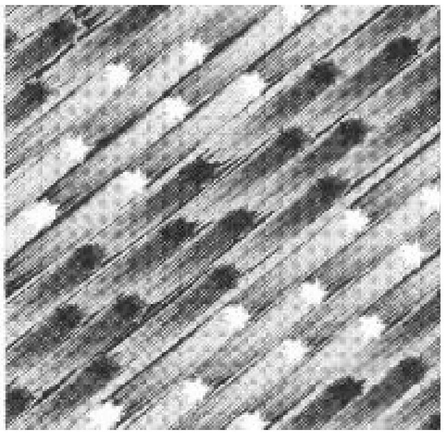


Implementing and Evaluating Performance in an ATM Backbone/TR Edge Device Environment

April 14, 1999

Joao Molina
NHD Installation Support
Cary, North Carolina

Table of Contents

Preface

The intent of this report is to share the results of the tests performed on a test network, assembled at the NHD Installation Support Lab, Weston Building, Cary, NC. The objective of the test was to verify functions and concepts and to test new fixes provided for any encountered problems. Diagrams of the physical and the logical network configuration, as well as some configuration data used in each individual piece of equipment are provided.

The Ganymede Chariot performance tool was utilized to measure the network's performance as far as data throughput is concerned, and the data collected is presented in the form of graphs.

It is preferable to have this document printed on a color printer, since all the diagrams are color coded for easier understanding.

For questions related to the performed tests, please send an e-mail to:

Joao Molina
molina@us.ibm.com
(919) 301-4446 / (8) 352-4446

Jeff Ottman
jottman@us.ibm.com
(919) 301-4448 / (8) 352-4448

Shannon Kurtz
rskurtz@us.ibm.com
(919) 993-4345 / (8) 589-4345

Overview

The object of this project was to ensure a smooth and successful customer installation by testing functions, products, fixes, and concepts, before they were actually installed at the customer production network.

The network design was a given variable in this project, and will not be discussed in this document. The only point I would like to make regarding the design is its focus on the redundancy and backup of resources.

With that objective in mind , a test network had been put in place, simulating three buildings of a campus network, each one with two 8260 Nways Multiprotocol Switching Hubs, two 8270 Nways Token-Ring Switches, and five workstations running a mixture of Windows/NT Workstation 4.0 and OS/2 Warp 4.0 (see figure 01).

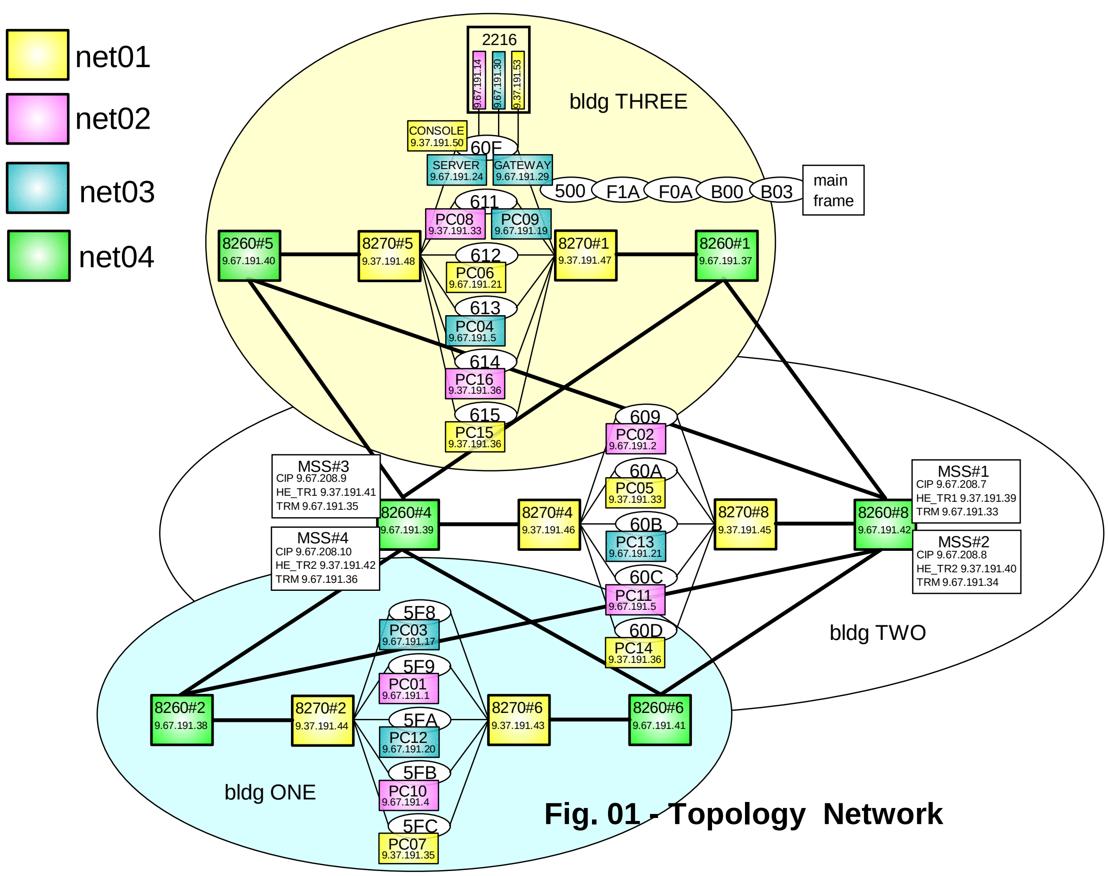


Fig. 01 - Topology Network

In each building, five workstations are connected to five different token-ring LAN segments, or domains. Each ring then connects to two 8270 Nways Token-Ring Switches, establishing a dual path configuration for the workstation(s) connected to that

LAN segment. Spanning tree is configured in the 8270s, in order to avoid loops in the network, and provide backup paths.

The root bridge is *8270 #4*, making the traffic between buildings to flow on the left hand side of the network using the token-ring switches *8270#5*, *8270#4*, and *8270#2*. The traffic will start flowing on the right hand side of the network, using the token-ring switches *8270#1*, *8270#8*, and *8270#6*, whenever the original path is not available.

Displaying the status of the token-ring LAN segments, we could see the spanning tree mechanism working. Figure 02 shows the SRB (Source Routing Bridge) spanning tree status for the token-ring switch *8270#1*.

Token-ring switch *8270#1* access to the LAN segments 611, 612, 613 and 614 is blocked to unicast frames, as we can see under the column “Segment Status”.

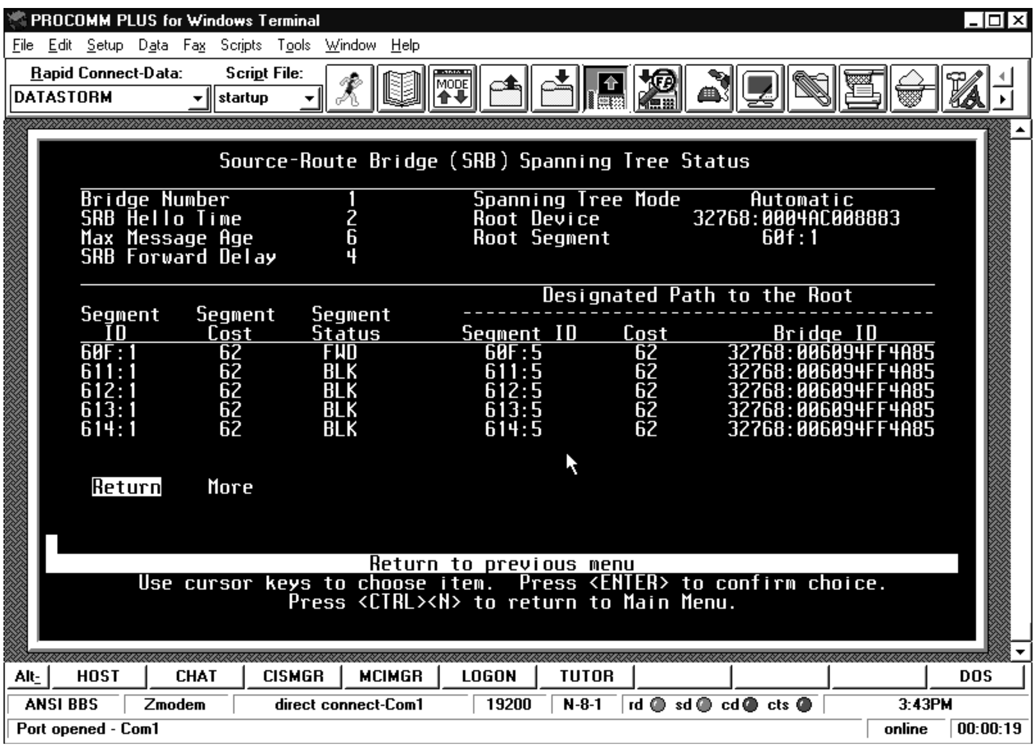


Fig. 02 - Bridge #1 Spanning Tree Status

Figure 3 shows that the token-ring switch *8270#5* LAN access to the same segments is allowed.

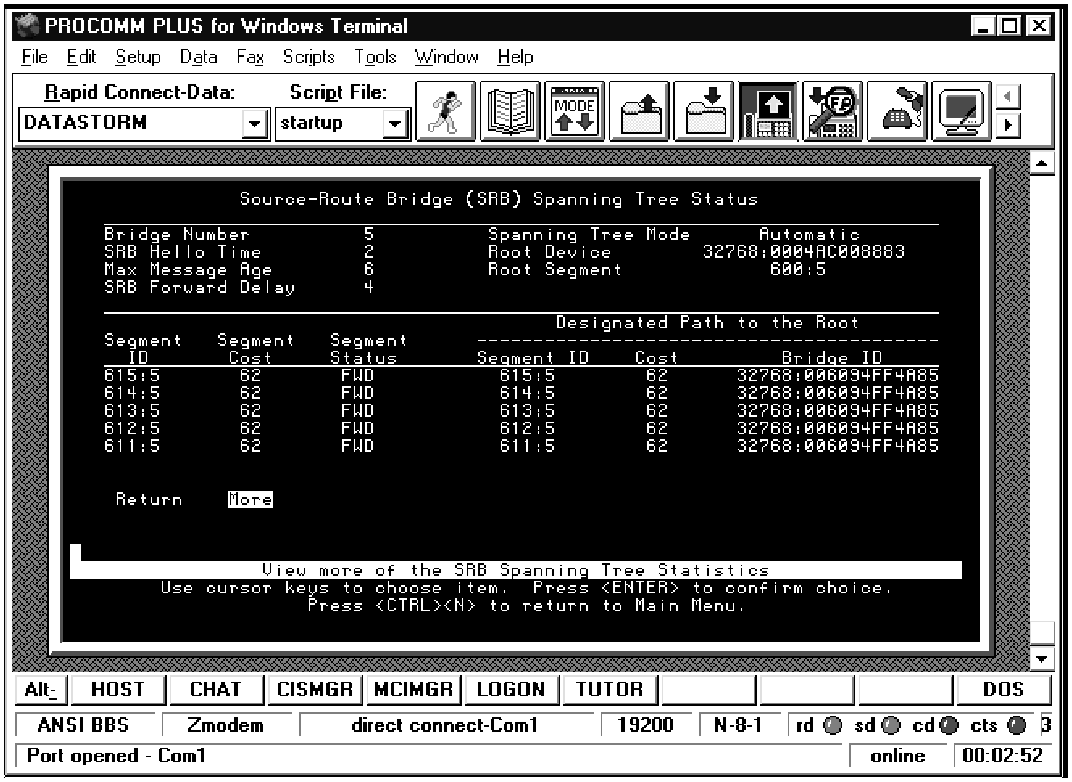


Fig. 03 - Bridge #5 Spanning Tree Status

The three simulated buildings are interconnected via PNNI connections. Each 8260 has two links to two different 8260s, providing redundancy and building the backbone network.

Three IP subnets, NET01, NET02, and NET03 have been created, with the 2216 doing the routing between the subnets over three token-ring interfaces, which are connected to the same LAN segment (60F). ATM is not configured on the 2216.

Four 8210 Nways Multiprotocol Switched Services (MSS) blades provide LECS and LES/BUS server functions for the network, creating the logical network, and providing redundancy for those servers, as well. Two 8210 MSS blades are plugged to *Hub #4* and two are plugged to *Hub #8*.

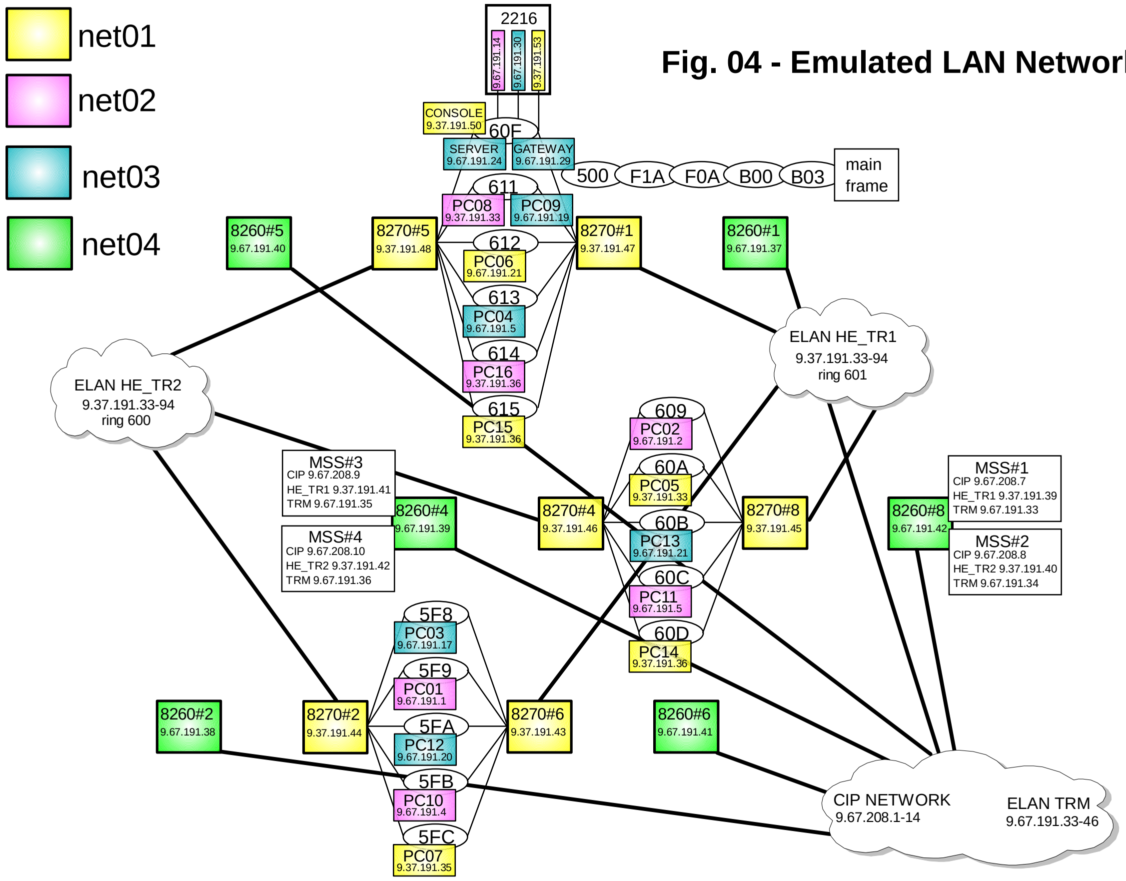


Fig. 04 - Emulated LAN Network

The three LES/BUS servers build three ELANs, which are HE_TR1, HE_TR2, and TRM. The TRM ELAN is used for network management purposes only, as opposed to managing the resources via a CIP (Classic IP) network. No routing/bridging between ELANs is enabled on the MSSs, so all inter ELAN traffic must be routed through the 2216.

The 8270s are ATM 155 Mbps connected to the 8260s, which form the ATM network backbone, running PNNI. The *8270 #5*, *8270 #4*, and *8270 #2* join ELAN HE_TR2 and build the primary path for workstations to reach resources on a different building or subnet. The *8270 #1*, *8270 #8*, and *8270 #6* join ELAN HE_TR1 and are the alternate path for workstations to reach resources on a different building or subnet when the primary path is down.

One of the applications running on the workstations is the “3270 Personal Communications”, which allows access to an SNA host, via a gateway. The gateway function is performed by the “IBM Communications Server”, running on a Windows/NT platform. That server has two Token-Ring adapters which connects to LAN segment 60F on the test network, and to ring 500 on the IBM network respectively.

A NETBIOS server, running the “IBM LAN Server 4.0”, is connected to LAN segment 60F as well. It provides file server functions to the 16 users/PCs. NETBIOS is the only protocol running on that server.

Products Release and Version

The code levels utilized in the test are:

8260	CPSW	operational	3.1.9
		boot	3.1.9
		FPGA	B52
		MSS FPGA	B50
8270	base		4.0.2
	UFC		1.14.1
2216	Common Code		3.0
8210	code		1.1 w PTF 6

TCP/IP Subnets

The following are the subnet address ranges utilized in the network:

subnet NET01
9.37.191.33 through 9.37.191.95, mask 255.255.255.192

subnet NET02
9.67.191.1 through 9.67.191.14, mask 255.255.255.240

subnet NET03
9.67.191.17 through 9.67.191.30, mask 255.255.255.240

subnet CIP
9.67.208.1 through 9.67.208.14, mask 255.255.255.240

subnet TRM
9.67.191.33 through 9.67.191.46, mask 255.255.255.240

Note that even though there is provision for IP addresses for the CIP network, they are not used.

The ATM network

The following table shows the functions performed by each 8210 MSS. Notice that every server has a backup running on a different 8260/8210.

8210 MSS	function
MSS#1 ARP server/client 9.67.208.7 HE_TR1 client 9.37.191.39 TRM client 9.67.191.33	- primary LES/BUS server HE_TR1 - ARP server
MSS#2 ARP client 9.67.208.8 HE_TR2 client 9.37.191.40 HE_ETH1 client 9.67.208.23 TRM client 9.67.191.34 backup default gw TRM 9.67.191.46	- the primary LES/BUS server HE_ETH1 - backup LECS configuration server - backup LES/BUS server HE_TR2 - primary LES/BUS server TRM - backup default gateway TRM
MSS#3 ARP client 9.67.208.9 HE_TR1 client 9.37.191.41 HE_ETH1 client 9.67.208.24 TRM client 9.67.191.35 pri default gw TRM 9.67.191.46	- the primary LECS configuration server - primary LES/BUS server TRM - backup LES/BUS server HE_TR1 - backup LES/BUS server HE_ETH1 - primary default gateway TRM
MSS#4 ARP client 9.67.208.10 HE_TR2 client 9.37.191.42 TRM client 9.67.191.36	- primary LES/BUS server HE_TR2

The following are the ATM addresses assigned to each addressable ATM resource in the network:

ARP server	39.53.45.41.52.53.5F.48.4F.46.46.01.08.48.21.00.01.12.0A.06
HE_TR1 LES/BUS	39.53.45.41.52.53.5F.48.4F.46.46.01.08.48.21.00.01.12.04.02
HE_TR2 LES/BUS	39.53.45.41.52.53.5F.48.4F.46.46.01.04.48.21.0D.01.14.05.10
TRM LES/BUS	39.53.45.41.52.53.5F.48.4F.46.46.01.04.48.21.0D.01.12.04.01
HE_TR1 backup	39.53.45.41.52.53.5F.48.4F.46.46.01.08.48.21.0D.01.12.07.03
HE_TR2 backup	39.53.45.41.52.53.5F.48.4F.46.46.01.04.48.21.00.01.14.08.11
TRM backup	39.53.45.41.52.53.5F.48.4F.46.46.01.08.48.21.00.01.14.14.01
Primary LECS	39.53.45.41.52.53.5F.48.4F.46.46.01.04.48.21.0D.01.12.00.00
Backup LECS	39.53.45.41.52.53.5F.48.4F.46.46.01.08.48.21.00.01.14.00.00

Performance Information

The Chariot software, a Ganymede product, was used to generate data traffic and measure data throughput.

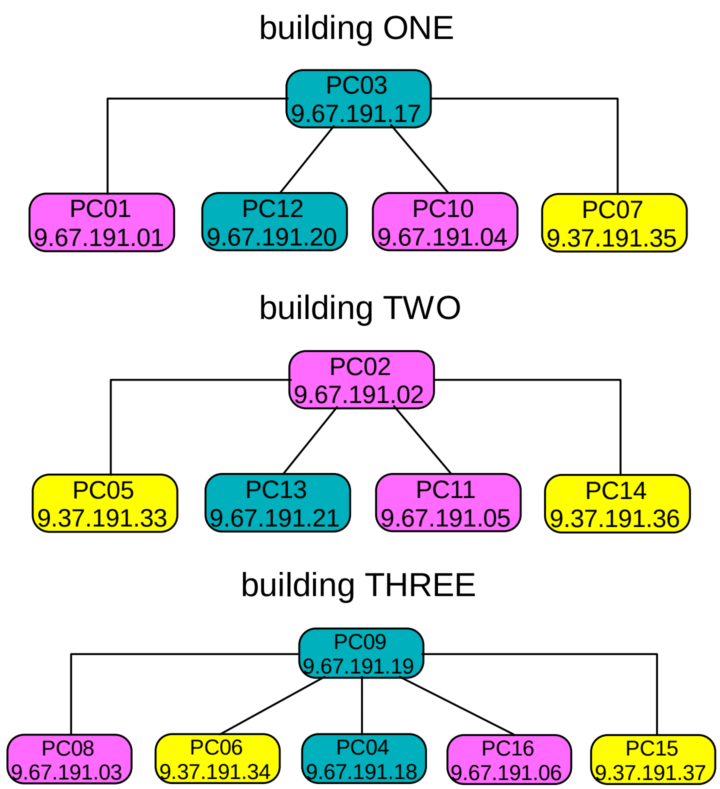
The tool allows measuring performance between two workstations, called end points. For each pair of end points, a script is selected and it emulates a specific application running on those two workstations. Multiple scripts can be executed simultaneously in the same workstation. Data collected is sent to a console workstation.

The selected script emulates a file transfer performed between the two endpoints. The file size is 100,000 bytes. The receiving workstation sends a request message, 100 bytes long, to the sender workstation which initiates the file transfer transaction. The sender performs the file transfer operation 100 times. At the end of each operation a record containing information about the operation is created and is sent to the console workstation.

The data collected was then stored in a binary spreadsheet file, .WK3, and was used to create the performance graphs shown in the next pages.

Two different scenarios were used to measure performance. Scenario A (figure 05) simulates data traffic inside the same building. One workstation in each one of the three buildings performs the role of a file server, while the other workstations are clients receiving the file been transmitted by the server. Notice that even though the workstations are in the same building, they may be on a different IP subnet than the server. Consequently, the data traffic has to be routed by the 2216, adding some delay to the total response time.

Fig. 05 - Chariot Scenario A



Scenario B (figure 06) simulates data traffic between buildings. One workstation, PC09, located in building THREE performs the role of a file server. All other workstations located in the same building and in the other buildings, are clients receiving files transmitted by the 'corporate server'. As in scenario A, workstation clients may be on a different subnet than the file server, requiring the router services.

Fig. 06 - Chariot Scenario B

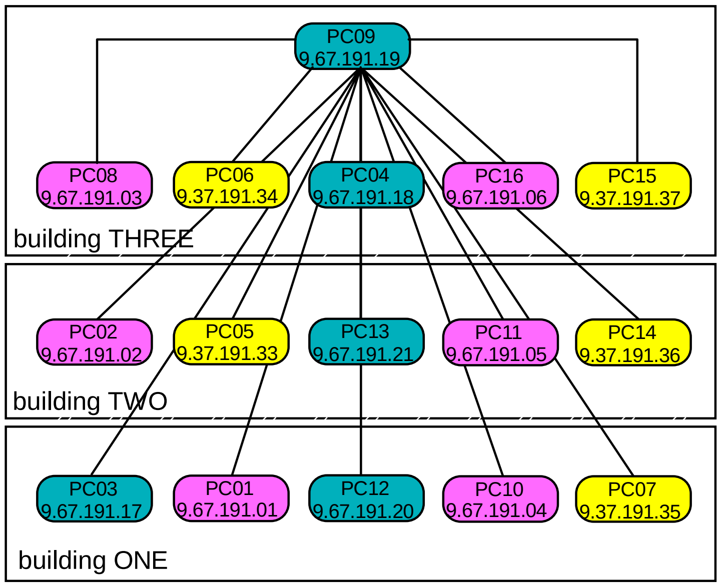
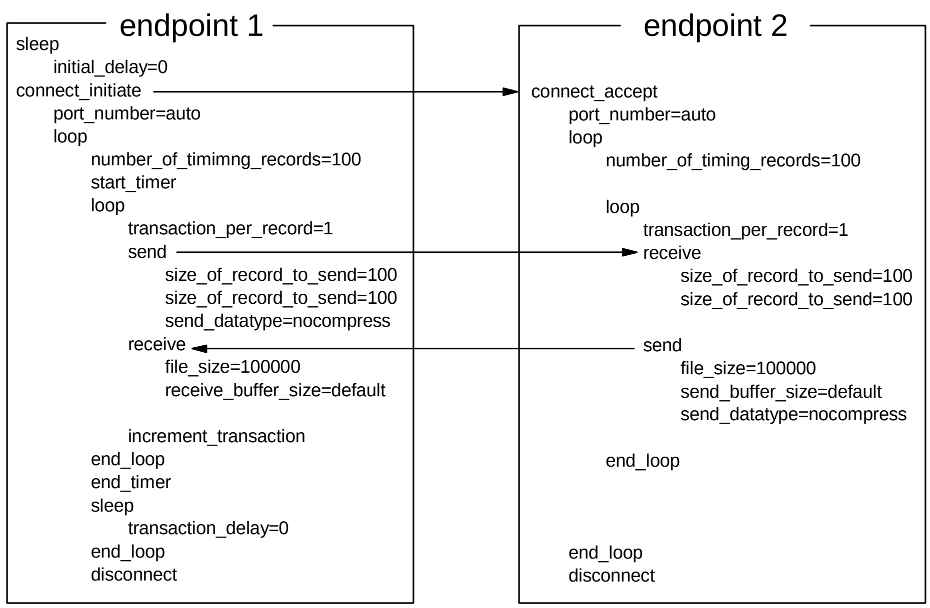


Figure 07 shows the scripts utilized to simulate the file transfer transaction. The scripts are coded and reside at the console workstation. They are shipped to the endpoint workstations when the simulation is started.

Fig. 07 - File Transfer Script



Performance Test Results

The main goals in running traffic through the network were to verify the network's behaviour under load, and to verify how fast the network could recover from outages such as loosing power on the token-ring switche(s), 8260(s), MSS(s), or having a piece of fiber accidentally disconnected. It wasn't our goal to stress the limits of the network.

The data collected by the Chariot software was plotted for each building. Figures 08, 09, 10 and 11 show data throughput for PC01, PC12, PC10, and PC07 running only one pair of PCs at a time, as described in scenario A.

Average throughput is about 850 Kbytes/sec for workstations on different subnets and 1,740 Kbytes for workstations on the same subnet, where data frames are switched, as opposed to routed. Those numbers are consistent among all the three buildings, and it shows that usually switched networks are with no doubts faster than routed networks.

Figures 08a, 09a, 10a and 11a show data throughput for PC01, PC12, PC10, and PC07 all running at the same time, as described in scenario B. In this case we verified that routed data throughput is about 350 Kbytes/sec, and the switched data throughput is about 900 Kbytes/sec. Those numbers show the influence of the network load on the data throughput.

The same approach was used for the other buildings when running the performance tests.

Fig. 08 - Data Throughput - Building ONE

(one workstation)

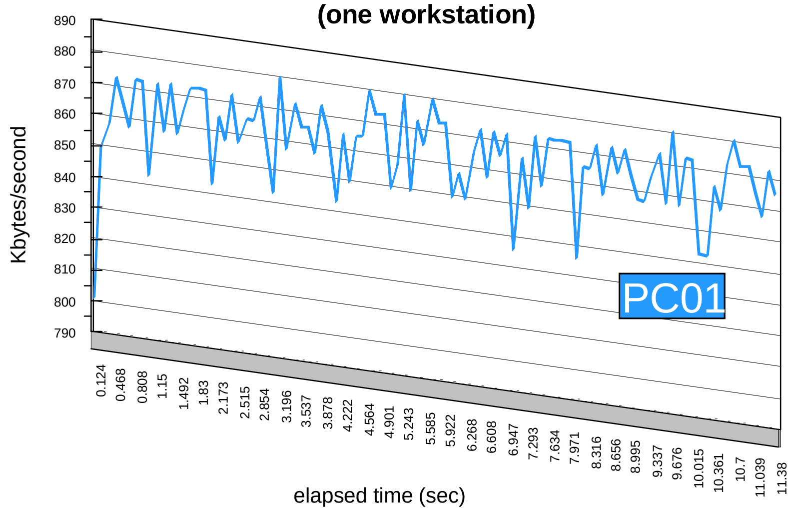


Fig. 08a - Data Throughput - Building ONE

(all the workstations)

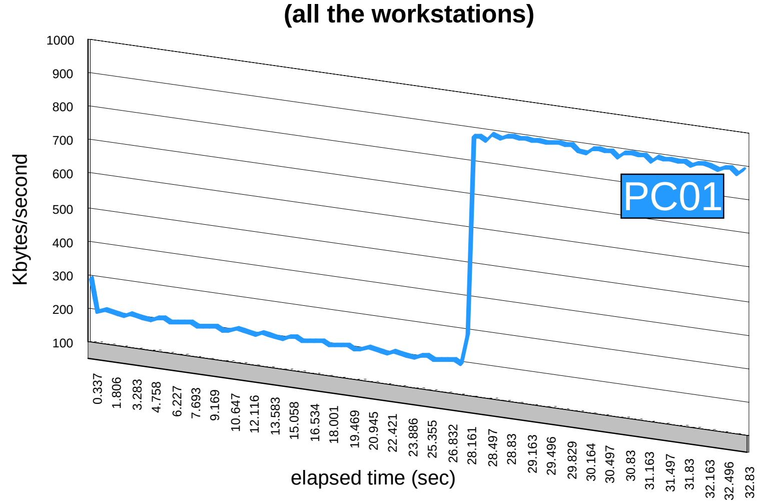


Fig. 09 - Data Throughput - Building ONE
(one workstation)

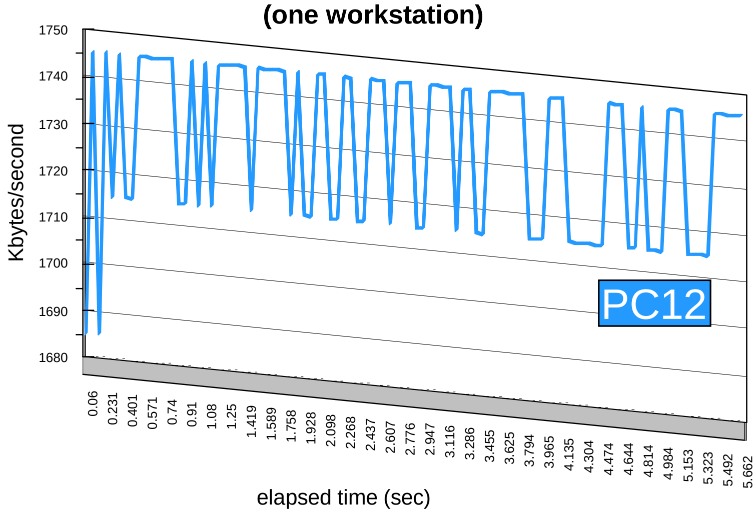


Fig. 09a - Data Throughput - Building ONE
(all the workstations)

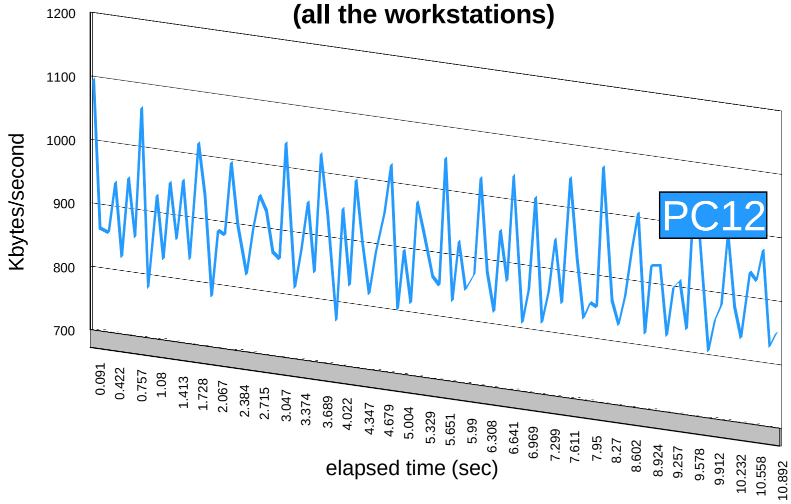


Fig. 10 - Data Throughput - Building ONE
(one workstation)

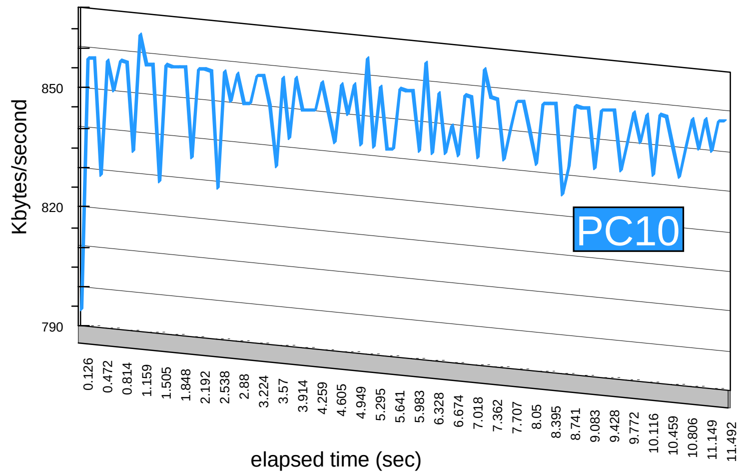


Fig. 10a - Data Throughput - Building ONE
(all the workstations)

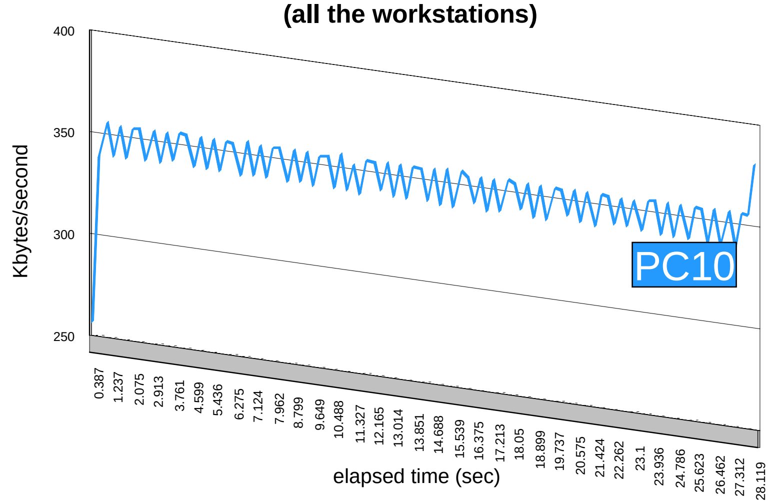


Fig. 11 - Data Throughput - Building ONE
(one workstation)

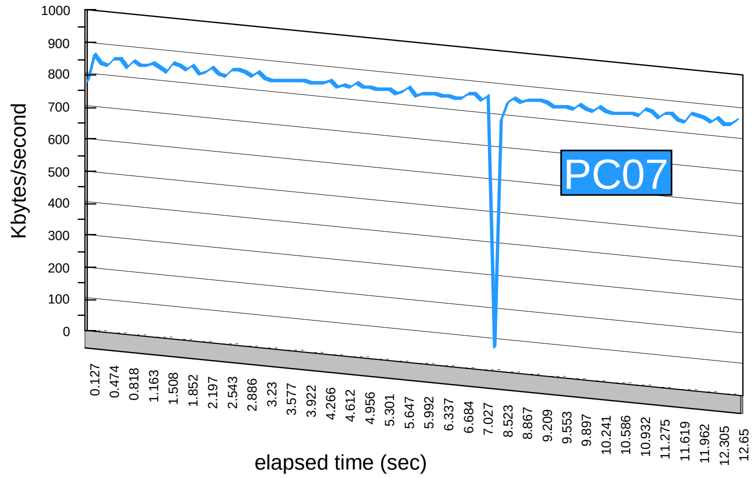


Fig. 11a - Data Throughput - Building ONE
(all the workstations)

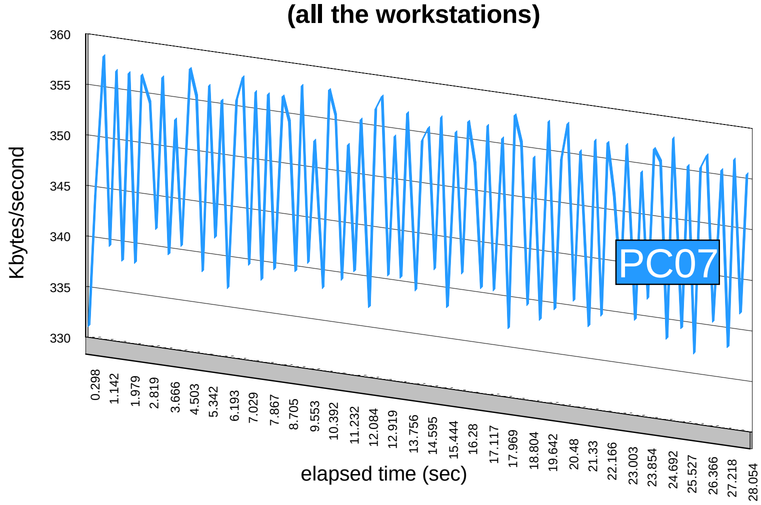


Fig. 12 - Data Throughput - Building TWO
(one workstation)

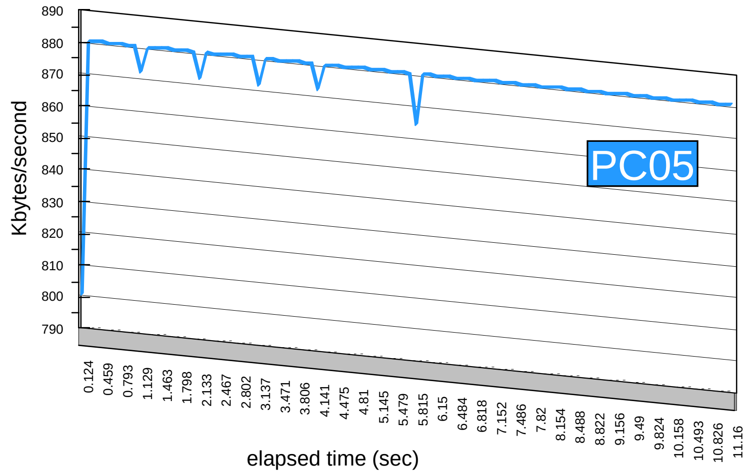


Fig. 12a - Data Throughput - Building TWO
(all the workstations)

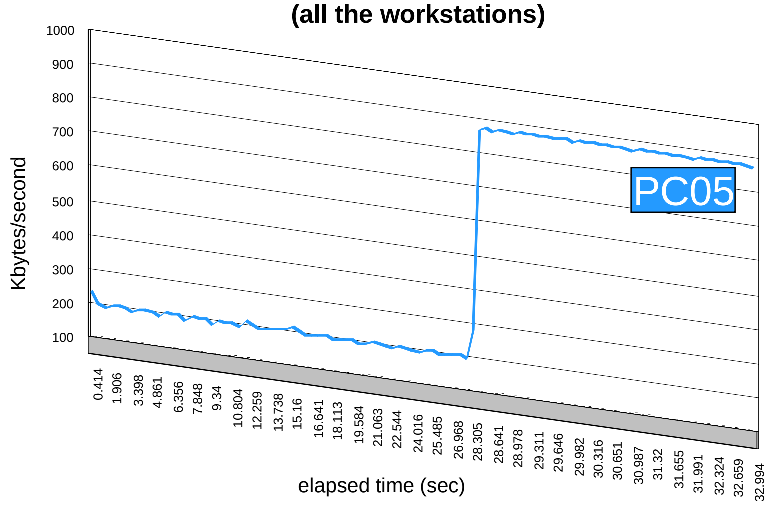


Fig. 13 - Data Throughput - Building TWO
(one workstation)

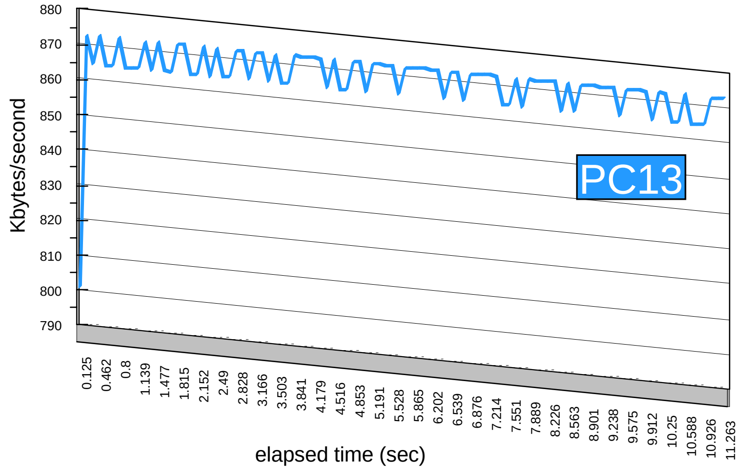


Fig. 13a - Data Throughput - Building TWO
(all the workstations)

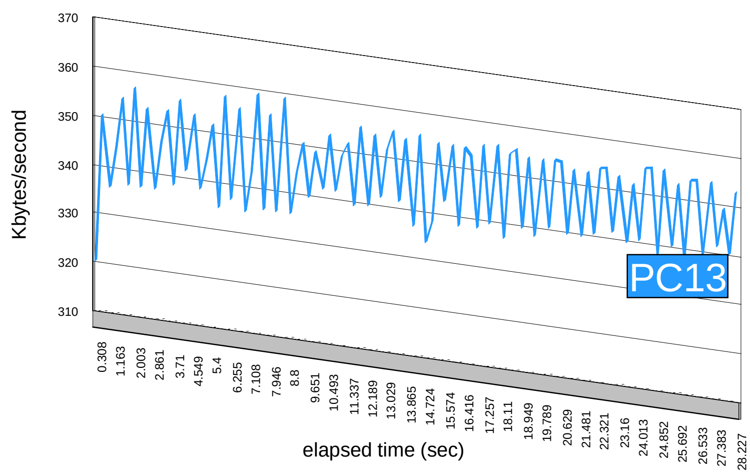


Fig. 14 - Data Throughput - Building TWO

(one workstation)

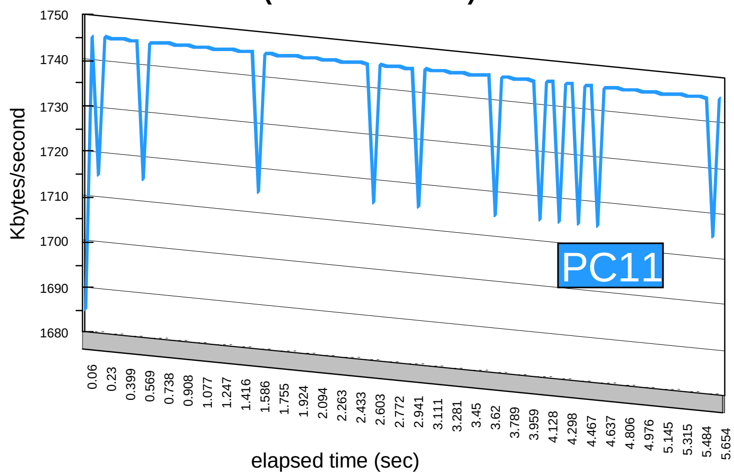


Fig. 14a - Data Throughput - Building TWO

(all the workstations)

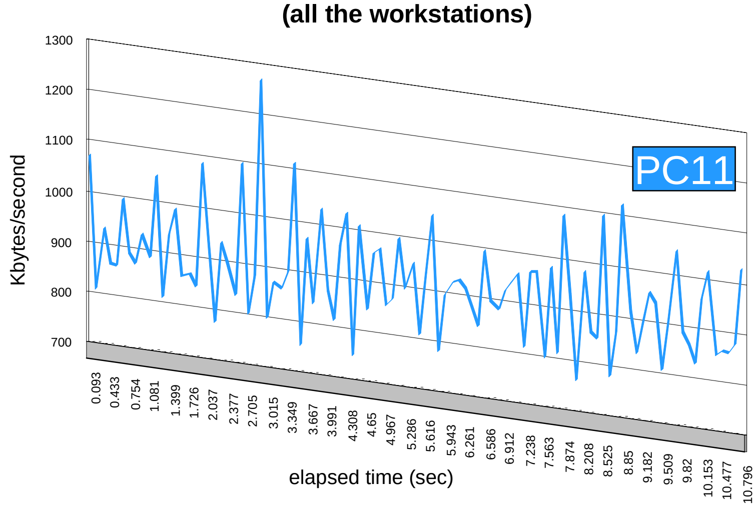


Fig. 15 - Data Throughput - Building TWO
(one workstation)

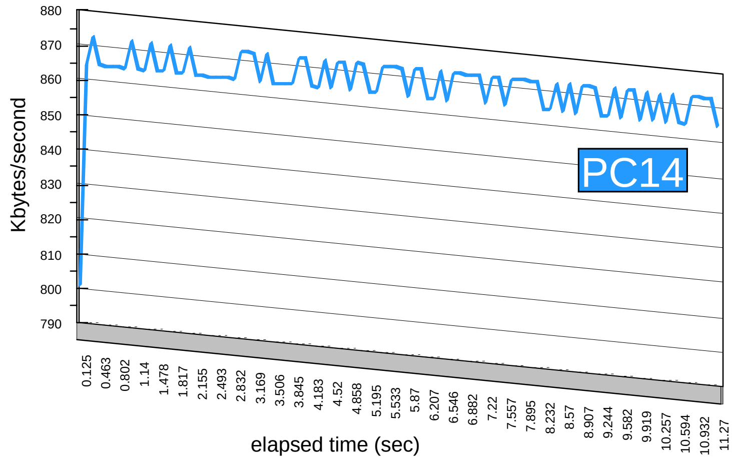


Fig. 15a - Data Throughput - Building TWO
(all the workstations)

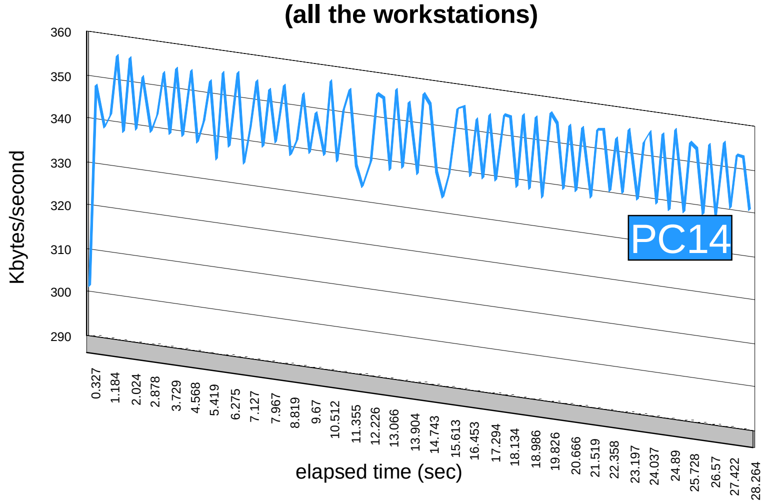


Fig. 16 - Data Throughput - Building THREE (one workstation)

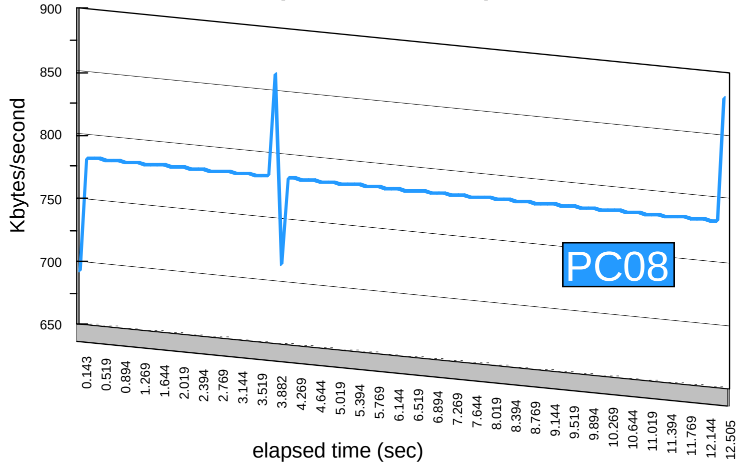


Fig. 16a - Data Throughput - Building THREE (all the workstations)

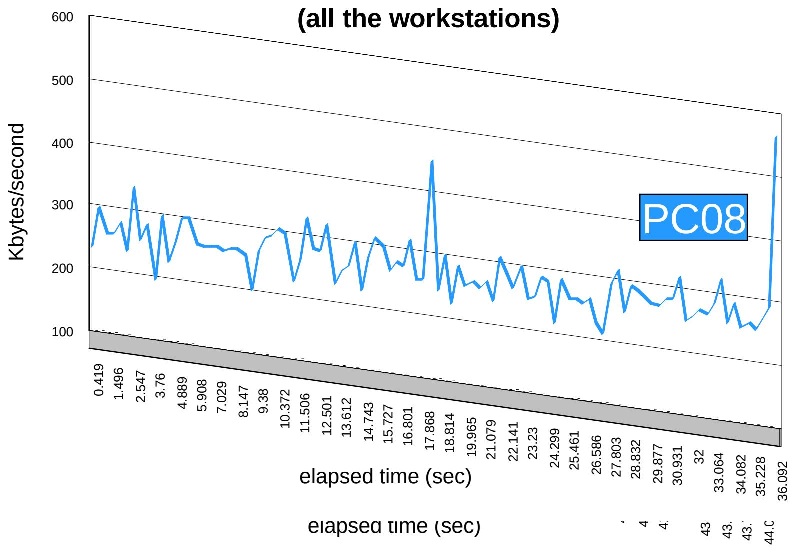


Fig. 17 - Data Throughput - Building THREE
(one workstation)

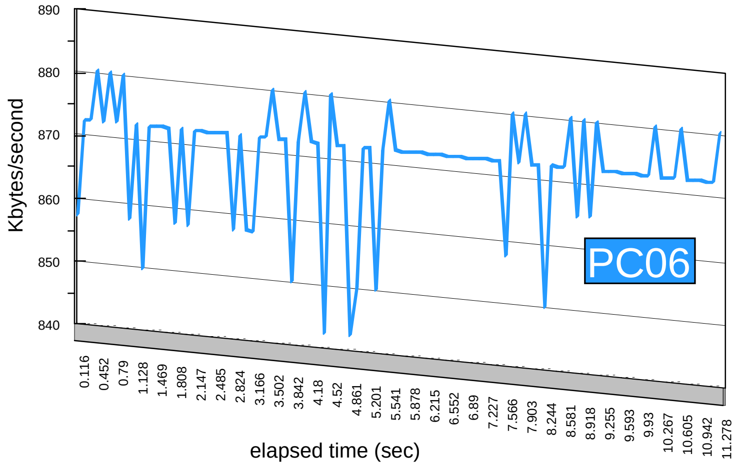


Fig. 17a - Data Throughput - Building THREE
(all the workstations)

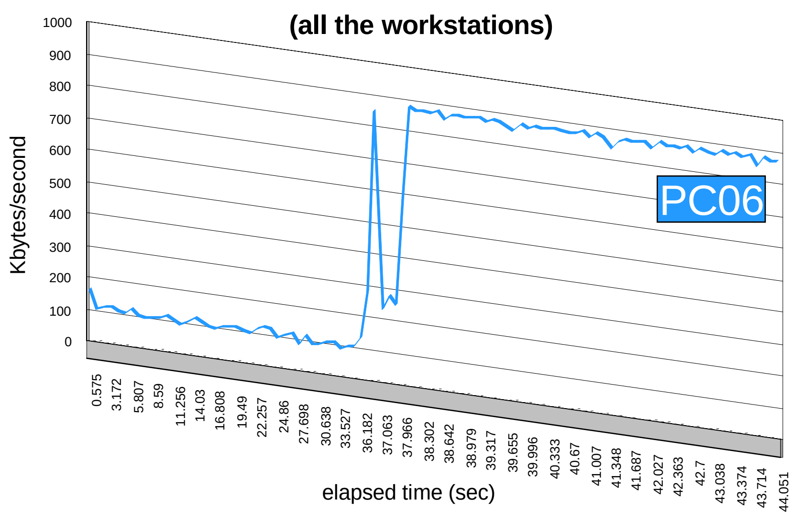


Fig. 18 - Data Throughput - Building THREE

(one workstation)

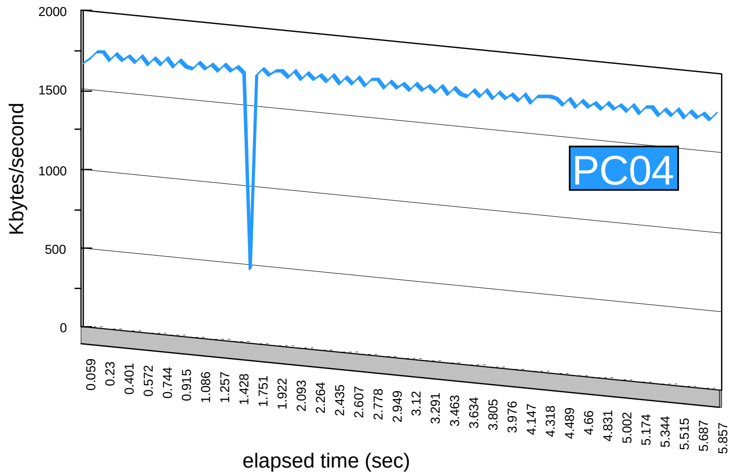


Fig. 18a - Data Throughput - Building THREE

(all the workstations)

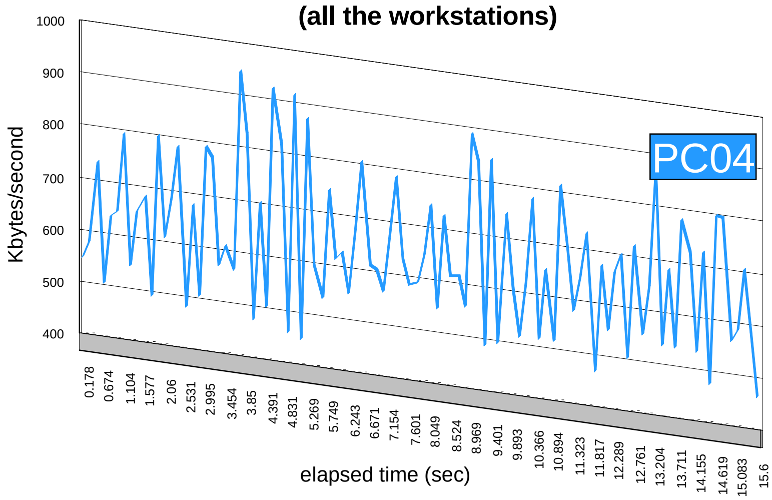


Fig. 19 - Data Throughput - Building THREE
(one workstation)

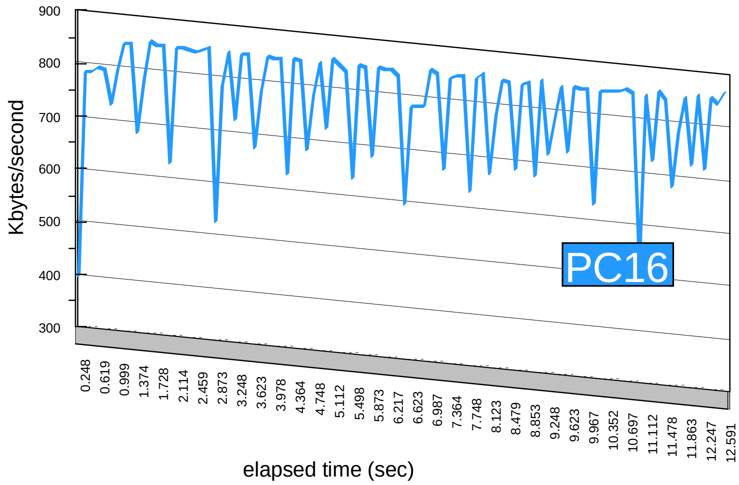


Fig. 19a - Data Throughput - Building THREE
(all the workstations)

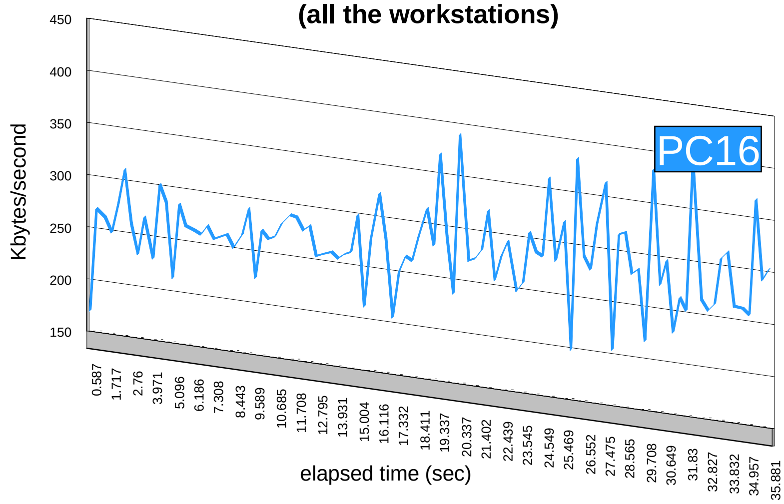


Fig. 20 - Data Throughput - Building THREE

(one workstation)

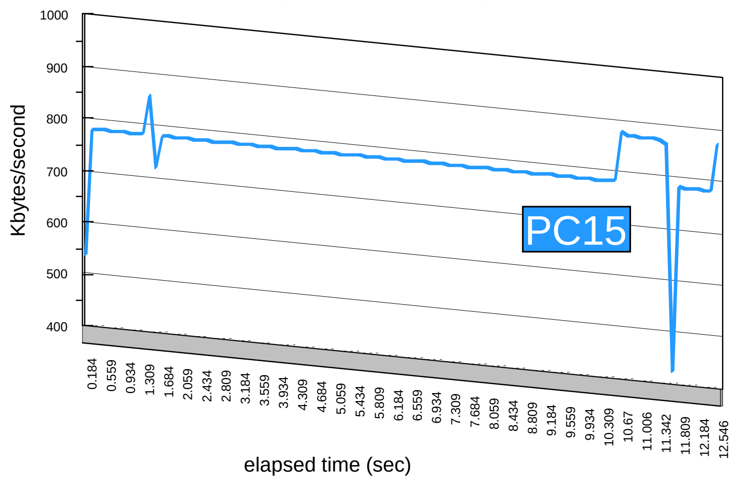
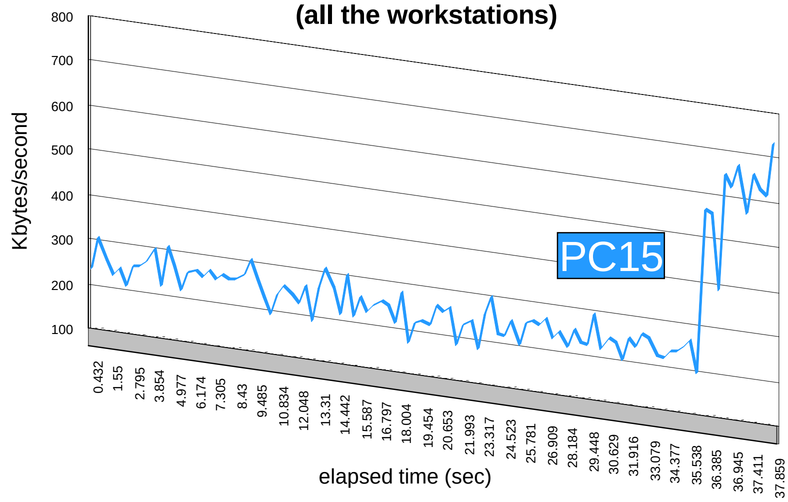


Fig. 20a - Data Throughput - Building THREE

(all the workstations)



Problem Summary

During the failure tests performed we found a number of problems with the different components of the network. Here below is a brief description of those problems.

- **CRC errors**

April 27, 1998

During rerun of all tests as shown in the test plan document, all SNA and IP connectivity was lost between workstations in building TWO and the rest of the network. The testing being performed was causing PNNI failures between hubs and alternating between sets. The same test has been performed before with no failure. Traces of the network have been taken and CRC error were seen on PNNI connections between hubs and the network. These are also seen in the 8270 diagnostic screens. The problem shifts to different parts of the network.

It was observed ATM cells with invalid PTI and GFC bits. The UNI specs show these values to not be allowed. We were not able to recreate the problem

April 29, 1998

PMR 28052 was opened with IBM service. It was suggested to upgrade to version 3.1.8 that had gone GA.

April 30, 1998

As requested by service, all the 8260s have been upgraded to GA 3.1.8. Also, all the 8270s have been upgraded to release 4.0 and UFC 1.14, and MSS was upgraded to PTF 6.

Since the upgrades we had not been able to recreate the problem. All the testing procedures were executed from the beginning as outlined in the test plan, with no occurrence of the CRC problem.

May 5, 1998

The PNNI test suite was run, and we didn't see any more CRC errors. The test suite involves repeatedly breaking PNNI links back and forth and in a round robin fashion.

The 8270 failure testing was performed and no CRC errors were detected.

- **RIF Field**

April 16, 1998

A large number of IP pings were setup between a workstation on building TWO, domain 60A, and various other workstations on the same subnet and different subnets in building TWO and THREE. The *8270 #8* was powered off and traffic should reroute over *8270 #4*. Pings between the workstations in the same subnet recovered in less than 2 minutes. IP traffic going to other subnets in any building did not recover for up to 30 minutes.

April 21, 1998

Same problem occurred by removing connections from the 8226 hub connected to the 8270. Traces indicated that the 2216 is sending an ARP reply with an 'old' RIF field that has not been updated to reflect the new information in the network

April 23, 1998

The traces have been sent to 2216 development that has acknowledged the origin of the problem

May 5, 1998

A fix has been applied to the 2216 but it has not fixed the problem.

June , 1998

2216 Common Code 3.0 was installed but did not fix the problem. By the time that this document was written, development had identified the problem and scheduled to have it fixed in Common Code 4.0.

- **IP Workstation Recovery Time**

April 17, 1998

Problems were detected with the recovery time of IP workstations recovering from a network failure. After a network failure, which could include pulling a link from a 8274, Windows NT workstations would take up to 5 minutes to recover.

Workstation recovery time is based on several timers in the workstations itself and 8270 and 2216. Reducing the ARP cache time out value to 3 minutes, resulted in routed traffic recovering in approximately 3 to 4 minutes.

April 30, 1998

We discovered that in Windows/NT, the ARP cache value cannot be changed and has the value of 10 minutes.

May 5, 1998

Talking to Microsoft Technical support, we're given a 'hot fix' that allows the ARP cache value to be changed (ArpCacheLife keyword in the registry). This fix will be included in the version 5.0 of Windows NT.

The fix was installed and we did see some improvement .

- **PCOMM does not Recover**

April 17, 1998

After a network failure, Windows NT Workstations running Personal Communications must be stopped and restarted to recover the SNA sessions. A fix was installed with no success. The problem can be recreated all the times.

- **OS/2 PCOMM Crashes After Network Failure**

Network failures involving the powering off of an 8270 or loss of connectivity to an 8270, causes PCOMM and OS/2 to crash. The workstation must be rebooted to be recovered. By the time the document was written, the problem had no fix for it.