CERN/EAST note 94-24
August 29th 1994
draft

# Communication benchmarks with HPs and IBMs connected around an Ancor Fibre Channel Fabric

Fabrice Chantemargue

## I. Introduction

This note is aimed at providing communication times for transferring data from memory to memory between *HP/9000* workstations, and between a *HP/9000* and an *IBM RS/6000*, via a 16 port Ancor Fibre Channel Fabric. Some similar measurements for communications between *IBM* workstations only have already been achieved and they are reported in [1]: therefore, they will not be addressed in this paper.
Figure *1* shows the experimental hardware configuration now available. The two *HP* machines are *700 series* with a processor running at 100 MHz. The *IBM RS/6000* workstation pool is composed of a *590* model with a *Power2* processor running at 66 MHz, a *250* model with a *PowerPC 601* at 66 MHz and a *C10* model with a *PowerPC 601* at 80 MHz. Two kinds of adapters, depending on the machine in which they are intended to be located, can be used to provide access to the Fibre Channel world:

- an *EISA CIM* (Extended Industry Standard Architecture Communications Interface Module) *250* Ancor adapter for the *HP* workstations,
- a *MCA CIM* (Micro Channel Architecture Communications Interface Module) *250 Ancor* adapter for the *IBM* workstation.

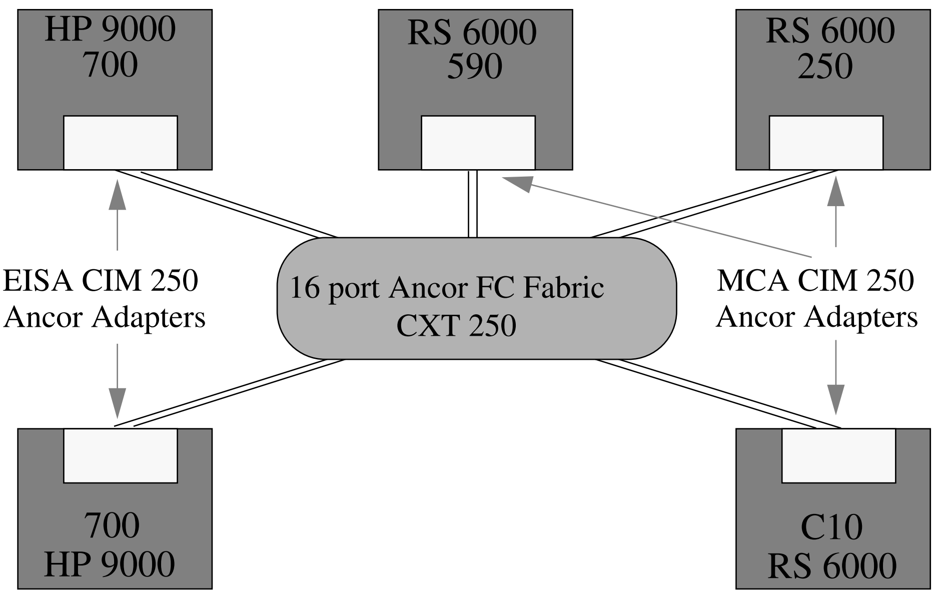


Figure 1: Experimental hardware configuration.

The *EISA CIM 250* interface to the *EISA* bus (*HP 9000/700* workstations) is a 32 bit bus Master capable of data transfer rates up to 33 MBytes/s and it should not represent a bottleneck in this system, where the speed of each 50 micron optic fibre is 266 Mbits/s. Let us just recall the *MCA CIM 250* interface to the *MCA* bus (*RS/6000* IBM workstations) is a 32 bit bus Master capable of data transfer rates up to 40 MBytes/s. For more details about the Fibre Channel Standard, the Fibre Channel fabric and adapters, see [2], [3], [4] and [5].
The driver available for *EISA CIM 250* adapters can only support today the *TCP/IP* protocol. The *Direct Channel* interface feature of the driver, which will allow the replacement of the *TCP* and *IP* levels of protocol with a direct connection to the driver, thereby achieving the highest possible bandwidth, is announced to be the next supported interface (no date on the availability). By the way, note that the driver available for *MCA CIM 250* boards supports both *TCP/IP* and *Direct Channel* protocols.

# II. Results using *TCP/IP*

The program used to test out the performance is '*ttcp*'. This program is in the public domain and was originally created by the US Army Ballistics Research Lab (BRL).
There are three input parameters to this program, which can be chosen independently:

- the size of the message to be sent,
- the number of times the message is sent (to perform an average measurement),
- the socket buffer size.

The size of the message to be transferred ranges from 1 byte up to 2 Mbytes. There is no limitation on the message size at the user level.
The number of times the message is sent is equal to $10^4$ if the message size is lower than $10^5$ bytes. In the case of a message size greater than $10^5$ bytes, the number of times the message is sent is equal to $10^3$. These values have been chosen like this in order to provide relevant time measurements.
Several values (in bytes) for the socket buffer size have been tested according to the experiments: 4K, 8K, 16K, 32K and 64K (1K is equal to 1024 bytes).

## A. Communications between *HP* workstations

Graphs show hereinafter how the communication time and therefore the communication speed scale with the size of the message to be transferred and the socket buffer size.
Figure *2a* displays speed measurements, whilst figure *2b* displays communication times. As the two *HP* machines are identical, there is practically no difference in the measured performances when the direction of transmission is inverted: that is the reason why we present only some measurements achieved for one direction of communication.

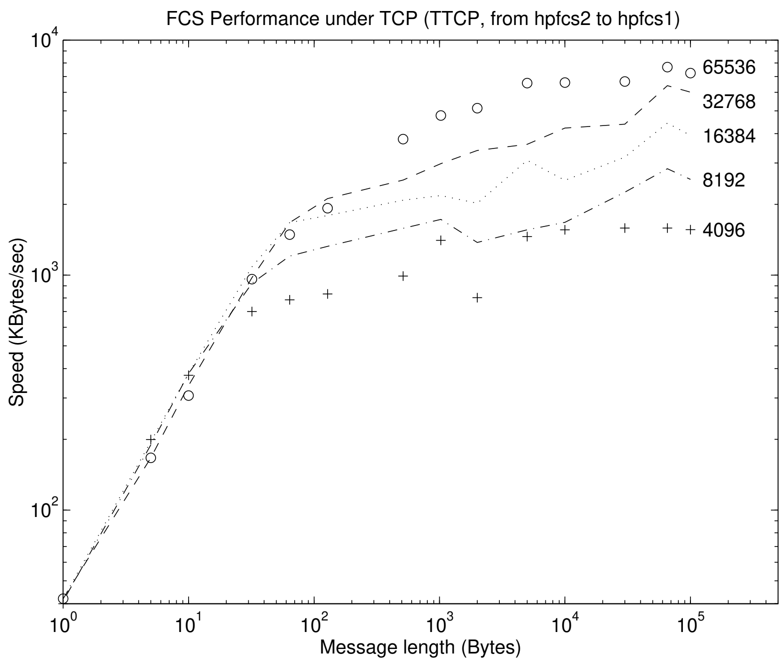


Figure 2a: Speed performance.

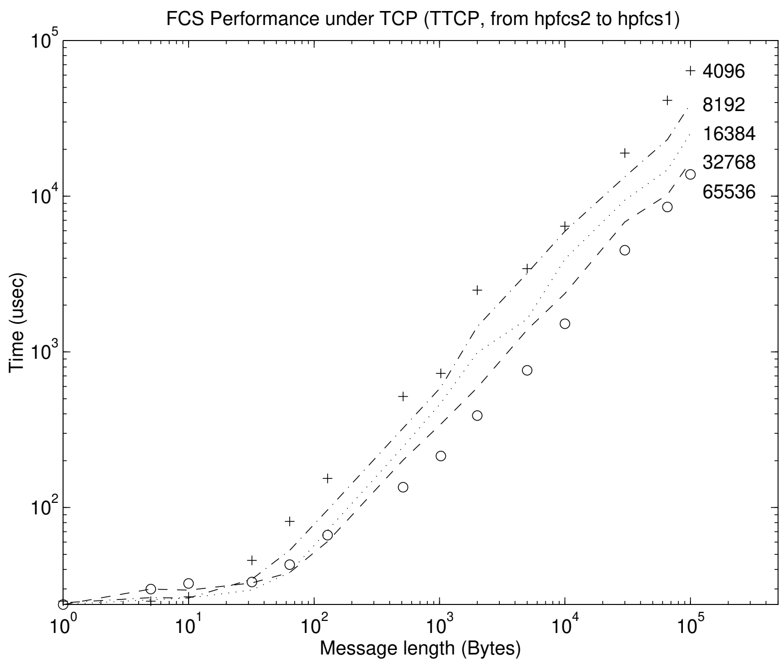


Figure 2b: Latency performance.

### B. Communications between *HP* and *IBM* workstations.

Some communication measurements have been achieved between a *HP/9000-70*0 and an *IBM RS/6000-590* workstations, in order to demonstrate the interoperability. The *IBM RS/6000-590*

has been chosen because it is the most powerful *IBM* machine connected around the Fibre Channel Fabric.
Graphs (see figures *3a*, *3b*, *4a*, *4b*) show how the communication time and the communication speed scale with the size of the message to be transferred and the socket buffer size. Figures *3a* and *3b* (respectively *4a* and *4b*) display results when the *HP* workstation acts as a sender and the *IBM* workstation acts as a receiver, and vice versa respectively.

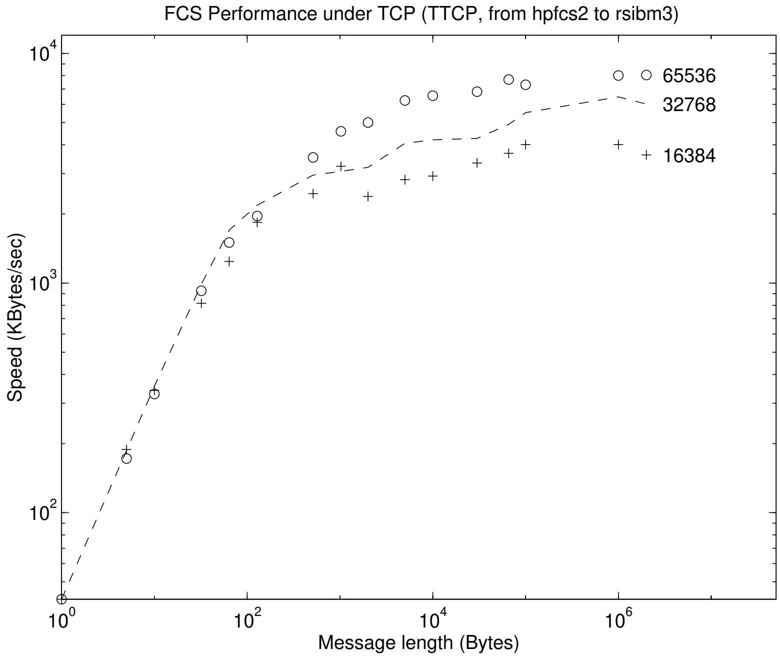


Figure 3a: Speed performance: *HP 9000/700* is sender, *IBM RS/6000 590* is receiver.

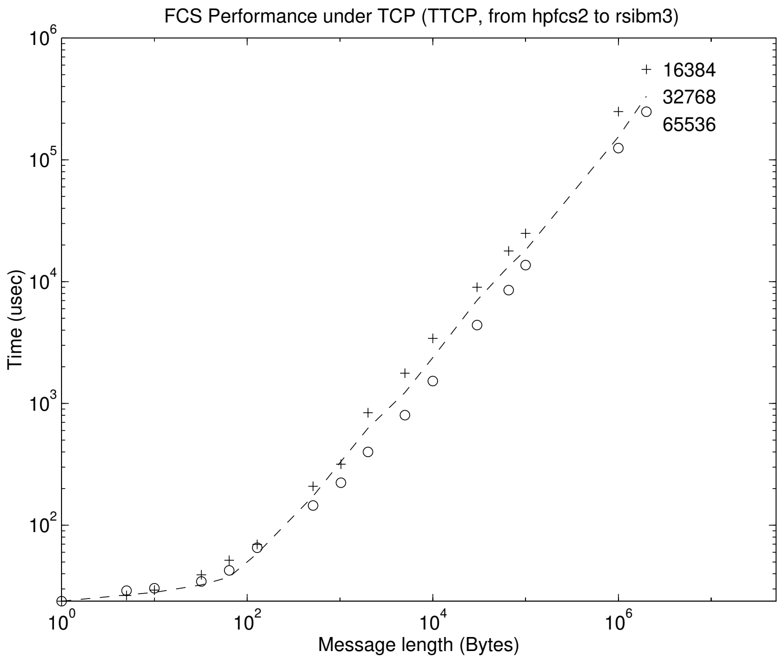


Figure 3b: Latency performance: *HP 9000/700* is sender, *IBM RS/6000 590* is receiver.

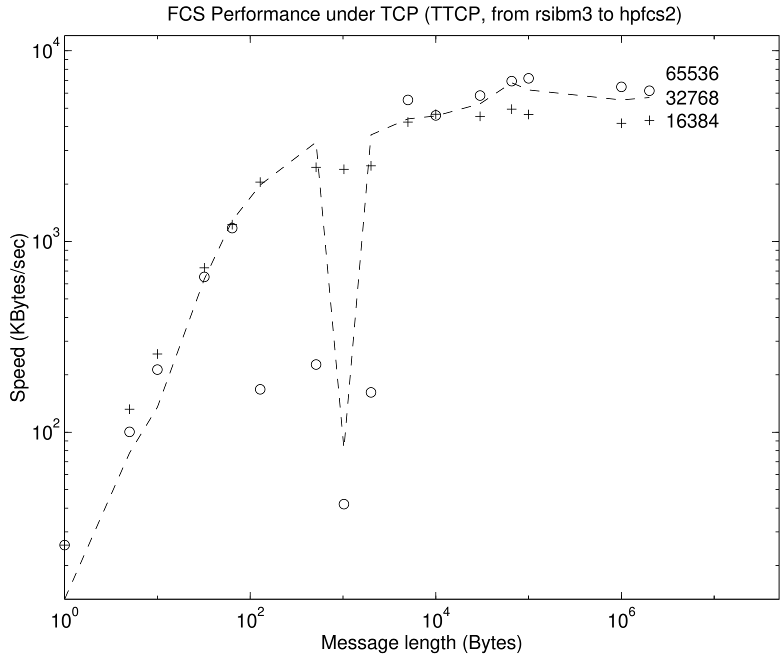


Figure 4a: Speed performance: *HP 9000/700* is receiver, *IBM RS/6000 590* is sender.

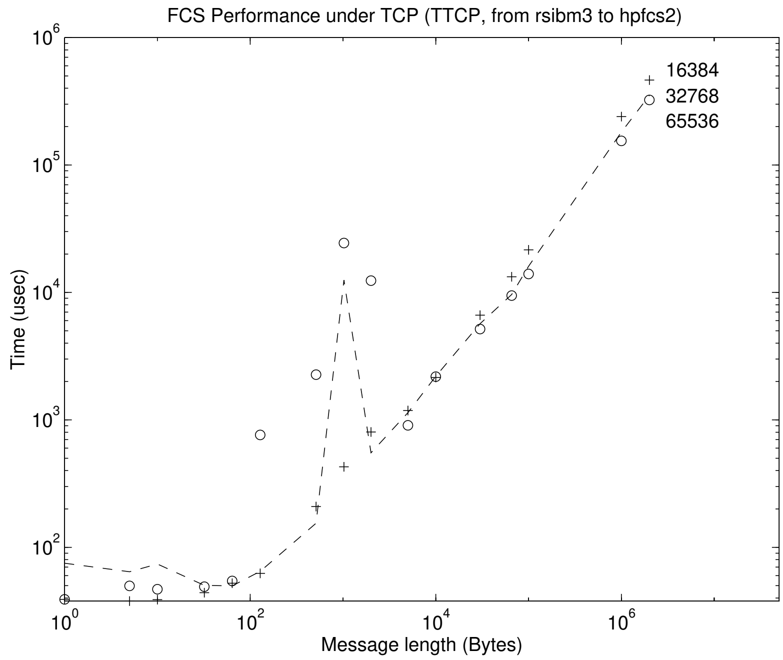


Figure 4b: Latency performance: *HP 9000/700* is receiver, *IBM RS/6000 590* is sender.

# III. Conclusion

All the results presented in this note represent maximum reachable values when implementing a parallel application in which some processing tasks will interleave the communication processes. The best measurements we obtained, are summarized below, according to the hardware configuration used:

- communications between *HP* workstations: the overhead of the system is around 25 microseconds, and the speed can reach 7.7 MBytes/s for transmitting 64 KByte messages (for a 64 Kbyte socket buffer size).

- communications between *RS/6000 590* and *HP/9000 700* workstations:
* the overhead of the system is around 25 microseconds, and the speed can reach 7.2 MBytes/s for transmitting $10^5$ byte messages (for a 64 KByte socket buffer size), when the *IBM* machine is receiving data;
* the overhead of the system is around 25 microseconds, and the speed can reach 8 MBytes/s for transmitting $10^6$ byte messages (for a 64 KByte socket buffer size), when the *IBM* machine is sending data.

For a *IBM-HP* communication direction, the shape of the measured performance chart (either speed or time) is alarming: it is not continuous. For instance, a drop of several hundreds of percent in the performance has been reported between sending a message of 128 bytes and a message of 1024 bytes (for 32KByte and 64KByte socket buffer sizes). Moreover this result is completely deterministic: it can be reproduced at any time. Similar behaviours when measuring communications involving an *IBM* workstation, in a similar experiment using the same *TCP* implementation have been previously reported (see [1] for more details). In that case, for a given parameterization of the socket buffer size and the message size, a deadlock can occur between the two machines, and it is then deblocked by time expiry on the receiver side. We believe this is a likely explanation, which unfortunately can not be verified here, as no network analyser for Fibre Channel is yet available.

The overhead of the system (interface plus fabric plus all the software) is around 25 microseconds. The overhead is defined as the limit for the communication time when the message size to be transferred tends towards zero. For our application, this definition gives an over optimistic expectation of overhead for protocols which wait to merge messages before sending. In that case, for instance instead of sending 10 times a 100 byte message, the system will send 5 times a 200 byte message. This problem has also been already mentioned in [1].

# IV. References

[1] “Communications Benchmarks with the Ancor Fibre Channel Fabric”, Fabrice Chantemargue, CERN/EAST note 94-22, July 28th 1994 (draft).
[2] Fibre Channel Association: “Fibre Channel: Connection to the Future”, ISBN: 1-878707-19-1, 1994
[3] “CXT 250 16-port Switch”, Installer’s/User’s manual, Publication N°. PUB 007 A Rev. D.
[4] “EISA CIM 250”, Driver Interface Manual, Publication N° PUB 010 Rev. B.
[5] “MCA CIM 250”, Installer’s/User’s Manual, Publication N° PUB 009 A Rev. Draft.