

Performance Evaluation of Fibre Channel on a Cluster of Workstations¹

Hong Ong Paul Farrell
Department of Mathematics & Computer Science
Kent State University
Kent, Ohio 44242, U.S.A.

¹This work was supported in part by the Ohio Board of Regents Research Challenge Fund and through use of the equipment of the Ohio Board of Regents Investment Fund Ohio Communication and Computing ATM Research Network (OARnet).

Contents

1	Introduction	1
2	An Overview Of Fibre Channel Technology	1
2.1	Channels And Networks	1
2.2	Fibre Channel ANSI standard	2
2.2.1	Protocol Layers	2
2.2.2	Classes Of Service	3
2.3	Fibre Channel Topologies	4
3	Overview Of The Testing Environment	5
3.1	Hardware Test Environment	5
3.2	Software Test Environment	6
4	Communication Benchmarks	7
4.1	Timing Collection For Evaluating Throughput	8
4.1.1	Timing Results	8
4.1.2	Timing Discussion	8
4.2	Timing Collection For Evaluating Latency	15
4.2.1	Timing Results	15
4.2.2	Timing Discussion	16
5	Concluding Remarks	16
5.1	Future Works	23

List of Figures

1	Fibre Channel ANSI Standard	2
2	Point-to-Point	5
3	Arbitrated Loop	5
4	Switched Fabric	6
5	Hardware Configuration	7
6	Bandwidth Performance from C110 to C110	8
7	Bandwidth Performance from C110 to C180.	9
8	Bandwidth Performance from C110 to J210.	9
9	Bandwidth Performance from C110 to HP9000/735	10
10	Bandwidth performance from C180 to C110.	10
11	Bandwidth performance from C180 to J210.	11
12	Bandwidth performance from C180 to HP 9000/735.	11
13	Bandwidth performance from J210 to C110.	12
14	Bandwidth performance from J210 to C180.	12
15	Bandwidth performance from J210 to HP 9000/735.	13
16	Bandwidth performance from HP 9000/735 to C110.	13
17	Bandwidth performance from HP 9000/735 to C180.	14
18	Bandwidth performance from HP 9000/735 to J210.	14
19	Bandwidth performance from C180 to C180.	15
20	Latency Performance from C110 to C110.	16
21	Latency Performance from C110 to C180.	17
22	Latency performance from C110 to J210.	17
23	Latency performance from C110 to HP9000/735	18
24	Latency performance from C180 to C110.	18
25	Latency performance from C180 to J210.	19
26	Latency performance from C180 to HP 9000/735.	19
27	Latency performance from J210 to C110.	20
28	Latency performance from J210 to C180.	20
29	Latency performance from J210 to HP 9000/735.	21
30	Latency performance from HP 9000/735 to C110.	21
31	Latency performance from HP 9000/735 to C180.	22
32	Latency performance from HP 9000/735 to J210.	22
33	Latency performance from C180 to C180.	23

Abstract

Fibre Channel is the general name of an integrated set of standards being developed by the American National Standards Institute (ANSI). It is designed to significantly improve the speed at which data is transfer between computer systems (i.e workstation, mainframe) and I/O peripherals (i.e storage devices and displays), while providing one standard for networking, storage, and data transmission. In general, Fibre Channel provides inexpensive means of rapidly transferring large volumes of information. The high bandwidth, reliability, flexible topologies, and connectivity offered by Fibre Channel have attracted many users to consider this high-speed network technology for their applications. This paper attempts to evaluate the performance of Fibre channel on a cluster of workstations.

1 Introduction

In recent years, high-performance computers have been receiving a tremendous amount of attention in the data communication industry. The improvement in performance of data communication has encouraged the increase of data intensive and distributed networking applications, such as multimedia, distance learning, and scientific visualization. However, until recently networks interconnecting computers and peripheral devices were unable to run at the speed required by these applications.

Fibre Channel is one of the emerging high speed network technologies that can meet many requirements related to the ever increasing need for data and communications-intensive applications being developed for business, government and academic institutions. Fibre Channel is the general name of an integrated set of standards [1] being developed by the American National Standard Institute (ANSI). The standards address the need for very fast transfer of large volumes of information, and define a high-speed interface channel that can be used to interconnect computer systems and peripherals devices. It provides one standard for existing channel and network protocols to be operated simultaneously over Fibre Channel. Consequently, the standard relieves system manufactures from the burden of supporting a variety of channel and network protocols for storage, and data transfer.

This paper evaluates the performance characteristics of Fibre Channel on a cluster of HP workstations connected by a Fibre Channel Fabric based on a 16 port Ancor switch. In the remainder of this paper, Section 2 gives an overview of Fibre Channel. Section 3 describes the hardware and software test environment. In section 4 the end-to-end communication characteristics (i.e latency and throughput) are shown and analyzed, and section 5 summarizes the experience that we gained from the experiment and indicates future work.

2 An Overview Of Fibre Channel Technology

To give a general overview of Fibre Channel technology, we will organize this section into the following subsections: Section 2.1 Channels and Networks, Section 2.2 Fibre Channel ANSI Standard, Section 2.3 Fibre Channel topologies.

2.1 Channels And Networks

Basically, there are two types of data communications namely channel and network [2]. A channel is a direct or switched point-to-point connection between communication entities. Its primary duty is to transport data from one point to another at the highest speed with lowest possible delay, and perform error correction in hardware. If a data transfer fails because of network congestion, a channel retries immediately without consulting software. It is inherently suitable for an environment where there is little or no decision making, and the destination addresses are predefined. Channels have much lower delay than networks since they are hardware intensive. In contrast, networks link distributed nodes with a communication protocol that supports communication among these nodes. It is usually software intensive which tends to have higher overhead in data transmission. As a result, it is slower than a channel. The benefit of networks is that they can handle a wider-range of tasks as they can automatically adjust to meet an environment with varying or unanticipated connections.

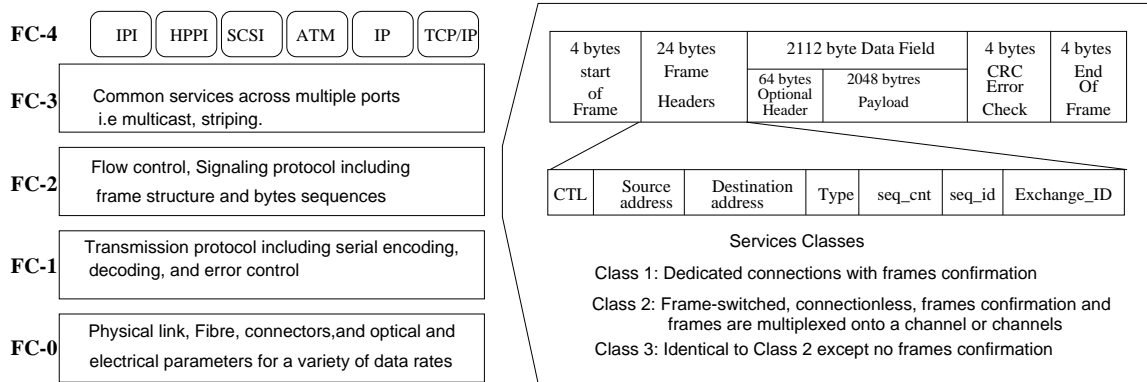


Figure 1: Fibre Channel ANSI Standard

Fibre Channel is designed to combine the advantages of both channel and network data communication. It employs a simple technique that avoids the issue of the difference between channel and network protocols. It provides a mechanism to transfer data reliably between one buffer at the source device to another buffer at the destination device. Fibre Channel does not process the buffer at all. It simply takes what is in the sending buffer and transports it to the receiving buffer. The individual network protocols can process the buffer before or after it is sent or received. Fibre Channel provides complete control over the transfer layer only and offers error checking.

2.2 Fibre Channel ANSI standard

In 1988, the American National Standard Institute (ANSI) X3T9.3 committee formed the Fibre Channel working group to develop a practical, yet expandable scheme to attain high speed data transfer among communication entities. The Fibre Channel working group was assigned full ANSI status in 1994.

The Fibre Channel standard is organized into five functional layers and three service classes. The following sub-sections attempt to give a brief discussion of this organization.

2.2.1 Protocol Layers

The Fibre Channel standard is defined as a multi-layered stack of functional levels (Figure 1), not unlike those used to represent network protocols. These layers do not map directly to OSI layers. The layers of the Fibre Channel standard define the physical media and transmission rates, encoding scheme, framing protocol and flow control, common service, and the upper-level applications interfaces.

- **FC-0** defines the physical media of the Fibre Channel including the physical characteristics of the media transmitters, receivers and connectors, the electrical and optical characteristics, the transmission rates, and other physical components of the standard. This layer deals with

performance and cost issues. It provides a variety of physical media to address variations in the physical cabling requirements. Coax cable and shielded twisted pair are defined for limited-distance applications. These wide-range of alternatives allow system administrators to tailor the installation to meet the specific requirements of users.

- **FC-1** defines the transmission protocol which includes the serial encoding, decoding, and error control. The 8B encoding and 10B decoding scheme is used to integrate the data with the clock information required by serial transmission techniques. Fibre Channel uses 10 bits to represent each 8 bits of upper level data. Thus, it must operate at a speed sufficient to accommodate this 25 percent overhead.
- **FC-2** defines the signaling protocol which includes the frame structure and byte sequences. All frames belonging to a single transfer are uniquely identified by sequential numbering from 0 through n . This scheme enables the receiver to determine which frame is missing, if there is any.
- **FC-3** defines a common set of services required for advanced features such as striping to increase bandwidth, allowing multiple ports to respond to the same alias address, and multicasting.
- **FC-4** is the highest level in the Fibre Channel standards set. It defines the mapping, between the lower levels of the Fibre Channel and the IPI and SCSI command sets, the HIPPI data framing, IP, and other Upper Level Protocols (ULPs).

As Fibre Channel is applied to other applications, many more application interfaces are expected to be defined.

As a result of the 10 bit encoding and the protocol overhead, the 266 Mbps Fibre Channel protocol tested in this paper has a maximum data throughput of 25 Mbytes/sec or 250 Mbps. This does not take into account any further theoretical loss in performance due to the overhead of higher level protocols such as IP, UDP or TCP or any limitations in attainable throughput due to processor or bus throughput constraints.

2.2.2 Classes Of Service

Fibre Channel defines three different classes of services to accommodate the wide range of communication needs. All classes of service can be applied to all topologies; and, each class can be integrated within a framework of other network protocols.

- **Class 1** service is a dedicated connection between two ports. It behaves in much the same way as today's dedicated physical channels. When two communication devices are linked together using Class 1, the communication entities can use the full bandwidth of the connection. All links which constitute the connection are used exclusively for the connection. No other network traffic affects this communication. Frames are guaranteed to arrive in the order in which they are transmitted regardless of the network topology. When the time needed to make a connection is short or data transmissions are long, class 1 is an ideal link.

- **Class 2** service does not have a dedicated connection. However, it provides guaranteed delivery with an acknowledgment of receipts. There is no delay in establishing a connection as in Class 1. Also, no uncertainty exists as to whether or not delivery was achieved. If delivery cannot be made because of congestion, a busy frame is returned, and the sender tries again. As with traditional packet-switched systems, the route between the two nodes is not dedicated which allows better utilization of the bandwidth of the link. Class 2 is ideal for data transfers between a shared mass-storage system physically located at some distance from several individual workstations.
- **Class 3** is a connectionless service that allows data to be sent rapidly to multiple recipients, but no confirmation of receipt is given. This class is most practical when it takes a long time to make a connection. The software layers must determine whether or not data has been lost. If the software detects that data has been lost, then it would retransmit the data. Class 3 service is very useful for real time broadcasts, where timeliness is a prime importance and where information not received on time has little value.

2.3 Fibre Channel Topologies

Fibre Channel provides three connection methods, namely point-to-point, arbitrated loop, and switched fabric to interconnect systems and I/O devices. Each connection method has its own strengths and weaknesses. Since the transmission medium is isolated from the control protocol, each implementation may use a topology best suited to the environment of use. In the following paragraphs, we describe the three topologies mentioned.

- **Point-to-Point**

The simplest of all connection methods is point-to-point. This method involves a simple connection between two systems. Point-to-point uses a single, full-duplex cable between the two devices. Because no intermediate devices exist and the connection is limited to two nodes, the point-to-point connection method provides the greatest possible bandwidth and the lowest latency. Figure 2 illustrates this connection method.

- **Arbitrated Loop**

An arbitrated loop connects up to 126 devices in a ring. Arbitrated loop is similar to token ring, where each device arbitrates for loop access, and once granted, has a dedicated connection between sender and receiver. The available bandwidth of the loop is shared between all devices. The primary reason to use arbitrated loop is for cost. Since no switch is required to connect multiple devices, the per connection cost is significantly less than it is with the switched fabric method. Arbitrated loops are inherently inefficient in large configurations. This is because every node in the loop must look at the data regardless of the destination. Because each node is logically connected in a circle, significant cable induced delays are possible with large configurations. Figure 3 shows the arbitrated loop method.

- **Switched Fabric**

A switched fabric topology provides the greatest connection capability and largest total aggregate throughput of the topologies discussed. In this connection method, each device is

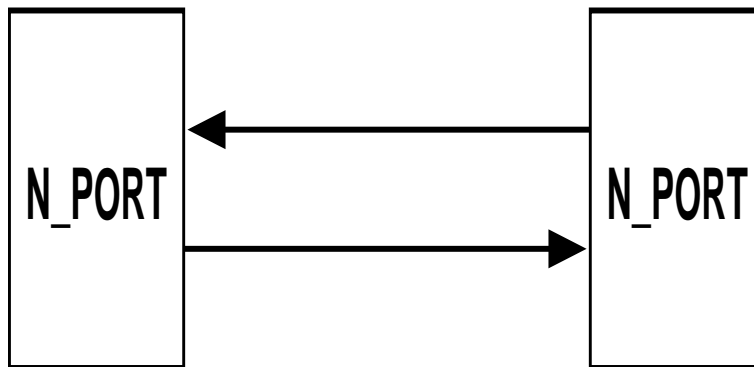


Figure 2: Point-to-Point

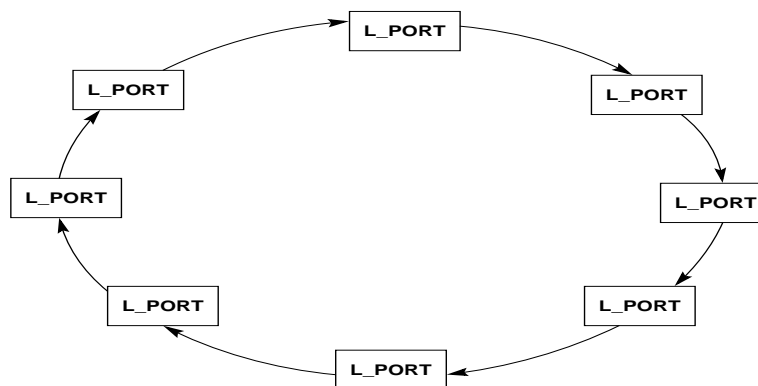


Figure 3: Arbitrated Loop

connected to a switch and receives a nonblocking data path to any other connection on the switch. This setup is equivalent to a dedicated connection to every device. As the number of devices increases to occupy multiple switches, the switches are, in turn, connected together. Multiple connection paths between switches are recommended to provide circuit redundancy and increase total bandwidth. Figure 4 illustrates this connection method.

3 Overview Of The Testing Environment

3.1 Hardware Test Environment

The experiments conducted to evaluate the communication throughput and latency were performed on a clusters of six HP workstations. The following Figure 5 shows the hardware configuration that was used in the tests described here. This consists of a C180, a J210, three C110, and a 9000/735.

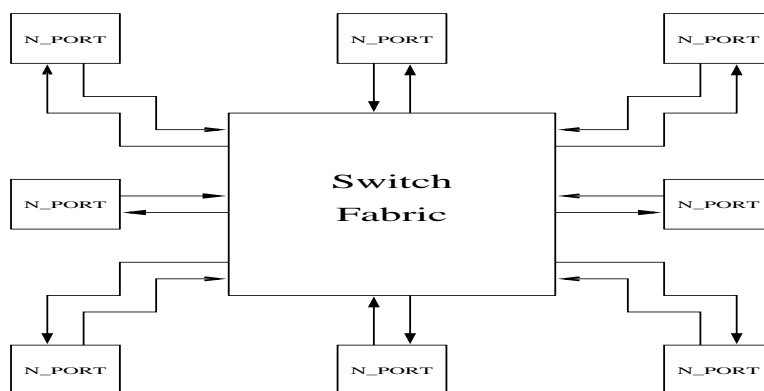


Figure 4: Switched Fabric

The C110 workstations are running at 120 MHz with 128 MB of RAM. The J210 workstation is also running at 120 MHz but with 3x128 MB of RAM. The C180 has the same amount of RAM as the C110 but running at 180 MHz. The 9000/735 workstation is running at 125 MHz with 128 MB of RAM. These workstations are connected to an Ancor FCS 266/16 Fibre Channel switch¹, and the Fibre Channel interface cards are installed in EISA slots. All the machines run the HP-UX 10.20 operating system except HP 9000/735 which runs HP-UX 10.01. The machines are in a normal departmental LAN environment rather than an isolated testbed. All the tests were performed when the machines were lightly to moderately loaded.

3.2 Software Test Environment

The software program used to test the communication performance was *Netperf* (version 2.1). This program can be obtained from <ftp.cup.hp.com>. *Netperf* is a software benchmark program that is widely used to measure the performance of many different types of networking. It has a very flexible and convenient user interface to test for both throughput, and latency. In [5], communication benchmarks are also obtained using this software. In the tests performed here, *Netperf* was run using the following command line

```
netperf -l 10 -H $REMOTEHOST -t TCP_STREAM --
-D -m $SEND_SIZE -s $SOCKET_SIZE -S $SOCKET_SIZE
```

to investigate the throughput of the switch and

```
netperf -l 10 -H $REMOTEHOST -t TCP_RR --
-D -m $SEND_SIZE -s $SOCKET_SIZE -S $SOCKET_SIZE
```

to obtain the delay between two communication nodes. The input parameters² to this program are:

¹For more detail description of this particular switch, please visit <http://www.ancor.com>

²For full description of *Netperf* parameters, please refer to [7]

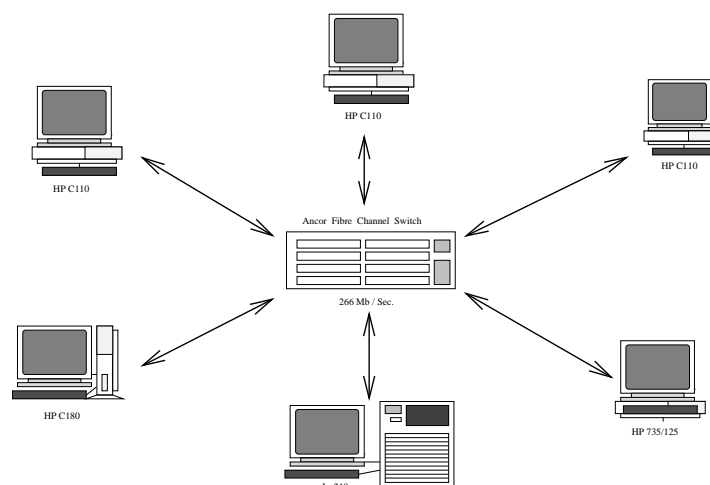


Figure 5: Hardware Configuration

```

-H - the host machine that we are testing.
-m - the size of the message to be sent.
-s - the socket buffer size on the local machine
-S - the socket buffer size on the remote machine
-l - the amount of time to run the test.
-D - set nodelay option

```

The size of the message to be transferred ranges from 1 byte to 2 Mbytes. There is no limitation of message size at the user level. The socket buffer size ranges from 4096 bytes to 262143 bytes which is the maximum socket size configurable in the HP-UX 10.20 kernel. Each test was run for a duration of 10 seconds. There are several parameters can be set to affect the performance of TCP/IP. *RFC 1323* [8] defines a set of TCP extension to improve performance over large *bandwidth* \times *delay* paths and to provide reliable operation over high-speed paths. On all machines, we configured the High-Speed TCP/IP Extension (*RFC 1323*) to increase the performance of TCP/IP over Fibre Channel. Other parameters such as the maximum socket buffer size, and the window size on both receiver and sender were modified to fully utilize the advantages of *RFC 1323*. The `TCP_NODELAY` option was set on the sender side for these measurements to avoid grouping and buffering messages before sending. It is necessary to turn on this option to ensure each message is sent immediately. Otherwise, we might over-estimate performance in the test results. In addition to the *Netperf* software, a script file was written locally to facilitate the testing.

4 Communication Benchmarks

In this section we will present the results that we obtained from running the experiments; and, we will also attempt to explain and analyze the results obtained where appropriate.

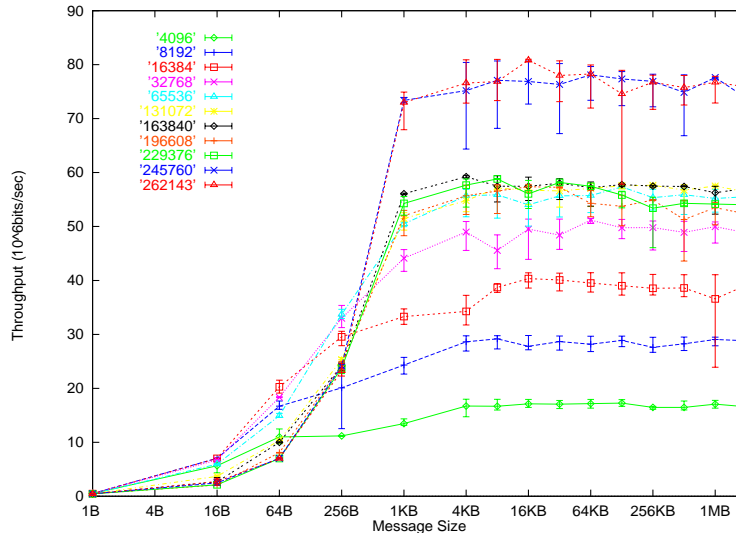


Figure 6: Bandwidth Performance from C110 to C110

4.1 Timing Collection For Evaluating Throughput

4.1.1 Timing Results

We will define the throughput or bandwidth to be the rate at which the Fibre Channel can transmit data. The results are reported in Mbits per second as it is widely used among vendors. In order to investigate the stability of the results, we repeated each test five times. We performed tests to determine how the performance varied with message size, socket buffer size, and CPU speed. Figure 6 through Figure 19 reported the average throughput obtained by running *Netperf* among the different sets of machines with different message and socket buffer size. The figures consist of individual graphs for each socket size. These graphs give the average throughput for each message size and also include error bars at each message size to indicate the maximum and minimum of the five tests taken. In the following section, we will discuss these throughput characteristics further.

4.1.2 Timing Discussion

Figure 6 through figure 18 showed the achievable user-level bandwidth characteristic of Fibre Channel on different sets of HP workstations. There are several interesting observations that are worth mentioning.

- As the message size increases, the Fibre Channel's throughput attains approximately 82 MBits/sec with message size equal to 2MBytes. However, with HP 9000/735 as the sender, the maximum attainable throughput is roughly equals to 67 MBits/sec. This can be explained by the fact that the faster machines (i.e C110, C180, and CJ210) can process the message

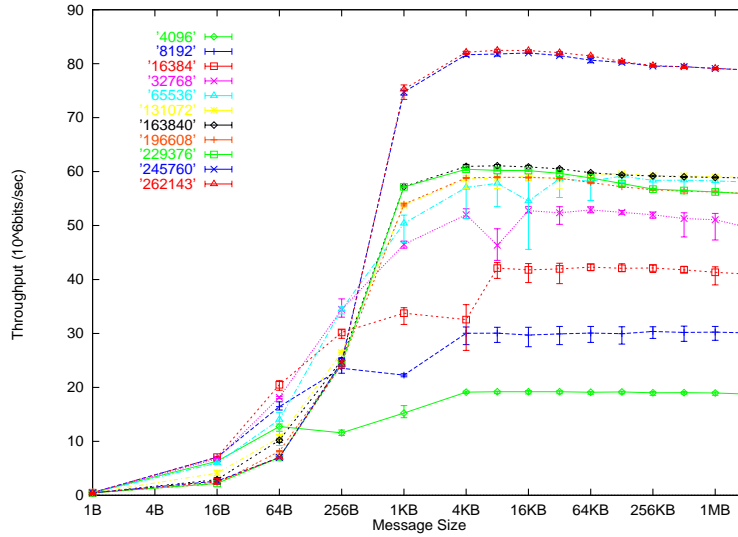


Figure 7: Bandwidth Performance from C110 to C180.

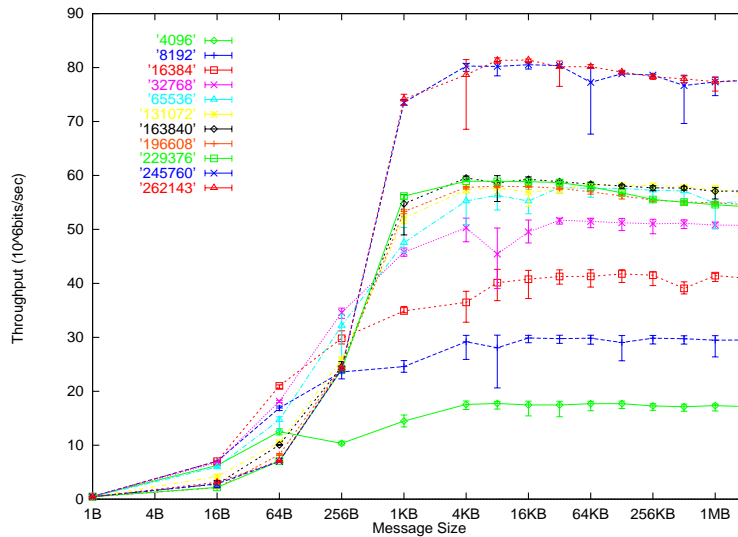


Figure 8: Bandwidth Performance from C110 to J210.

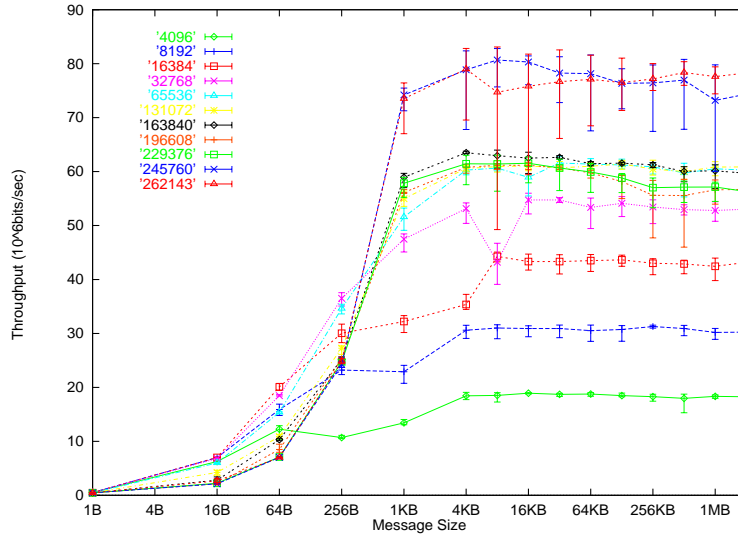


Figure 9: Bandwidth Performance from C110 to HP9000/735

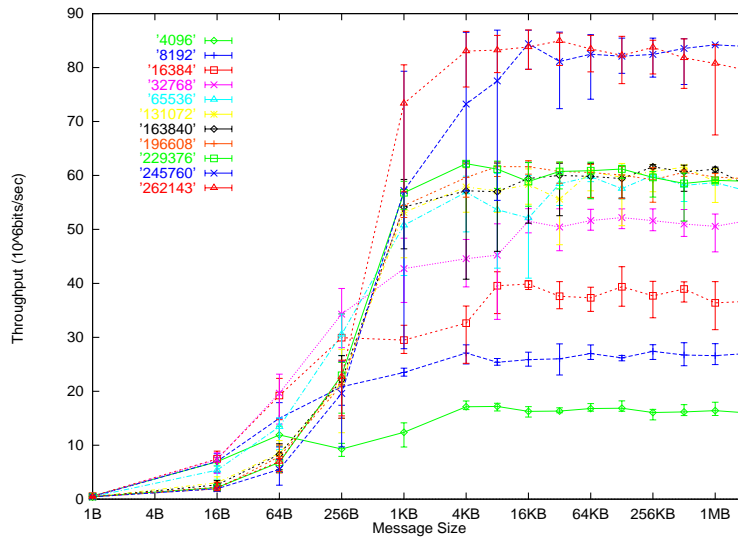


Figure 10: Bandwidth performance from C180 to C110.

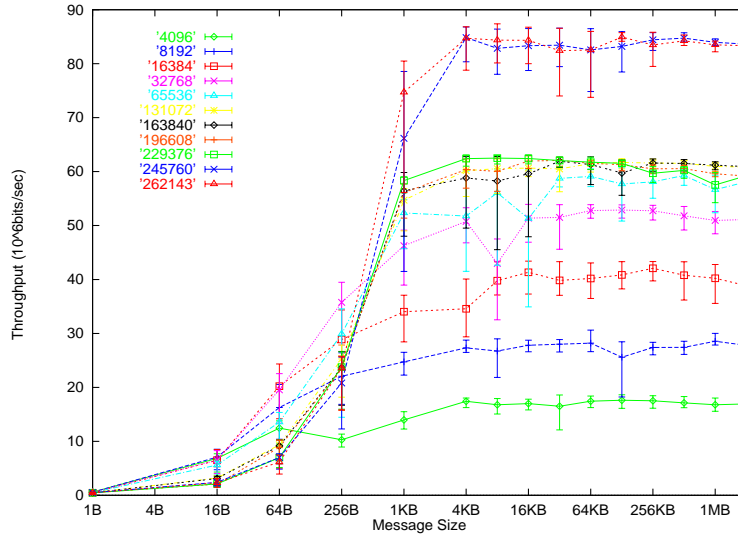


Figure 11: Bandwidth performance from C180 to J210.

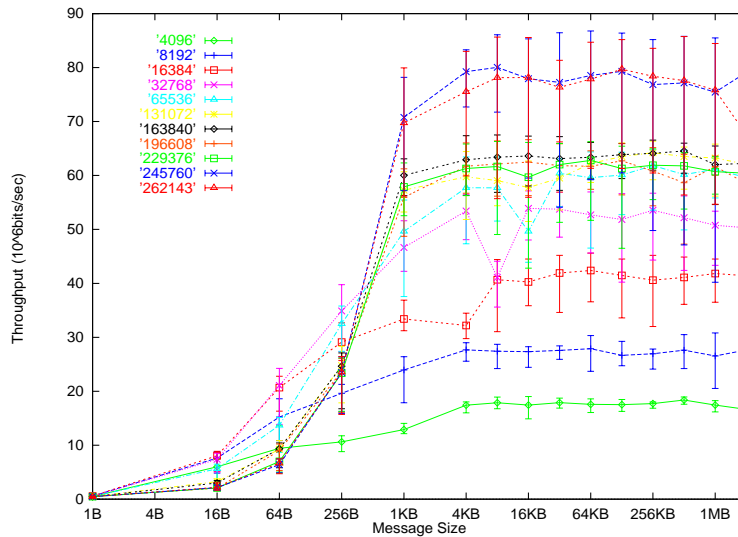


Figure 12: Bandwidth performance from C180 to HP 9000/735.

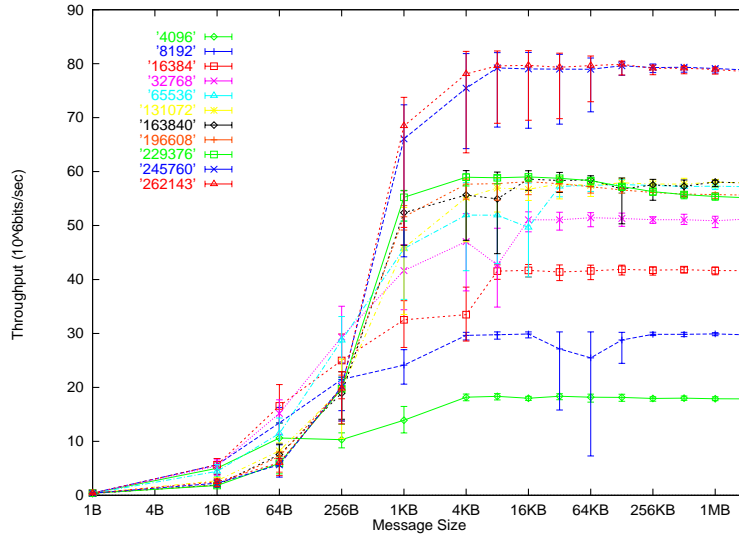


Figure 13: Bandwidth performance from J210 to C110.

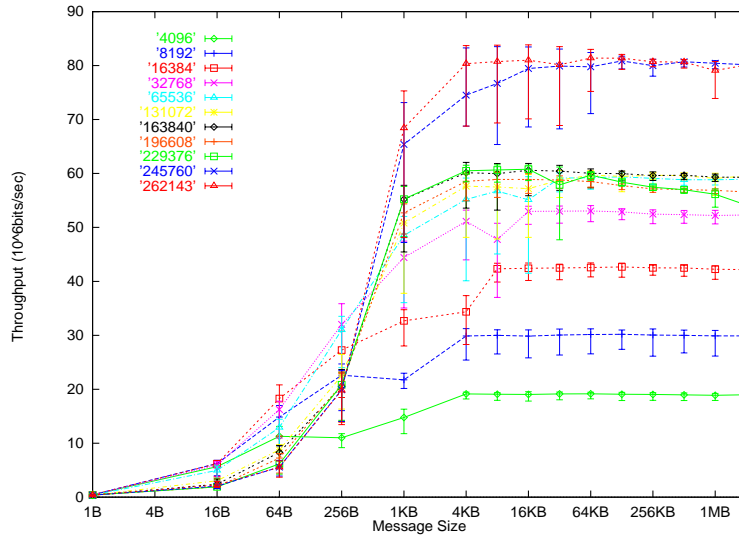


Figure 14: Bandwidth performance from J210 to C180.

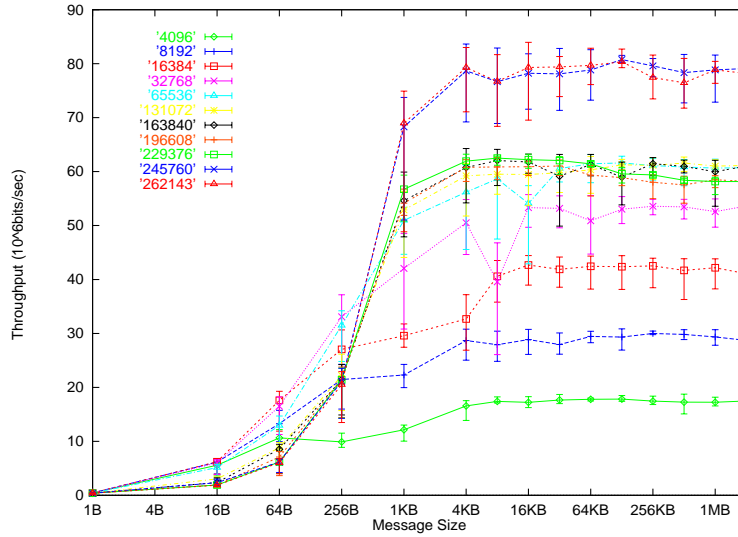


Figure 15: Bandwidth performance from J210 to HP 9000/735.

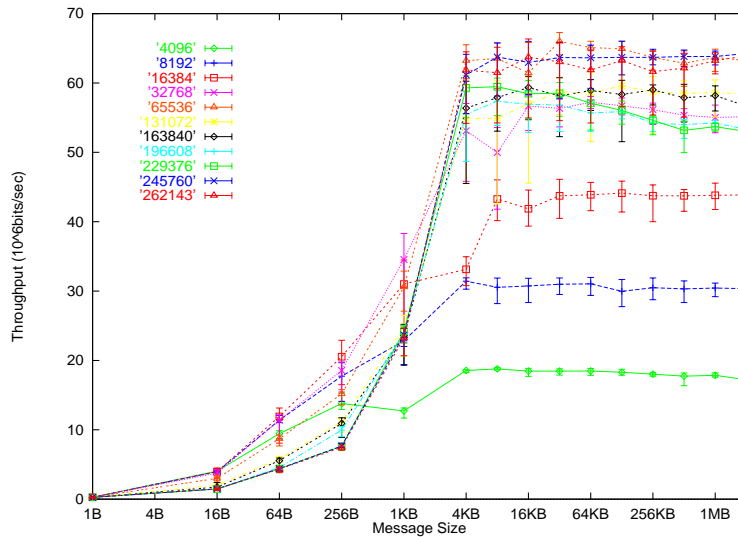


Figure 16: Bandwidth performance from HP 9000/735 to C110.

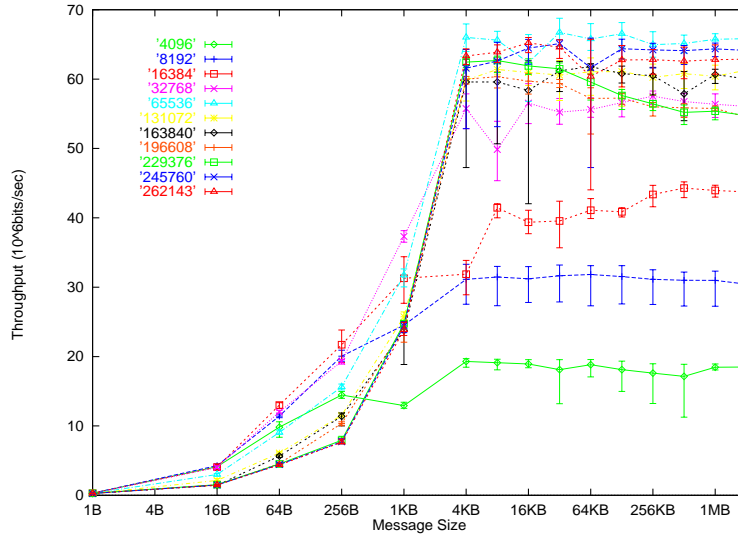


Figure 17: Bandwidth performance from HP 9000/735 to C180.

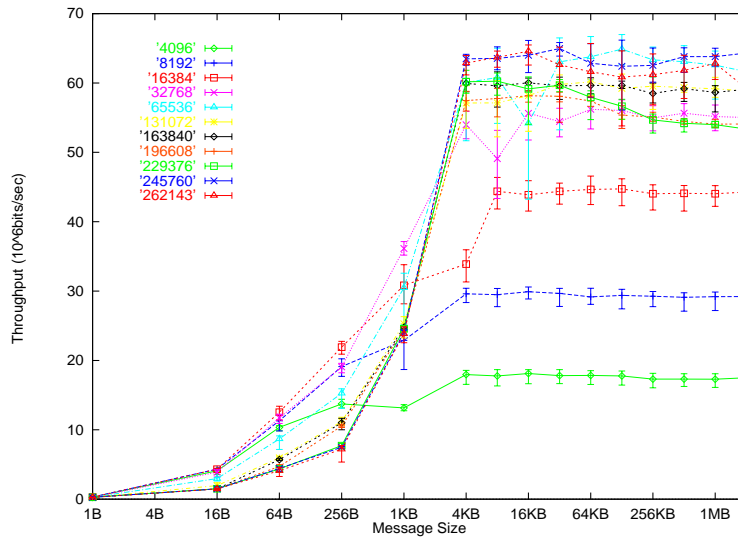


Figure 18: Bandwidth performance from HP 9000/735 to J210.

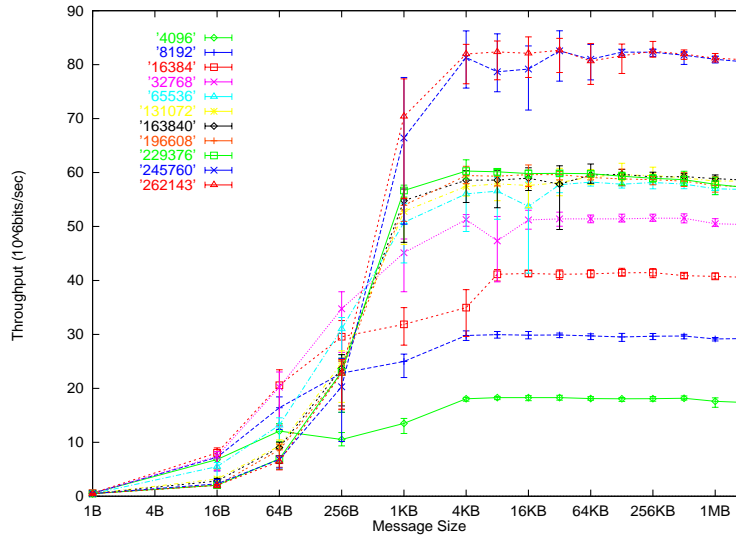


Figure 19: Bandwidth performance from C180 to C180.

buffers faster at the protocol level. Hence, they are able to transfer the message buffer to the Fibre Channel interface faster thus achieving higher throughput.

- To attain throughput of 82Mbps the socket buffer size must be set sufficiently large. In the case of our tests this was 245760 or 262143 bytes. With socket size varying between 64KB and 224KB throughput of approximately 60 Mbps was attained even on the fastest machines. This implies that socket buffer size must be set close to the maximum in order to have good throughput performance. In general, TCP performance depends on the $bandwidth \times delay$ product. This product is the amount of data that can occupy the communication links. In order to keep the Fibre Channel communication link full, a larger TCP window size is required. *RFC 1323* addressed this issue by defining a *TCP Window Scale* option to expand the size of the TCP window.
- With socket buffer size equals to 32KB, there is a sudden drop in throughput when the message buffer reaches 8KB. This phenomenon is consistent throughout all experiments and was also observed in [3]
- Message size greater than 512 bytes has much higher throughput and the peak throughput is attained for message as small as 4096 bytes.

4.2 Timing Collection For Evaluating Latency

4.2.1 Timing Results

We will define the communication latency to be the time from when a sending application initiates transfer until the receiving application finishes receiving it. The average latency graphs are drawn

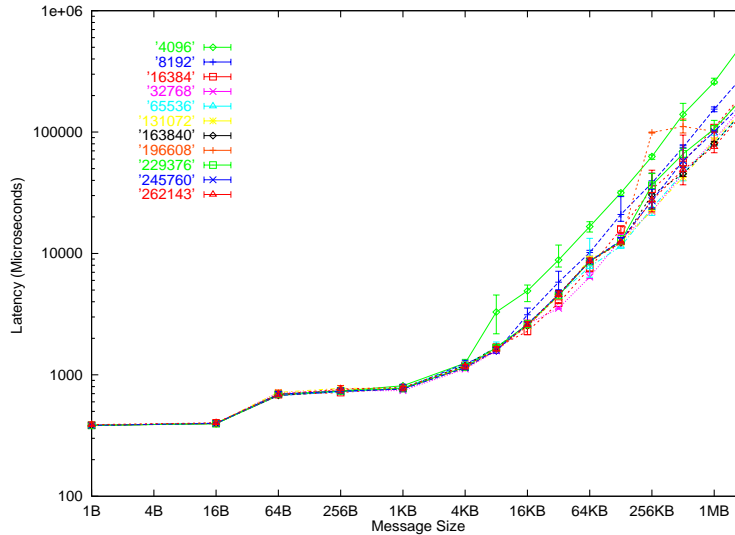


Figure 20: Latency Performance from C110 to C110.

with error bars to show the maximum and minimum delay obtained from running five tests on the different sets of machines. Figure 20 through Figure 33 show the results obtained. In Section 4.2.2, we will discuss the timing results collected.

4.2.2 Timing Discussion

From figure 20 through figure 33, we can see that there is not much variation in latency across the different machines sets. In addition, at least per socket sizes greater than 16KB, there is negligible difference in delay for the same message size with different socket size. Taking the time to send a one byte message as a measure of latency of the network, we can say that the latency of the Fibre Channel network is approximately 340μ secs. Over the various machines, it varies between 336μ secs and 354μ secs on unloaded machine, with outlying of 398μ secs to 460μ secs on moderately loaded machines such as the J210 or HP 9000/735.

5 Concluding Remarks

In this paper, we performed experiments to evaluate a high-speed network based on Fibre Channel protocol. Graphs were presented to report the throughput and latency characteristic of Fibre Channel. The delay characteristic of Fibre Channel is deterministic and the maximum user-level throughput could be achieved by taking advantage of the features of *RFC 1323*. High-speed networks such as Fibre Channel have shifted the features of communication overhead from physical media to protocol processing. Therefore, it is critical to improve the performance of protocol processing in order to achieve high-speed transmission.

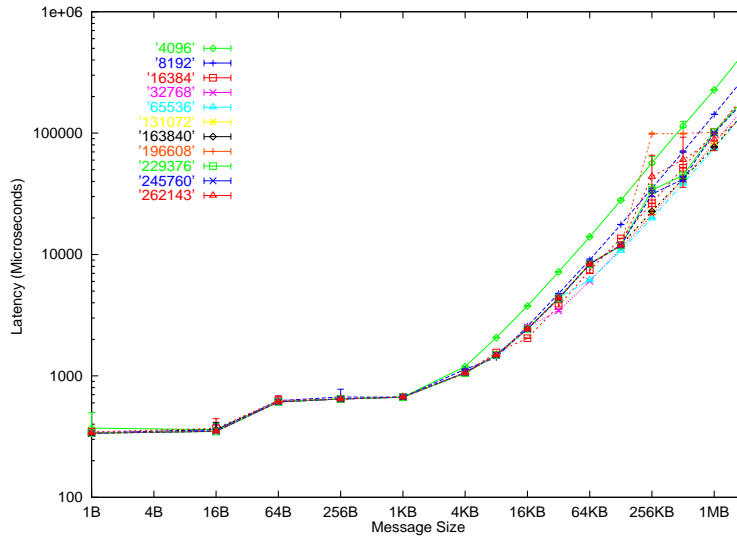


Figure 21: Latency Performance from C110 to C180.

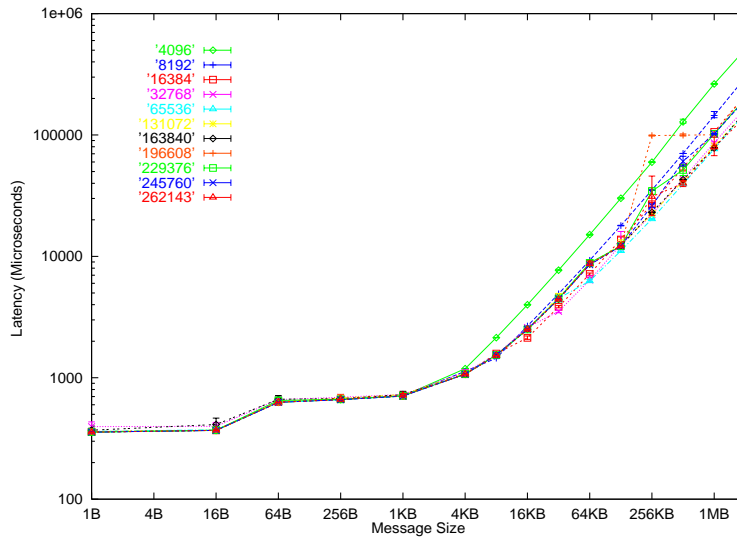


Figure 22: Latency performance from C110 to J210.

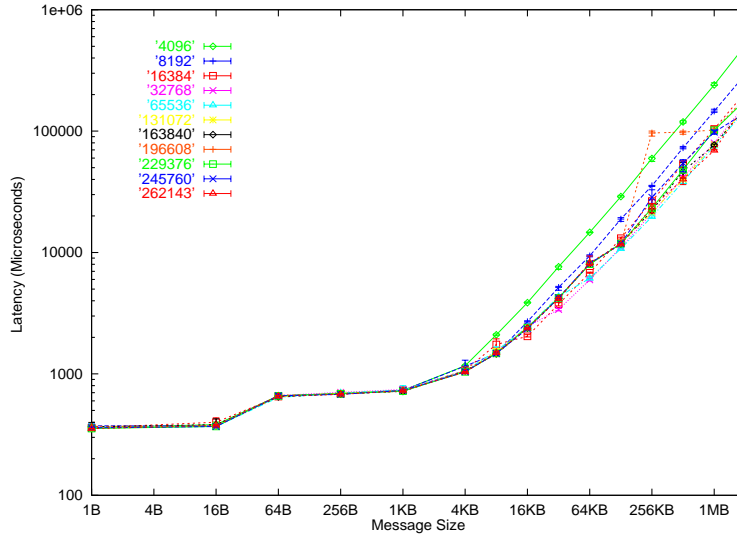


Figure 23: Latency performance from C110 to HP9000/735

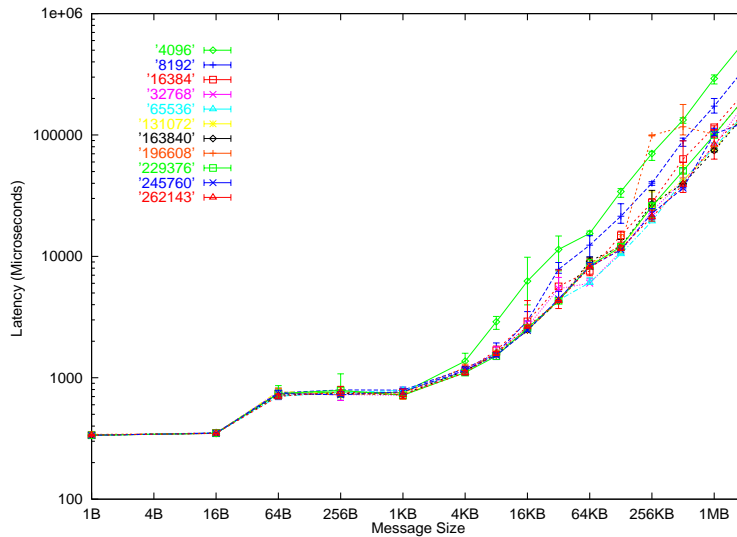


Figure 24: Latency performance from C180 to C110.

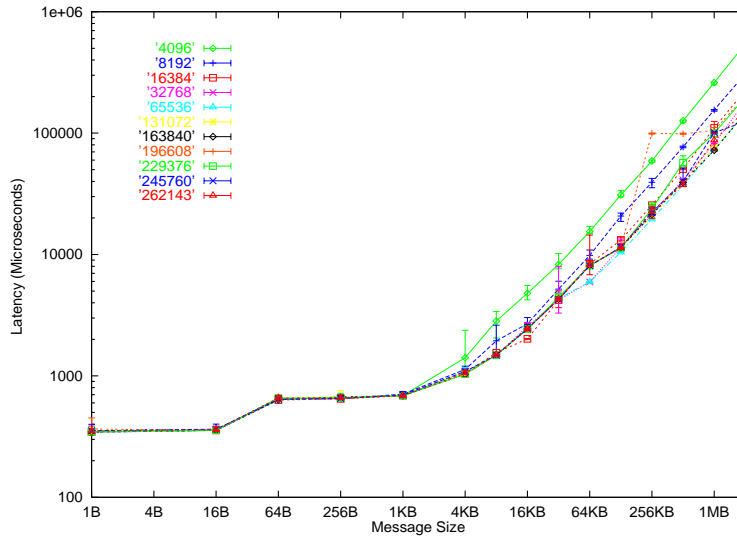


Figure 25: Latency performance from C180 to J210.

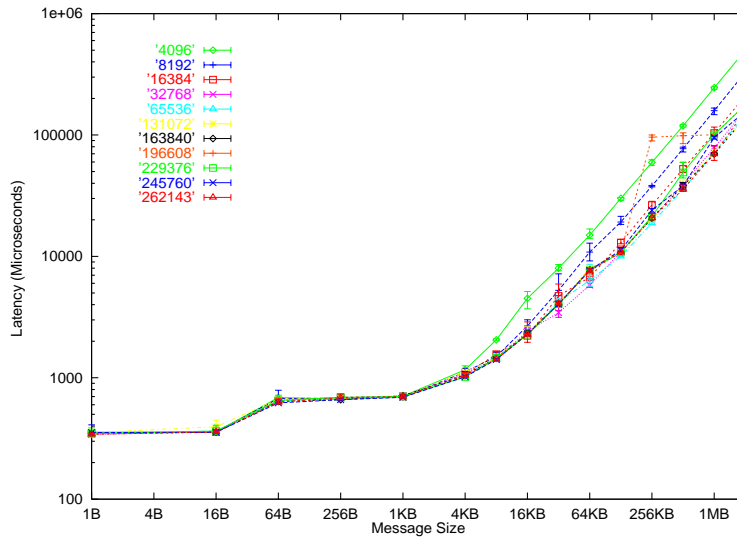


Figure 26: Latency performance from C180 to HP 9000/735.

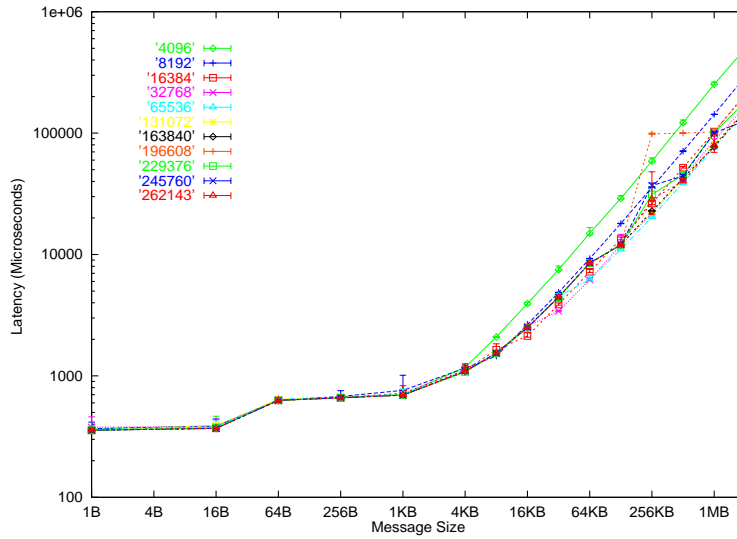


Figure 27: Latency performance from J210 to C110.

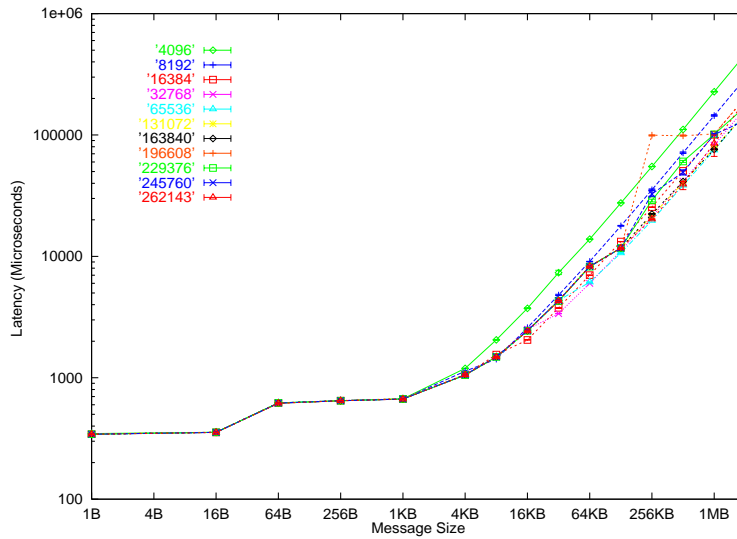


Figure 28: Latency performance from J210 to C180.

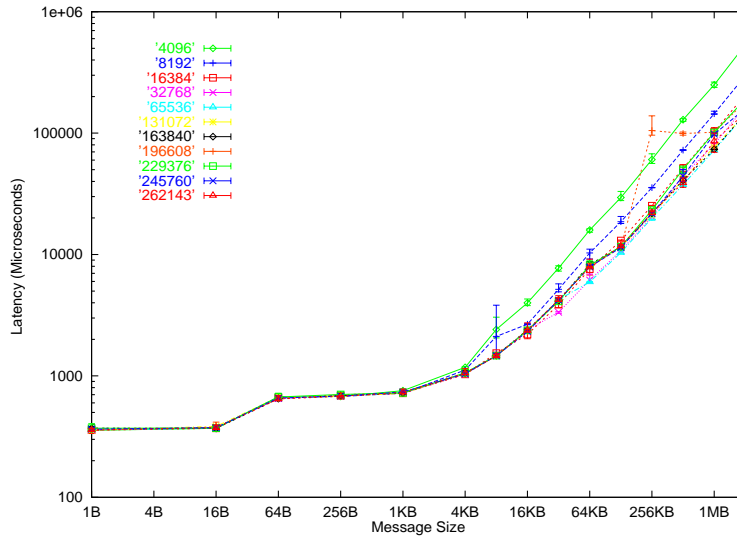


Figure 29: Latency performance from J210 to HP 9000/735.

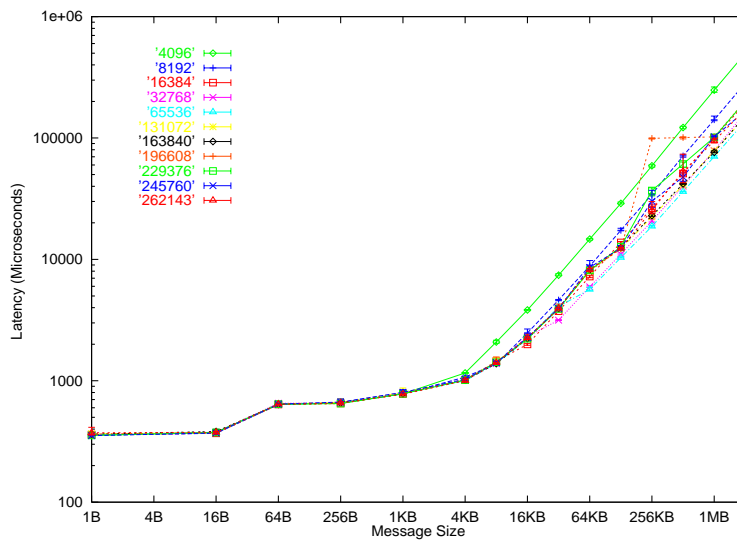


Figure 30: Latency performance from HP 9000/735 to C110.

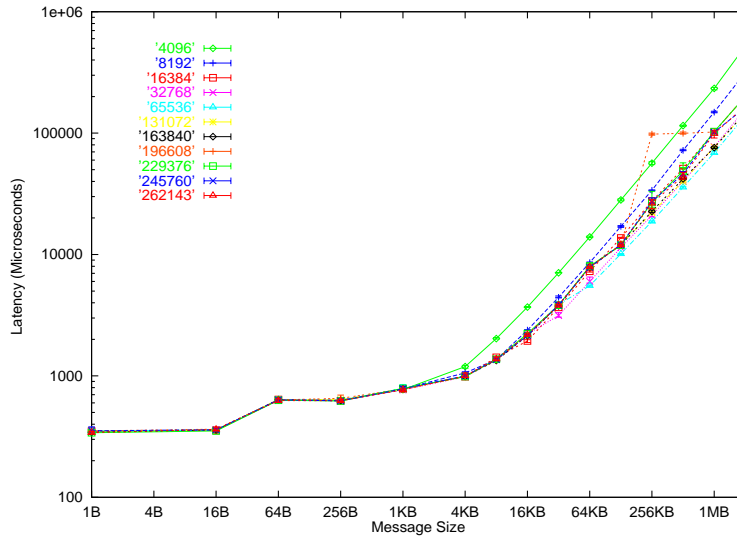


Figure 31: Latency performance from HP 9000/735 to C180.

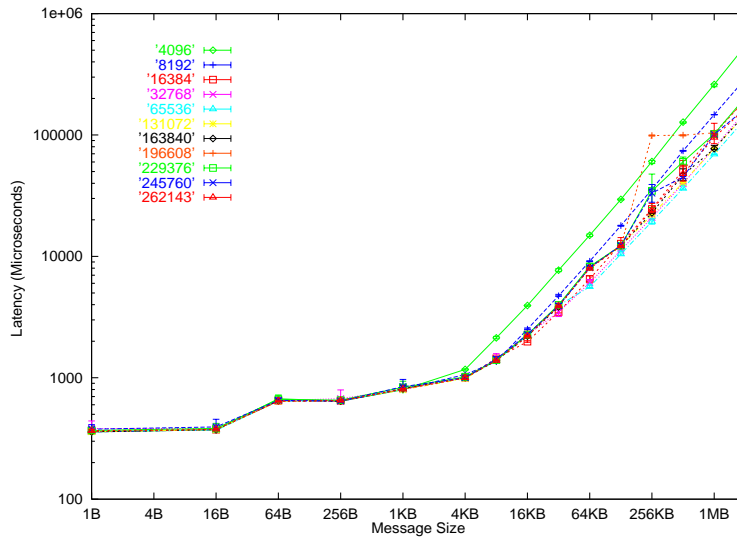


Figure 32: Latency performance from HP 9000/735 to J210.

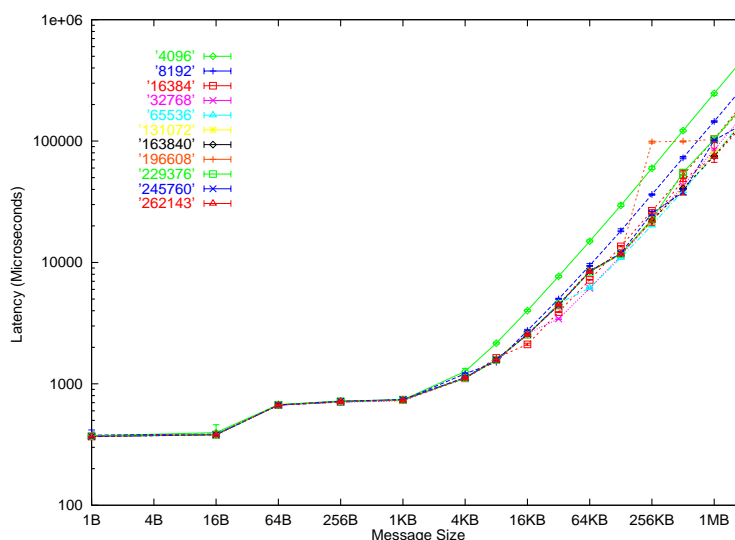


Figure 33: Latency performance from C180 to C180.

5.1 Future Works

Our future experiments will include monitoring the Fibre Channel device driver in order to understand how it interact with other components of a network sub-systems and testing how message-passing library such as PVM and MPI operates under Fibre Channel.

References

- [1] ANSI X3.230-1994, "Fibre Channel Physical and Signalling Interface (FC-PH) Rev4.3"
- [2] The University of New Hampshire InterOperability Lab. "Basic Training in Fibre Channel Technology". NetWorld+Interop'96 in Las Vegas 1996.
- [3] James A. MacDonald. "Performance Evaluation of Two High-Performance Local Area Network: Fibre Channel vs. ATM". March, 1996.
- [4] Carlo Miron, Fabrice Chantemargue, Susana Munoz. "Fibre Channel Performances with IBM Equipment". June, 1995.
- [5] Fabrice Chantemargue. "Communication benchmarks with the Ancor Fibre Channel Fabric". CERN/EAST note 94-22, July 1994.
- [6] Fabrice Chantemargue. "High Performance TCP/IP Communication benchmark with HPs and IBMs connected around an Ancor Fibre Channel Fabric". November 1994.

- [7] "Netperf: A Network Performance Benchmark. Revision 2.1" Information Networks Division. Hewlett-Packard Company, February 15, 1996.
- [8] Jacobson, Braden, & Borman "TCP Extensions for High Performance" RFC1323, May 1992.