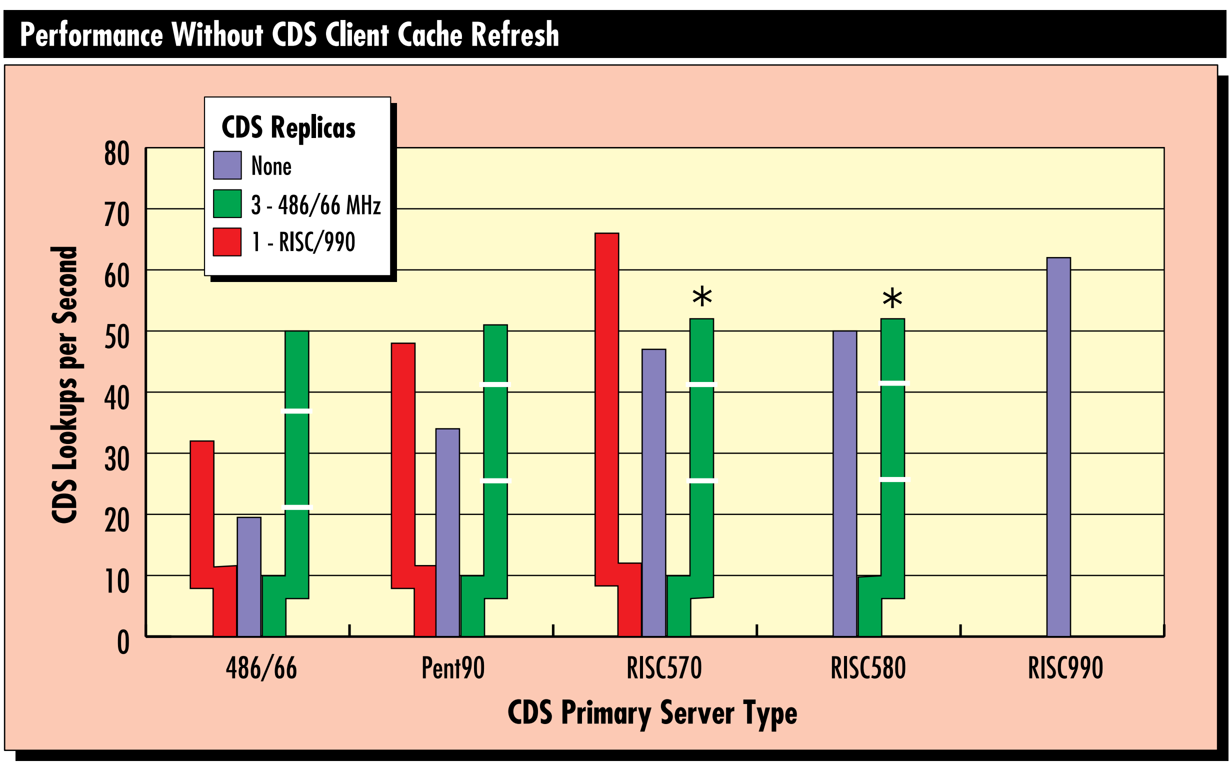


Figure 6. Performance without CDS client cache refresh

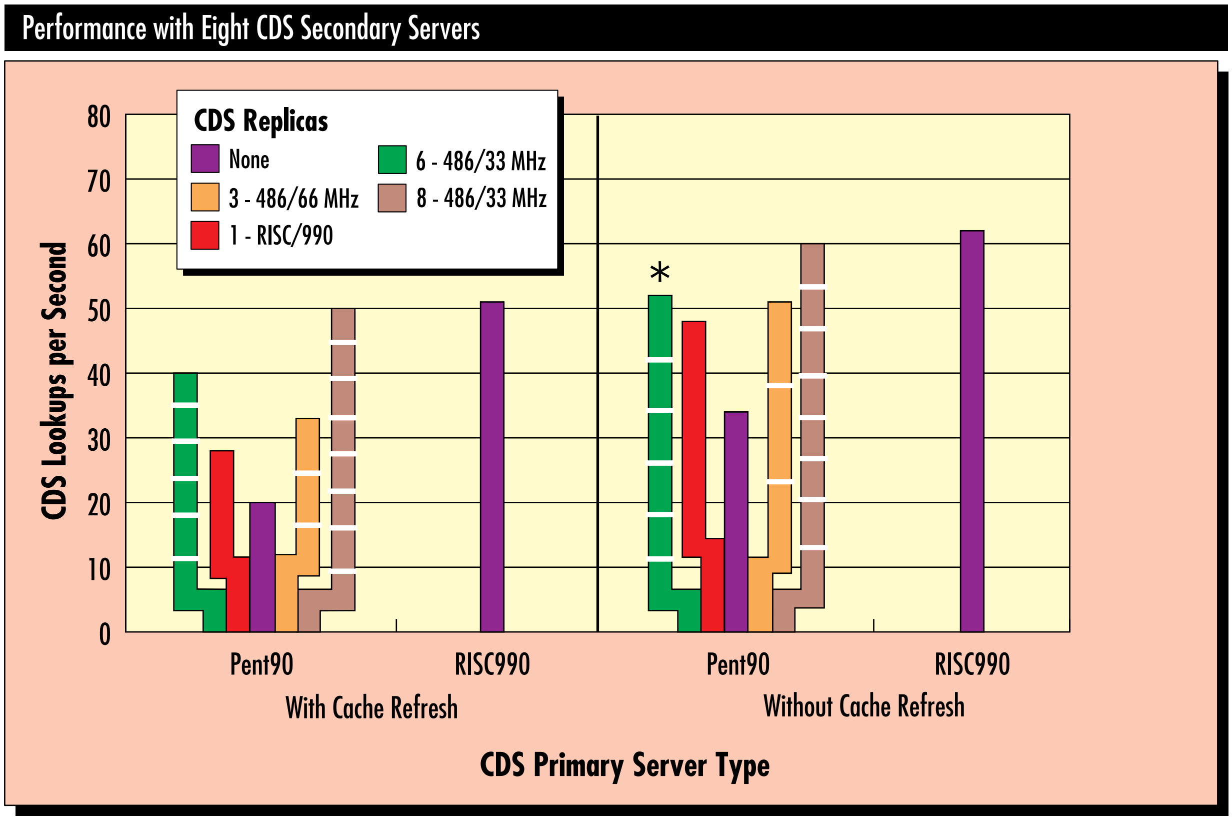


Figure 7. Performance with eight CDS secondary servers

more CDS replicas to absorb additional workload and CDS namespace population will increase the performance and capacity limits of the DCE cell.

Satisfying CDS Lookup Needs

Since BigCo did not provide specific requirements for object and directory populations and performance objectives, we can conclude from the populations of other LAN Server customers that the population is representative. Based on BigCo's network statistics, our POS benchmark application generated about the same network utilization, suggesting a similar workload. The POS benchmark performs an unnaturally high number of CDS lookups, higher than other customer applications. Therefore, the range of results achieved in these tests and the potential for improvement using additional replicas will satisfy the CDS lookup needs of BigCo.

The '/.:/hosts' Directory Tree

The DCE configuration program creates one CDS directory (`/.:/hosts/<hostname>`) and three or four CDS objects on behalf of each DCE client. In a large system such as BigCo, this would require an additional 150 MB to 180 MB of disk space for the CDS namespace. When the primary CDS server is configured, it creates the '`/.:/hosts`'directory, which is located by default on the primary CDS server. Because of the namespace disk requirement and potential degradation in CDS lookup performance, this directory should be moved to a dedicated CDS secondary server.

The `/.:/hosts` directory entries are accessed during DCE initialization by some of the management tools such as the DCE Graphical User Interface (GUI) configuration tools. Otherwise, there is little access to this directory tree. Since the performance of this replica is not critical, the memory size on this server does not need to be large enough to contain the full namespace database.

Immediately after the DCE test cell was created, we deleted the `/.:/hosts` tree from the primary CDS server and re-created it on a replica, then continued with the configuration of the other servers and clients. No other directories or objects were replicated on this secondary server.

For a 10,000-client DCE cell, we recommend 1 GB of disk space for the `/.:/hosts` directory and OS/2 `swapper.dat` file. The `swapper.dat` file can grow to three times the size of the local CDS namespace. During periodic maintenance by the CDS server daemon, the CDS namespace files can occupy up to 2.5 times the actual size of the namespace.

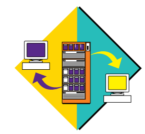
The existence of `/.:/hosts/<hostname>` and its objects is a dependency of all OSF DCE platforms. Furthermore, the literal CDS pathname is hard-coded in the OSF source and cannot be distributed across multiple replicas without causing a compatibility problem with other DCE platforms.

High Performance with Ample Capacity

The DCE services supporting IBM Warp Server can support workloads consistent with BigCo's security and directory needs.

The scalability of DCE Cell Directory Services through replication can support populations and arrival rates well above the levels measured in this study.

The IBM Warp Server product will exploit DCE technology. It is a large step forward in consolidating administration and interoperability in large enterprise environments.



Bob Russell, IBM Corporation, 11400 Burnet Road, Austin, TX 78758. Mr. Russell is an advisory performance analyst for IBM PSP Systems Performance. He joined IBM in 1963 as a customer engineer in the Electric Typewriter Division. His subsequent assignments have included branch office administration and various systems analysis positions. Currently, he is evaluating the performance of distributed client/server products on the OS/2 and AIX platforms.

POWER SPOT

IBM AIX Multi-Vendor Program

AIX is being offered by several Original Equipment Manufacturers (OEMs) on their own systems. The new IBM AIX Multi-Vendor Program (AIX MVP) defines processes to ensure that trademarked AIX systems, whether from IBM or other vendors, will support a consistent programming interface. Solution providers can then develop applications based on an interface that will be portable between those systems. This helps to protect customer application investments and to broaden the market for application developers.

With AIX MVP, software developers will have a broader range of binary compatible hardware platforms running AIX than ever before. Developers can take advantage of the growing number of these compatible platforms with a single port of their applications and can use any participating AIX MVP system as a reference platform for building and maintaining their applications.

The AIX MVP Program offers the following:

- Compliance tests that can be used by AIX source licensees and compatible hardware manufacturers to ensure that their systems adhere to a consistent AIX system definition
- Support structure for AIX application developers to help resolve trademarked AIX platform compatibility issues, regardless of whether an IBM or an OEM platform is involved
- Support for the common marketing activities of solution providers, IBM, and OEM AIX platform providers through a unified AIX Application Catalog on the World Wide Web, accessible directly from OEM providers' Web pages
- Participation of AIX licensees in the evolution of AIX through their participation in an Advisory Committee, which will guide the AIX MVP program. The committee will be chaired by IBM and comprised of members from representative companies in the AIX MVP Program

There is no charge to solution providers or customers to participate in the program.

Groupe Bull® and Motorola™ are among the current AIX licensees participating in the program. AIX MVP is planned to encompass AIX source licensees and compatible hardware manufacturers licensed to use the AIX trademark on their client and server offerings.

For More Information

Additional information is available through the World Wide Web on the Internet. To access, open the following URL: http://www.ibm.com.

The IBM Fax Information Service allows you to receive facsimiles of prior IBM product press releases. Dial 1-800-IBM-4FAX and enter "99" at the voice menu.