Performance and Recommended Use of AD385A 10 Gigabit Ethernet SR Cards

From Results on an HP rx8640 Server

February 2008

Table of Contents

Introduction

The availability of next generation 10 Gigabit devices makes deployment of 10 Gigabit Ethernet even more attractive. 10 Gigabit is ideal for data center deployment and can provide up to ten times the performance of 1 Gigabit Ethernet at approximately four times the cost. 10 Gigabit Ethernet can free up I/O slots and enable server consolidation in deployments where I/O slot availability is a limiting factor. 10 Gigabit also provides increased performance and flexibility for server virtualization. Please refer to the document *10 Gigabit Ethernet in Servers: Benefits and Challenges* available in the Information library under http://www.hp.com for more information on 10 Gigabit technology and deployment.

AD385A (Figure 1) is HP's next generation 10 Gigabit Ethernet adapter supporting PCI-X 2.0a Mode 2 (266MHz) operation. It carries forward all the salient features of the previous 10 Gigabit adapter (AB287A), including ECC protection for all internal memory and storage. The "Features and Benefits of AD385A" section in this document provides more details.

This document highlights the excellent performance of the AD385A. The performance measurements were done on an HP Integrity rx8640 server.

This document also provides suggestions based on the performance and technology that will help you to optimize the use of these products.

Figure 1 AD385A PCI-X 10 Gigabit Ethernet Card

Performance Summary

The AD385A card provides exceptional performance when used in accordance with the recommendations in this paper. Tests with a Maximum Transmission Unit (MTU) of 9000 bytes and 1500 bytes show that the AD385A provides link-rate throughput for both transmit and receive tests when the recommended models are used.

During scalability tests, the AD385A exhibited excellent linear scaling in performance when additional adapters were added to the system. This demonstrates convincingly the capabilities of the rx8640 server, HP-UX 11iv3 operating system and the AD385A adapter.

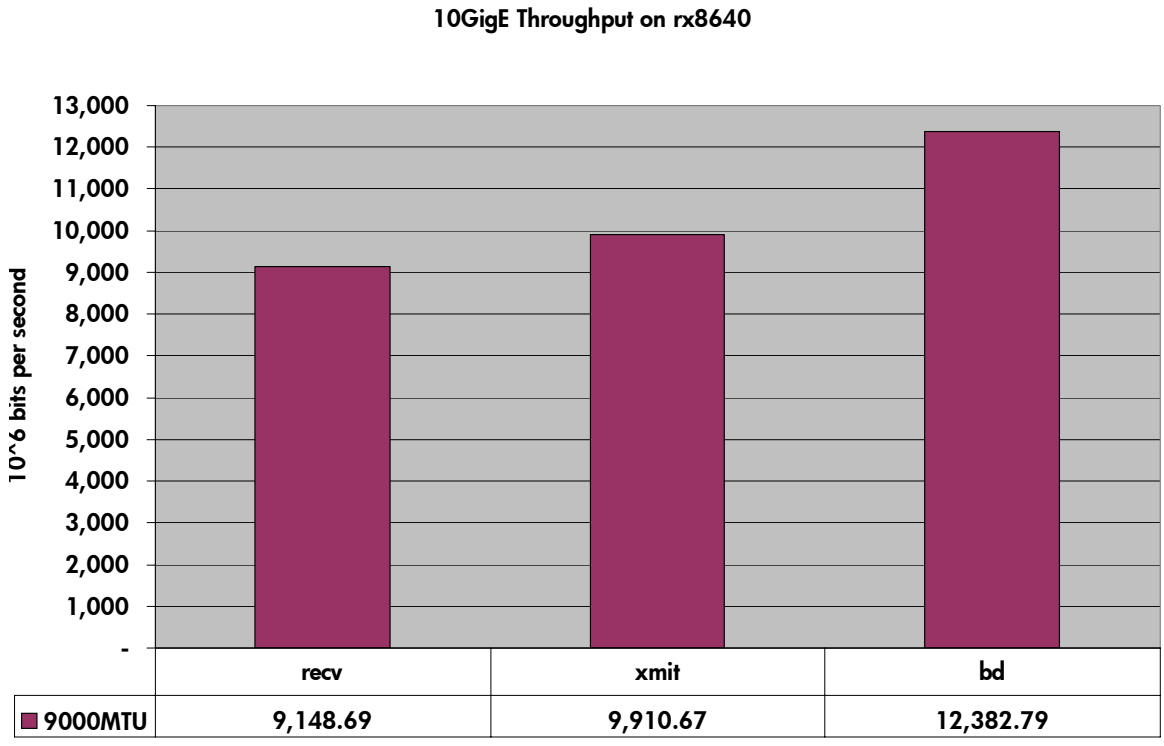
AD385A Throughput Performance

Following are the highlights of the excellent performance achieved when operating at 9000 MTU:

- Receive traffic achieved a sustained throughput of 9.15 Gigabits/second, which is close to the link rate of a 10 Gigabit Ethernet link. The service demand (which is the amount of time it takes the CPU to handle one kilobyte of data) was 1.21µs/KByte.
- The transmit throughput was 9.90 Gigabits/second, which is the link rate of a 10 Gigabit Ethernet link. The service demand was 0.86µs/KByte.
- The bi-directional throughput was 12.4 Gigabits/second. The bi-directional throughput is limited by the bandwidth capabilities of the PCI-X 266MHz implementation in the rx8640 server and the maximum bandwidth of the PCI-X 266MHz bus. The service demand was 1.07µs/KByte.

Figure 2 shows the throughput achieved by the10 Gigabit adapter operating at 9000 MTU. Table 1 shows the corresponding service demand.

Figure 2 10 Gigabit Ethernet Throughput at 9000 MTU



Following are the highlights of the impressive performance achieved during traffic testing at 1500 MTU (as shown in Figure 3):

- Receive traffic achieved 9.11 Gigabits/second throughput with a service demand of 2.20µs/KByte.
- A throughput of 9.47 Gigabits/second was achieved with Transmit traffic, which is the link rate at 1500 MTU for a 10 Gigabit Ethernet link; the service demand was 1.35µs/KByte.
- Bi-directional traffic achieved 12.02 Gigabits/second throughput with a service demand of 1.74µs/KByte. The bi-directional throughput is limited by the bandwidth capabilities of the PCI-X 266MHz implementation in the rx8640 server and the PCI-X 266MHz bandwidth limits.

Figure 3 10 Gigabit Ethernet Throughput at 1500 MTU

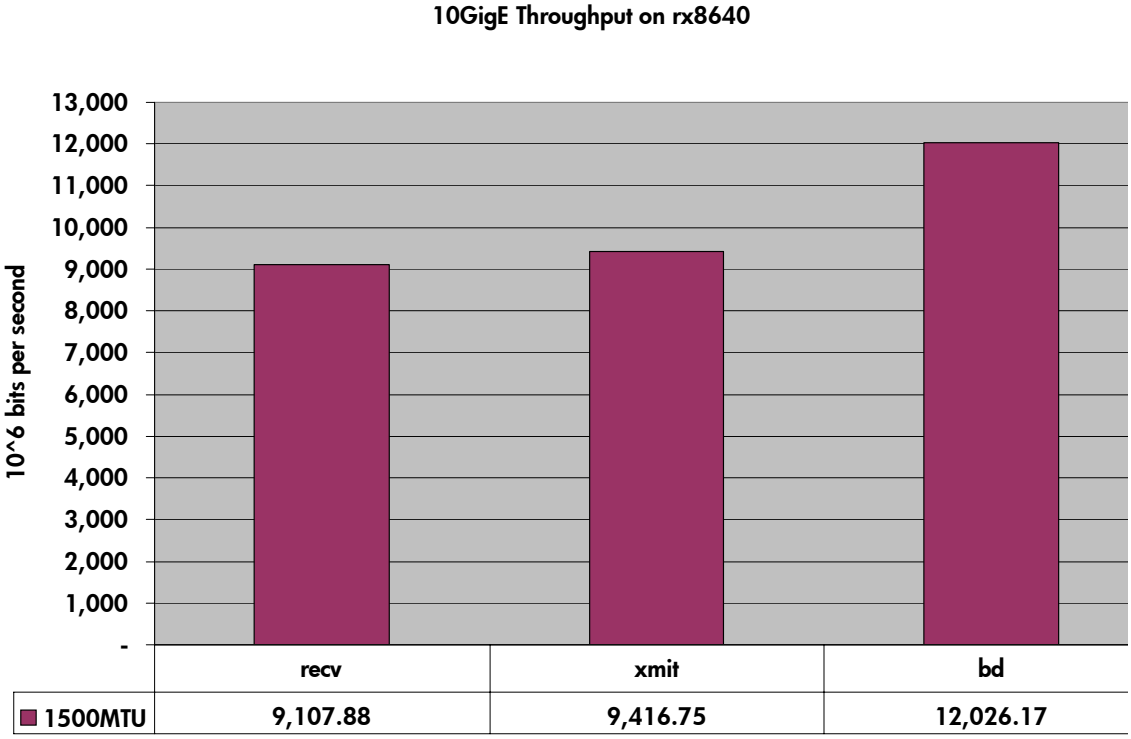


Table 1 shows the service demand.

Table 1 10 Gigabit Ethernet Service Demand

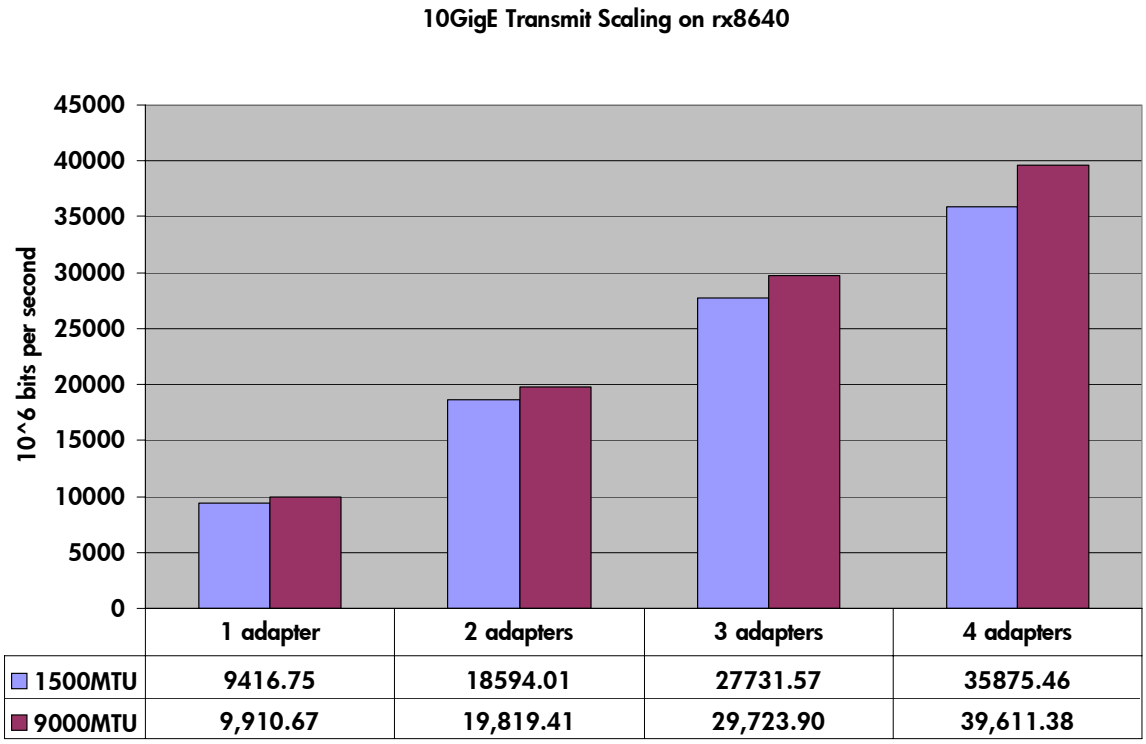
	Service Demand (µs of CPU time consumed/KB)	
	1500 MTU	**9000 MTU**
Transmit	1.35	0.86
Receive	2.20	1.21
Bi-directional	1.74	1.07

Scalability Tests

The AD385A exhibited linear scaling in performance when additional adapters were added to the system. During scalability tests, adapters were installed on slots 3 and 4 on both I/O bays of the rx8640.

The aggregate transmit throughput was 39.6 Gigabits/sec with four adapters installed on the server, when operating at 9000 MTU. The transmit performance with one, two, three and four AD385A adapters is shown in Figure 4. The performance scales linearly with the number of adapters.

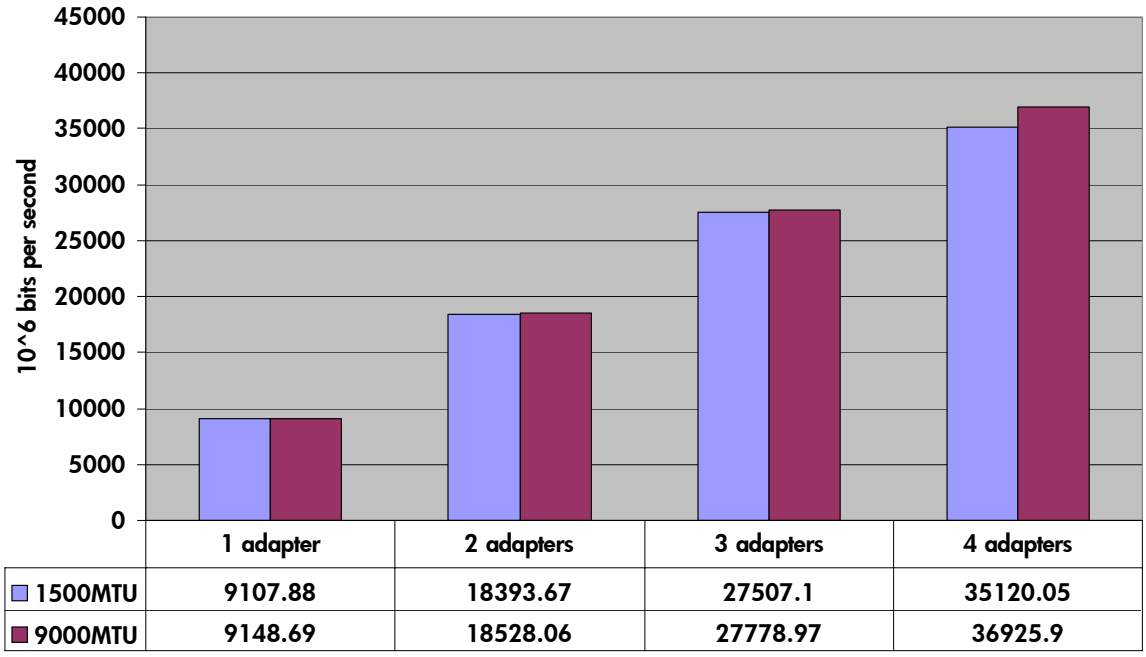
Figure 4 10 Gigabit Ethernet Transmit Scalability Tests



	1 adapter	2 adapters	3 adapters	4 adapters
1500MTU	9416.75	18594.01	27731.57	35875.46
9000MTU	9,910.67	19,819.41	29,723.90	39,611.38

An aggregate throughput of 37 Gigabits/second was achieved for receive traffic with four adapters when operating at 9000 MTU. The receive performance with one, two, three and four AD385A adapters is shown in Figure 5.

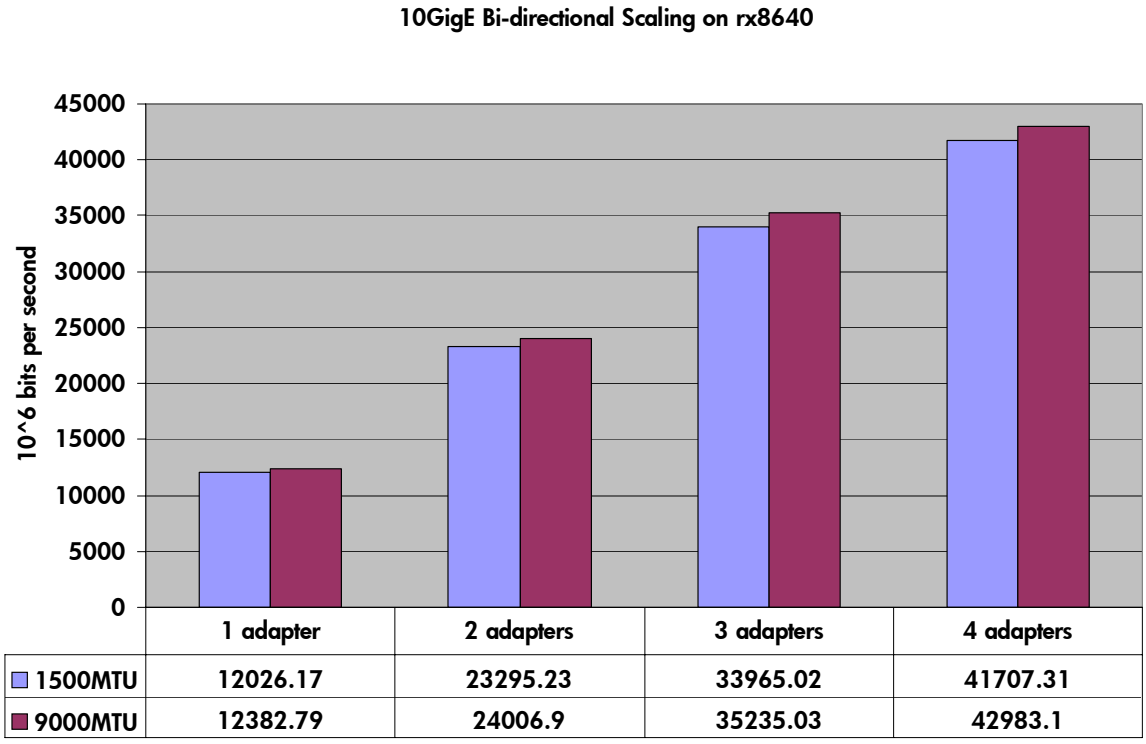
Figure 5 10 Gigabit Ethernet Receive Scalability Tests



	1 adapter	2 adapters	3 adapters	4 adapters
1500MTU	9107.88	18393.67	27507.1	35120.05
9000MTU	9148.69	18528.06	27778.97	36925.9

The aggregate bi-directional throughput was about 43 Gigabits/second with four adapters when operating at 9000 MTU. The bi-directional throughput with one, two, three, and four AD385A adapters is shown Figure 6.

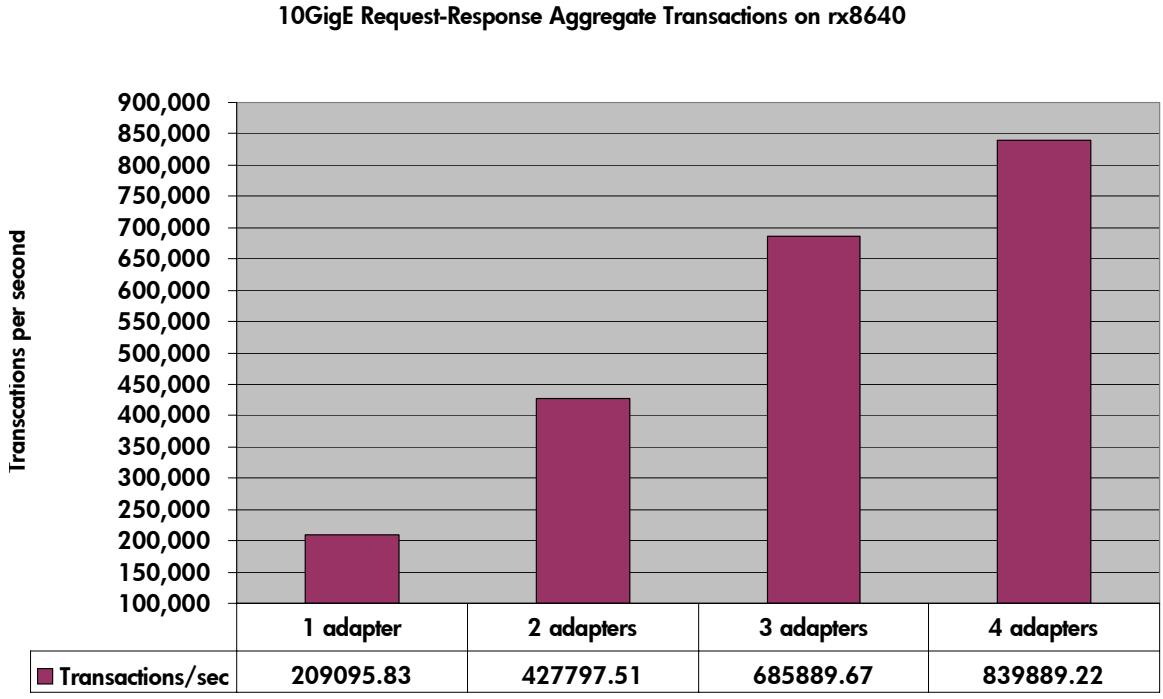
Figure 6 10 Gigabit Ethernet Bi-directional Scalability Tests



	1 adapter	2 adapters	3 adapters	4 adapters
1500MTU	12026.17	23295.23	33965.02	41707.31
9000MTU	12382.79	24006.9	35235.03	42983.1

Request-Response tests were also run to demonstrate the capability of the adapter to handle traffic of this nature. Again, tests were run with one, two, three and four AD385A adapters. The aggregate transactions that could be sustained with the processors running close to saturation were captured. These are shown in Figure 7.

Figure 7 10 Gigabit Ethernet Request-Response Scalability Tests



Recommended Use Based on Performance and Design

HP recommends the following usage model to achieve the best performance:

- Run the AD385A cards in the highest performing PCI-X slots. Slots 3, 4, 5, and 6 on each of the I/O bays are the recommended high performance PCI-X slots in the HP Integrity **rx8640** used in our performance testing. Please refer to the appropriate Server Installation Guide for information on how to identify high performance slots.
- Configure Jumbo Frames (9000 MTU) if possible. Jumbo Frames can provide excellent throughput with reduced CPU overhead. However, Jumbo Frames may not be usable in all topologies. Please refer to the HP document "Boosting server-to-server Gigabit Throughput with Jumbo Frames" for more information on advantages and recommended usage models.
- The10 Gigabit driver is optimized for multiprocessor support. Best performance will be achieved when running multiple data streams (TCP connections, for example) on a system with four or more processors configured. The software will automatically assign appropriate processor resources to each data stream. This is the default configuration in HP-UX 11i v3.
- The10 Gigabit driver also supports the ability to look at received packets, identify those belonging to the same TCP connection, perform reassembly and send the reassembled segments to the upper layers. This helps to reduce CPU utilization. This feature is called TCP packet reassembly in the driver. The performance gain that will be seen with this feature is workload dependent and is limited to TCP applications. Applications such as backup would see good improvements, but other workloads may not see the same level of improvement. It is disabled by default and can be enabled using nw_mgr. Please see the 10 Gigabit driver release notes and the nw_mgr man page on how to enable this feature and the restrictions to bear in mind when enabling this feature.

The recommendations mentioned here were followed to achieve the results shown in this article. Card throughput will be affected with configurations other than those recommended.

Please contact an HP representative for additional help in understanding how to best deploy the AD385A card.

How We Measured 10 Gigabit Ethernet Efficiency

This article highlights the AD385A **throughput**. Throughput is the **data transfer rate**, or the quantity of data transferred from one system to another in a given amount of time. In this article, it's shown for one-way transfers as well as two-way. Throughput measures how well programs run with a certain workload and how quickly user requests can be handled.

This article also provides the Service Demand for each throughput test. Service demand is the amount of time (in microseconds) it takes one CPU to handle one kilobyte of data. It is a normalized measurement because it eliminates disparities due to differences in quantities, types, or frequencies of CPUs. Service Demand is an important capacity planning & performance metric that is sometimes considered when comparing different server models.

The performance results shown in this article were measured with the netperf benchmarking software.

Tests were run with AD385A adapters installed in a 2-cell rx8640 with sixteen processors (four dual-core processors in each cell). Adapters were installed in slots 3 and 4 on both I/O bays.

The driver supports multi-queue using destination port based steering. It programs a card resident table to look-up the destination port of the received packets and directs them to appropriate queues. The driver updates the table dynamically as operating conditions change. This feature provides good performance improvement when there is a mix of applications that do not use the same destination port.

Details of the systems used and the software versions are shown in Table 2. The message and socket sizes used with netperf for the 10 Gigabit Ethernet transmit, receive, and bidirectional tests are as follows. 1500 MTU tests were run using a message size of 64Kbytes and socket size of 256K. At 9000 MTU, the tests used a message size of 128K and socket size of 512K. TCP RR tests were run using a request size of 1-byte and a response size of 1-byte.

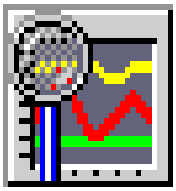
Performance will vary when this product is used on different systems or software.

NOTES:

- The observed performance results are consistent across all of the same type of I/O slots of the system; On HP Integrity rx8640, the high performance slots are slots 3, 4, 5 and 6 on each of the I/O Partitions. Performance on other slots will be lower.
- The core I/O card in the rx8640 carried minimal site LAN traffic during performance tests.
- Processor affinity was used to assign netperf processes to specific CPUs.

Table 2 summarizes the products used to measure the performance:

Table 2 Products Used in the Performance Measurement Tests

Servers Tested		rx8640 Server. 2-cell configuration. Eight 1.6 GHz CPU Itanium2 dual-core processors with 18 MB cache Four sockets per cell System Memory: 32GB. Eight DIMM slots in each of the cells were populated. Operating System - HP-UX 11i v3 of Sep 2007 (B.11.31.0709).
Cards Tested		AD385A PCI-X 266MHz 10 Gigabit Ethernet card • PCI-X (64-bit, 266MHz, 3.3v) • 10GigE LAN Driver (ixgbe) version: 10GigEthr-00 B.11.31.0712
Clients generating the test load for 10 Gigabit Ethernet	 	Four rx2660 servers, each with: • Two, 1.6 GHz Itanium2 CPUs • HP-UX 11i v2 (B.11.23.0712) • AD385A PCI-X 10 Gigabit card • 10 GigE LAN Driver (ixgbe) version B.11.23.0712 Two rx2600 servers, each with • Two, 1.5GHz Itanium2 CPUs • HP-UX 11i v2 (B.11.23.0712) • AD385A PCI-X 10 Gigabit card • 10 GigE LAN Driver (ixgbe) version B.11.23.0712
Benchmark software for 10 Gigabit Ethernet tests		Netperf 2.4.3 is the benchmarking software suite that generated LAN and disk traffic for the 10 Gigabit Ethernet performance tests. For more information about netperf or to get a free copy, go to http://www.netperf.org

Features and Benefits of the AD385A

Features and benefits of the AD385A include:

- PCI-X operation in 266MHz, 64-bit mode.
- Conforms to IEEE 10GBASE-SR using multi-mode fiber. Operating distances from 7 to 984 feet (2 to 300 meters).
- Supports several features that improve CPU utilization:
  - Jumbo Frames with a maximum transmission unit (MTU) of 9000 bytes.
  - On board TCP Segmentation Offload (TSO) of IPv4.
  - On board Checksum Offload (CKO) for TCP, UDP and IPv4.
- All onboard memory and storage are ECC protected; with single bit errors automatically corrected and double bit errors detected.
- Supports HP Serviceguard and Auto-Port Aggregation (APA) for high availability. APA is currently supported only in Lan Monitor mode for high availability.
- Supports PCI-X online addition/replacement (OLA/R) on specified systems.
- Supports 64-bit management information base (MIB) statistics. Directs the data link service (DLS) provider to return 64 bit statistics.
- Supports subnetwork-access protocol (SNAP) network-layer protocol encapsulation.

For More Information

For more information about the products described in this paper such as a current list of tested HP products or supported systems, please go to:

http://www.hp.com/products1/unixserverconnectivity.

This paper is the latest in a series of white papers detailing the performance of HP's link and server products. For a complete list of white papers on HP's networking and I/O products including Gigabit Ethernet and Fibre Channel solutions, go to :http://docs.hp.com/en/netcom.html.

For further assistance including a detailed analysis of your specific requirements and needs, please contact your local HP Sales Representative.

For More Information

The only warranties for HP products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. HP shall not be liable for technical or editorial errors or omissions contained herein.

HP-UX® , Serviceguard®, and Superdome® are registered trademarks of the Hewlett-Packard Corporation. Itanium 2 is a trademark of Intel Corporation. PCI-X is a registered trademark of the PCI SIG.

All other trademarks and registered trademarks are the property of the respective corporations.