Performance Analysis
November 2005

IBM eServer™ X3 Architecture™ Performance: 10 Gigabit Ethernet Adapters

By Dustin Fredrickson
IBM Systems and Technology Group

Executive Overview

The IBM 10 GbE SR Server Adapter[1] is one of the first 10 Gigabit Ethernet adapters available with support for PCI-X 2.0. When used in IBM eServer xSeries® systems based on IBM eServer X3 Architecture, this adapter delivers breakthrough 10 Gigabit Ethernet performance—up to 10 times the performance of today's gigabit Ethernet adapters and significantly better performance than equivalent PCI-X 1.0-based 10 Gigabit Ethernet solutions.

The IBM 10 GbE SR Server Adapter is based on Xframe® II second-generation technology from Neterion. Used in the x260 server, it takes advantage of X3 Architecture and PCI-X 2.0 Double Data Rate (266MHz) bus technology to deliver full line-rate 10 Gigabit Ethernet throughput—a level not possible with PCI-X 1.0-based solutions.

This paper reports the results of measurements conducted to quantify the performance increase for IBM eServer xSeries systems and Xframe II adapters utilizing PCI-X 2.0 bus interfaces as compared to PCI-X 1.0-based 10 Gigabit solutions. Measurements were taken in an IBM testing facility, using back-to-back x260 and x366 systems, and a multi-client configuration (one x260 server communicating with two x366 clients).

The measurement results confirm the increased throughput for the PCI-X 2.0-based adapter, which achieved throughput rates of up to 9.9Gbps (send), 9.2Gbps (receive), and 14.4Gbps (bidirectional), indicating significant performance increases when using the Xframe II adapter in a PCI-X 2.0 slot as compared to a PCI-X 1.0-based 10 Gigabit adapter. For example, the x260 system utilizing the 10 Gigabit Ethernet adapters achieved sustained bidirectional throughputs of approximately 7.7 and 14.4 Gbps with Xframe and Xframe II, respectively, a bidirectional performance improvement of over 88% for the PCI-X 2.0-based solution.

The server configurations and measurement methodologies are documented in this paper, along with the key results achieved in the various transfer scenarios. All measured rates for the Xframe II adapter in a PCI-X 2.0 slot are provided as comparisons to rates achieved with the previous Xframe I version of the adapter utilizing the fastest PCI-X 1.0 bus speed of 133MHz. All comparisons were made using the same test machines and configurations.

[1] Option part number 30R5001.

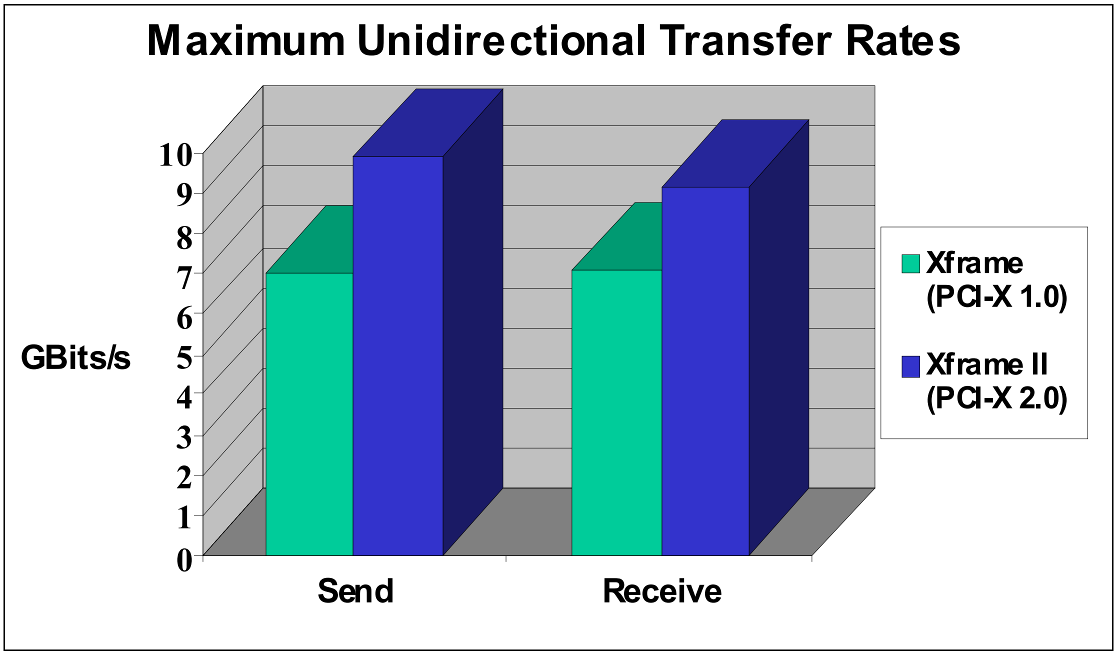
Contents

Performance Results

In all test cases, the throughput and CPU utilization were measured on the x260 server, while performing either sends or receives, or both simultaneously. Although not explicitly shown in the data, similar performance levels were experimentally validated on equivalently configured x260 and x366 systems. Throughput and CPU utilizations reported in all cases represent characteristics of large block application level TCP/IP data transfers, and represent the maximum sustainable capabilities of each configuration.

Unidirectional Communications Performance

The x260 system utilizing the 10 Gigabit Ethernet adapters achieved sustained unidirectional send throughputs of approximately 7.1 and 9.9 Gbps with Xframe and Xframe II, respectively, a unidirectional performance improvement of over 40% for the PCI-X 2.0 based solution.



Server Send Performance

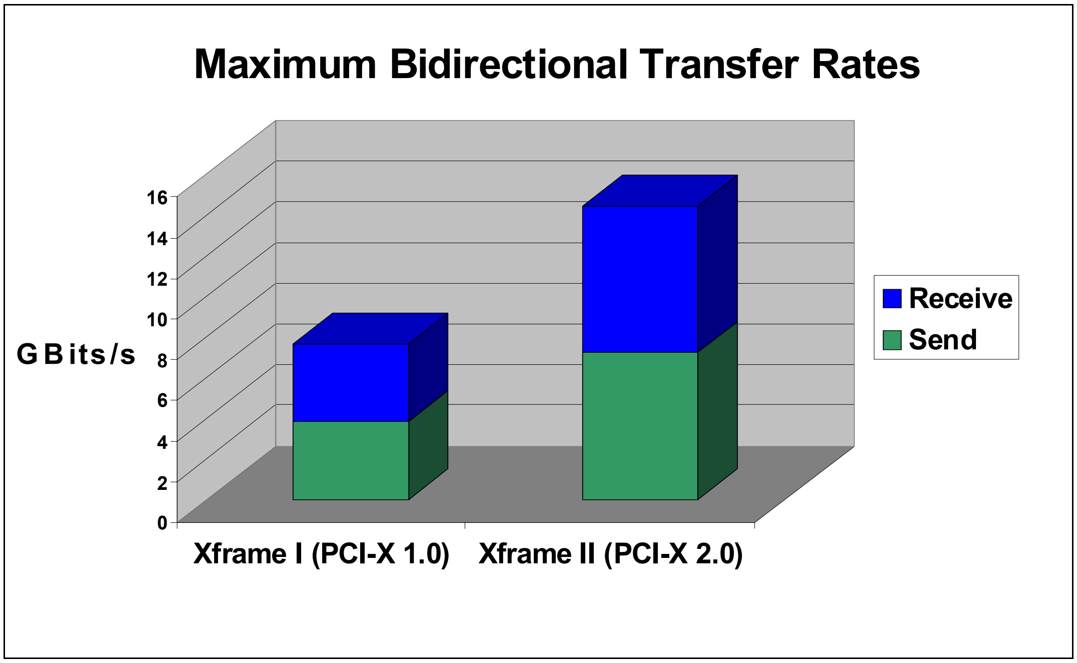
	Xframe PCI-X 1.0	**Xframe** PCI-X 1.0	**Xframe II** PCI-X 2.0 DDR
SEND PERFORMANCE	**1 Server, 1 Client**	**1 Server, 2 Clients**	**1 Server, 2 Clients**
Throughput (Gbits/second)	**7.07**	**6.42**	**9.93**
Average CPU Utilization (%)	**6.50%**	**7.29%**	**9.55%**

Server Receive Performance

	Xframe PCI-X 1.0	**Xframe II** PCI-X 2.0 DDR
RECEIVE PERFORMANCE	**1 Server, 1 Client**	**1 Server, 1 Client**
Throughput (Gbits/second)	**7.06**	**9.17**
Average CPU Utilization (%)	23.66%	27.58%

Bidirectional Communications Performance

The x260 system utilizing the 10 Gigabit Ethernet adapters achieved sustained bidirectional throughputs of approximately 7.7 and 14.4 Gbps with Xframe and Xframe II, respectively, a bidirectional performance improvement of over 88% for the PCI-X 2.0-based solution.



	Xframe PCI-X 1.0	**Xframe II** PCI-X 2.0 DDR
Measured on Server (X260)	**1 Server, 1 Client**	**1 Server, 1 Client**
Send Throughput (Gbits/second)	3.84	7.26
Receive Throughput (Gbits/second)	3.82	7.18
Total Concurrent Throughput (Gbits/second)	**7.66**	**14.44**

Test Methodology

All performance measurements were conducted using the NTttcp benchmark tool, a popular mechanism for evaluating transfer rates over Ethernet connections.

To optimize the transfers, the systems were configured as follows:

- Large Send Offload (LSO): enabled (default) on both systems
- Checksum-Offload: enabled (default)
- Ethernet payload size: 9000 bytes, Jumbo frames (default: 1500 bytes)
- Resulting packet size: 9018 bytes (with Ethernet header and CRC)
- Application Transfer size: 537600 bytes (to minimize TCP/IP stack processing overhead)

System Configurations

Figure 1: Back-to-back setup for sending and receiving. A single 10 Gbit network cable connects the adapters installed on each side. Measurements were taken on the x260 server.

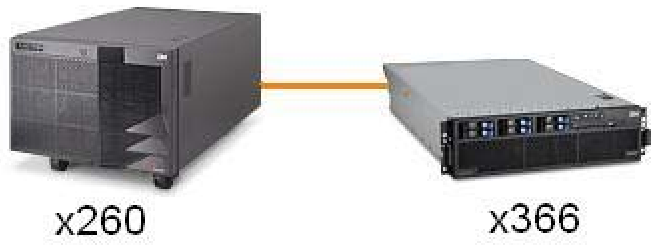
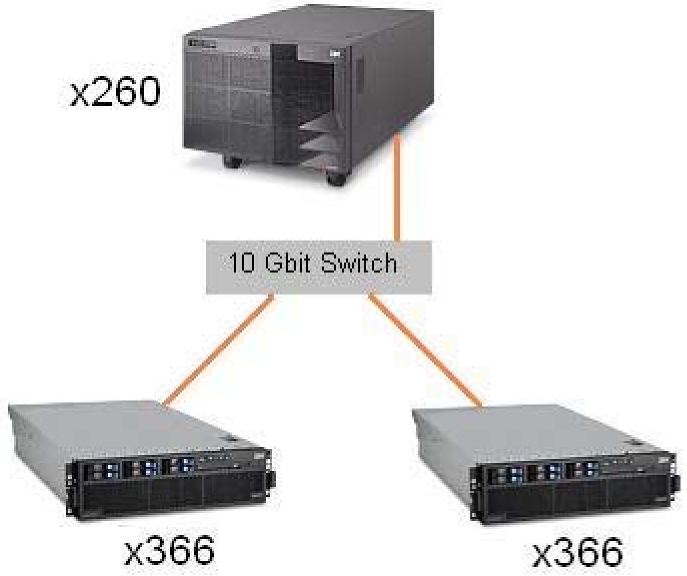


Figure 2: One server to two clients. A 10 Gigabit Ethernet switch is used to connect each system. Measurements were taken on the x260 server.



Server Configuration	
System	x260
CPU	(2) Intel Xeon Processors
Processor Clock Rate	3.66 GHz
Memory	8GB (8x 1GB DIMMs distributed on 4 memory cards) for sends; 32GB (16 x 2GB DIMMs) for receive tests (Note: while only a small amount of memory is used during the tests, the 32GB configuration shows a small performance benefit due to optimizations that benefit high-stress receive mode.)
Local Bus Specification	PCI-X 1.0 (Xframe), PCI-X 2.0 (Xframe II)
Ethernet Adapters	Neterion Xframe and Xframe II 10 GbE Adapters
Network Driver	2.0.10

Network Driver Tuning	Ethernet Payload Size = 9000
	MaxReadByteCount = 4096
	MaxSplitTransaction = 8

Network Type	10 Gigabit Ethernet
Threading	Multi-threaded for sending to two clients, or for simultaneous send/receive
Operating System	Microsoft Windows Server 2003, Enterprise x64 Edition (Build 3790, Service Pack 1)

Client Configuration	
System	x366
CPU	(2) Intel Xeon Processors
Processor Clock Rate	3.66 GHz
Memory	8GB for most configurations, 32GB for back-to-back Server Send (client receive) tests. (Note: While only a small amount of memory was used during the tests, the 32GB configuration showed a small performance benefit due to optimizations that benefit high-stress receive mode.)
Local Bus Specification	PCI-X 1.0 (Xframe), PCI-X 2.0 (Xframe II)
Ethernet Adapters	Neterion Xframe and Xframe II 10 GbE Adapters (matched to server)
Network Driver	2.0.10

Network Driver Tuning	Ethernet Payload Size = 9000
	MaxReadByteCount = 4096
	MaxSplitTransaction = 8

Network Type	10 Gigabit Ethernet
Threading	Single-threaded
Operating System	Microsoft Windows Server 2003, Enterprise x64 Edition (Build 3790, Service Pack 1)

Performance Test Procedure

Setup

- NTttcp installed on send and receive systems.
- Xframe adapters installed on both sides for the first tests; Xframe II adapters installed for subsequent tests.
- All applications closed (nothing running in the background) on both systems.
- Registry entries added at HKLM\System\CurrentControlSet\Services\Tcpip\Parameters:
  - Registry entry Tcp1323Opts, with type as REG_DWORD, and value set to 1.
  - Add a registry entry called TcpWindowSize, also with type REG_DWORD, and value set to 524280 (default 64K).
- Systems rebooted for settings to take effect.

Use of NTttcp

For each transfer measurement, the NTttcp tool initiated a test using Jumbo packets. All receiver threads were started, and then sender threads started simultaneously. For bidirectional tests, a sender and receiver thread is started on both client and server systems simultaneously. The –n parameter was adjusted for a run time of approximately 2 minutes per data point.

The measurements were repeated several times, without interruption, and the best three results were averaged to obtain the numbers reported in this paper.

NTttcp Command Lines

Unidirectional: One Server to Two Clients:	
Server, Thread 1:	ntttcps -m 1,2,192.168.1.2 -a 8 -l 537600 -n 150000 -p 5101
Server, Thread 2:	ntttcps -m 1,2,192.168.1.3 -a 8 -l 537600 -n 150000 -p 5201
Client 1:	ntttcpr -m 1,2,192.168.1.2 -a 8 -l 537600 -rb 5376000 -n 150000 -p 5101
Client 2:	ntttcpr -m 1,2,192.168.1.3 -a 8 -l 537600 -rb 5376000 -n 150000 -p 5201

Bidirectional: One Server to One Client:	
Server, Sender:	ntttcps -m 1,2,192.168.1.2 -a 8 -l 537600 -n 150000 –p 5101
Server, Receiver:	ntttcpr -m 1,2,192.168.1.1 -a 8 -l 537600 -rb 5376000 -n 150000 –p 5201
Client, Receiver:	ntttcpr -m 1,2,192.168.1.2 -a 8 -l 537600 -rb 5376000 -n 150000 –p 5101
Client, Sender:	ntttcps -m 1,2,192.168.1.1 -a 8 -l 537600 -n 150000 –p 5201

Unidirectional: One Server to One Client:	
Sender:	ntttcps -m 1,2,192.168.1.2 -a 8 -l 537600 -n 150000
Receiver:	ntttcpr -m 1,2,192.168.1.2 -a 8 -l 537600 -rb 5376000 -n 150000

NTttcp Command Line Parameters

-m	Number of threads (1 for single stream), on the specified processor (0, 1, 2, or 3 on a quad system), and the local IP address for the channel being exercised.
-a	Number of buffers to be pre-posted (asynchronous I/O buffers)
-l	Length of each buffer in bytes (application I/O size)
-rb	TCP receive window size in bytes
-n	Controls the duration of the test transfer (adjusted to approximate 2 minutes).
-p	Port number (only necessary for multi-threaded tests)

For More Information

To read more about the NTttcp benchmark, visit the Microsoft TechNet Web site at http://technet.microsoft.com/. NTttcp version 2.4 (used in this analysis) or later is available publicly via the Microsoft Windows Driver Kit or the Microsoft Windows Driver DevCon 2005 CD.

© IBM Corporation 2005

IBM Systems and Technology Group
3039 Cornwallis Road
Research Triangle Park, NC 27709

Produced in the USA
11-05
All rights reserved

Visit http://**ibm.com**/pc/safecomputing periodically for the latest information on safe and effective computing. Warranty Information: For a copy of applicable product warranties, write to: Warranty Information, P.O. Box 12195, RTP, NC 27709, Attn: Dept. JDJA/B203. IBM makes no representation or warranty regarding third-party products or services including those designated as ServerProven or ClusterProven.

IBM, the IBM logo, eServer, xSeries and X3 Architecture are trademarks of the International Business Machines Corporation in the United States and/or other countries. For a complete list of IBM Trademarks, see www.ibm.com/legal/copytrade.shtml.

All other products may be trademarks or registered trademarks of their respective companies.

Information about non-IBM products is obtained from the manufacturers of those products or their published announcements. IBM has not tested those products and cannot confirm the performance, compatibility, or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

MBit, GBit, and TBit = 1,000,000, 1,000,000,000 and 1,000,000,000,000 bits, respectively, when referring to network transfer rates.

Performance is based on measurements using industry standard or IBM benchmarks in a controlled environment. The actual throughput that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve performance levels equivalent to those stated here.

IBM reserves the right to change specifications or other product information without notice. References in this publication to IBM products or services do not imply that IBM intends to make them available in all countries in which IBM operates. IBM PROVIDES THIS PUBLICATION “AS IS” WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. Some jurisdictions do not allow disclaimer of express or implied warranties in certain transactions; therefore, this statement may not apply to you.

OPW01607-USEN-00