

**富士通社製 PC サーバ(PRIMERGY)と、
Exablaze 社製 超低遅延 10Gb イーサ
ネット・アダプター(ExaNIC)との
接続検証報告書**

2015年7月31日
ビットリーブ株式会社

目次

1.	はじめに	2
2.	検証の目的	3
3.	検証の場所と期間	3
4.	検証環境	3
(ア)	構成概要	3
(イ)	サーバ仕様	4
5.	検証内容	5
6.	検証結果	6
(ア)	装着確認	6
(イ)	動作確認	6
(ウ)	遅延確認	7
(エ)	スループット確認	9
7.	まとめ	10
8.	参考情報、お問い合わせ先	10

1. はじめに

Exablaze 社製の超低遅延 10GB イーサネット・アダプター「ExaNIC」は、アプリケーションからネットワークそして再びアプリケーションまでの遅延が 64 バイトのパケットで約 1 マイクロ秒と世界で最も低遅延の製品です。

これまで、この製品は高頻度取引などの遅延に厳しい金融業のアプリケーションで主として利用されてきました。超低遅延の性質は金融業に限らず様々なシステムやアプリケーションの性能を向上することができます。たとえば、サーバ間のデータ転送速度を向上させたいデータセンターや、大量のパケットを扱うがゆえに1つ1つのパケットの処理速度を最適化したいメディア業やストレージ業、およびテレコム業などのアプリケーションが想定されます。

今回の検証では、様々な業種で標準的に利用され、数多くの高速な計算処理システム構築の実績がある富士通社のサーバに「ExaNIC」を搭載したときに、正常に動作し、かつ超低遅延の性能となることを確認します。

結果としては、イーサネット・アダプターとして問題なく動作し、アプリケーションからネットワークそして再びアプリケーションまでの遅延は 64 バイトのデータで約 1 マイクロ秒、1,518 バイトで約 3 マイクロ秒となることが確認できました。また、スループット測定においても TCP/IP を介した実効データ転送性能が約 7GB/秒と良好な結果となりました。

本報告が高速なシステム構築をお考えのお客様のご検討の一助になれば幸甚です。

表1 接続検証機器の仕様

型番	仕様
ExaNIC X2 	ハーフハイト PCI Express カード データレート : 10GbE/1GbE ポート : 2 SFP+, PPS および電源 ホスト I/F : PCIe × 8 Gen 2.0@5.0 GT/s 時刻同期 : ホスト、PTP、オプションの PPS PPS 入力 : RS-422、RS-232、TTL オペレーティングシステム: Linux IC 設計 : FPGA
ExaNIC X4 	フルハイト PCI Express カード データレート : 10GbE/1GbE ポート : 4 SFP+, PPS および電源 ホスト I/F : PCIe × 8 Gen 2.0@5.0 GT/s 時刻同期 : ホスト、PTP、オプションの PPS PPS 入力 : RS-422、RS-232、TTL オペレーティングシステム: Linux IC 設計 : FPGA

※サーバは「PRIMERGY RX2540 M1(以下、RX2540 M1)」を用いて検証しました。

※本検証ではライザーカード無しに RX2540 M1 に装着できる、ハーフハイトタイプの「ExaNIC X2」を用いました。「ExaNIC X2」と「ExaNIC X4」の性能は同等です。

※富士通純正品の PCIe(× 8) フルハイトライザーカード PY-PRE821、または PY-PRE822 が適用できる構成の RX2540 M1 には、フルハイトタイプの「ExaNIC X4」カードが装着可能です。

2. 検証の目的

富士通社のサーバに「ExaNIC」が装着できること、問題なくイーサネット・アダプターとして利用できること、および超低遅延の性能となることを確認する。

3. 検証の場所と期間

検証場所 : 富士通検証センター(東京・浜松町)

検証期間 : 2015年7月22日(水曜)~2015年7月24日(金曜)

4. 検証環境

(ア) 構成概要

富士通社サーバ「RX2540 M1」にハーフハイトの「ExaNIC X2」を装着する。その構成のサーバを2台配置して、その間を1メートルの光ファイバーケーブルで直結接続して検証する。図1に接続検証時の構成を、図2に実際の接続イメージを示す。



図1 接続検証の構成



図2 実際の接続イメージ

(イ) サーバ仕様

サーバ1、サーバ2の機器仕様、ネットワーク仕様、オペレーティングシステム仕様は同一のものを使用している。以下の表にサーバ機器仕様、ネットワーク仕様、オペレーティングシステム仕様、接続仕様を示す。

表2 サーバ機器仕様

製品名	FUJITSU PRIMERGY RX2540 M1
ラックユニット	2U Server
物理プロセッサコア総数	16
論理プロセッサコア総数	16
ハイパースレディング	無効
プロセッサ種別	Intel(R) Xeon(R) CPU E5-2630 v3 @ 2.40GHz
キャッシュ	20MB
Intel QPI	N/A
メモリーインターコネクト	PCIe Gen3
メモリー	32GB
ディスク	600GB SATA
BIOS	Vendor: FUJITSU // American Megatrends Inc. Version: V5.0.0.9 R1.22.0 for D3289-A1x Release Date: 04/20/2015

表3 ネットワーク仕様

NIC/HCA1	Exablaze ExaNIC X2 Adapter
使用方法	Message Traffic
ドライババージョン	Exanic 1.5.0
ファームウェアバージョン	2015年03月30日
ネットワークスタック設定	MTU=1500
カーネルバイパスバージョン	Exasock 1.5.0

表4 オペレーティングシステム仕様

バージョン	Red Hat Enterprise Linux 6.6(for Intel64) (2.6.32-504.el6.x86_64)
カーネルコマンドライン記述	root=UUID=a6cec699-0696-4b54-a87e-12d298cecc79 rd_NO_LUKS rd_NO_MD crashkernel=auto KEYBOARDTYPE=pc KEYTABLE=jp106 LANG=ja_JP.UTF-8 rd_NO_LVM rd_NO_DM rhgb quiet isolcpus=1,2,3,4,5,6,7,8
動作サービス	NetworkManager ,abrt-dump-oops ,abrt-d ,capi ,cgred ,dhcpcd ,dhcrelay ,dnsmasq ,dovecot ,BE2SNMP ,gatherd ,reposd ,htcacheclean ,httpd ,csid ,kpropd ,krb5kdc ,lvmetad ,m dmonitor ,multipathd ,mysqld ,named ,rpc.svcgssd ,rpc.mountd ,nfsd ,rp c.rquotad ,nmbd ,nslcd ,ntpd ,numad ,odjjobd ,postmaster ,qpidd ,quota _nld ,radvd ,rdc ,rngd ,rpc.gssd ,rpc.idmapd ,sandbox ,sasauthd ,smartd ,smbd ,snmpd ,snmptrapd ,spamd ,spice-vdagentd ,sssd ,svnserve ,tcsd ,tgtd ,vsftpd ,wdaemon ,winbindd ,wpa_su plicant ,ypbind.
ファイルシステム	Ext3
シールドディング、バインディング	N/A
OS 設定追記	N/A

表5 接続仕様

スイッチ	1メートル 光ファイバーケーブルを直接接続
------	-----------------------

5. 検証内容

以下の表 6 の手順に従い検証を行う。検証で用いるツールの情報を表 7 に示す。

表 6 検証手順

項番	検証内容
1	[装着確認] ・富士通サーバ「RX2540 M1」に「ExaNIC」を装着して、ぶつかりや緩みがないことを確認する。
2	[動作確認] ・「RX2540 M1」に OS (Red Hat Enterprise Linux) をインストールした状態から、「ExaNIC」のドライバソフトウェアをインストールして、カードを認識するか、ポートを認識するか、ネットワークに接続するか、正常に通信ができるかを確認する。
3	[遅延確認] ・サーバ 1 の「ExaNIC」からサーバ 2 の「ExaNIC」を 10GBE のファイバーケーブルで直接接続して、SockPerf ツールにより遅延を測定する。 ・サーバ 2 には SockPerf ツールをサーバモードで動作させ、サーバ 1 から送信バイト長毎に 10 万回の測定を行う。 ・送信バイト長は RFC2544 に従い 64 バイト、128 バイト、256 バイト、512 バイト、768 バイト、1024 バイト、1280 バイト、1518 バイトの範囲で TCP と UDP で計測する。
4	[スループット確認] ・サーバ 1 の「ExaNIC」からサーバ 2 の「ExaNIC」を 10GBE のファイバーケーブルで直接接続して、Iperf ツールによりスループットを測定する。 ・10 秒、60 秒、100 秒、3600 秒間パケットを送信して、1 秒間の処理ビット数を計測する。

表 7 検証ツール

項番	検証内容
1	[遅延確認 SockPerf] ・バージョン 2.5 ・コマンド
2	[スループット確認 Iperf] ・バージョン 3.0.3 ・コマンド

6. 検証結果

(ア) 装着確認

「RX2540 M1」にハーフハイトの「ExaNIC X2」を装着して、ぶつかりや緩みがなく問題無いことを確認した。

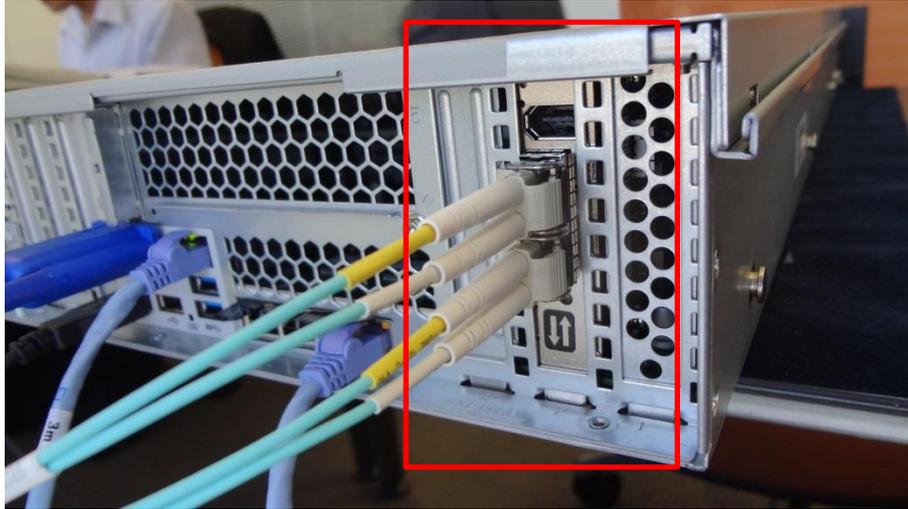


図3 「ExaNIC X2」の装着状態

※RX2540 M1 PCI スロットの Low Profile カード搭載部

(イ) 動作確認

OS (Red Hat Enterprise Linux 6.6(for Intel64))をインストールした[RX2540 M1]に[ExaNIC]のドライバソフトウェアをインストールして、リンクアップすることを確認した。[RX2540 M1]に標準で装着された4ポートのイーサネット・アダプターがあるため、[ExaNIC]は eth4 と eth5 とし設定を行っている。

正常な通信は遅延確認とスループット確認にて TCP と UDP で確認を行う。

```
[root@rx2540ml-21 TestResult]# exanic-config
Device exanic0:
Hardware type: ExaNIC X2
Board ID: 0x02
Temperature: 66.1 C  VCCint: 1.01 V  VCCaux: 1.83 V
Fan speed: 7347 RPM
Function: network interface
Firmware date: 20150327 (Fri Mar 27 06:34:05 2015)
Bridging: off
Port 0:
Interface: eth4
Port speed: 10000 Mbps
Port status: enabled, SFP present, signal detected, link active
MAC filters: 64 IP filters: 128
Promiscuous mode: off
Bypass-only mode: off
MAC address: 64:3f:5f:01:23:30
IP address: 192.168.10.1 Mask: 255.255.255.0
RX packets: 62 ignored: 0 error: 0
TX packets: 20
Port 1:
Interface: eth5
Port speed: 10000 Mbps
Port status: enabled, SFP present, signal detected, link active
MAC filters: 64 IP filters: 128
Promiscuous mode: off
Bypass-only mode: off
MAC address: 64:3f:5f:01:23:31
IP address: 192.168.11.1 Mask: 255.255.255.0
RX packets: 122715986 ignored: 0 error: 0
TX packets: 1656978074
[root@rx2540ml-21 TestResult]#
```

図4 リンクアップの確認状態

(ウ) 遅延確認

sockperf コマンドでアプリケーションからネットワークそして再びアプリケーションまでの遅延について、64 バイト、128 バイト、256 バイト、512 バイト、768 バイト、1024 バイト、1280 バイト、1518 バイトの範囲で TCP と UDP で計測した。メーカーの公表値を図 5 に示す。本計測では、この計測値との乖離も確認する。

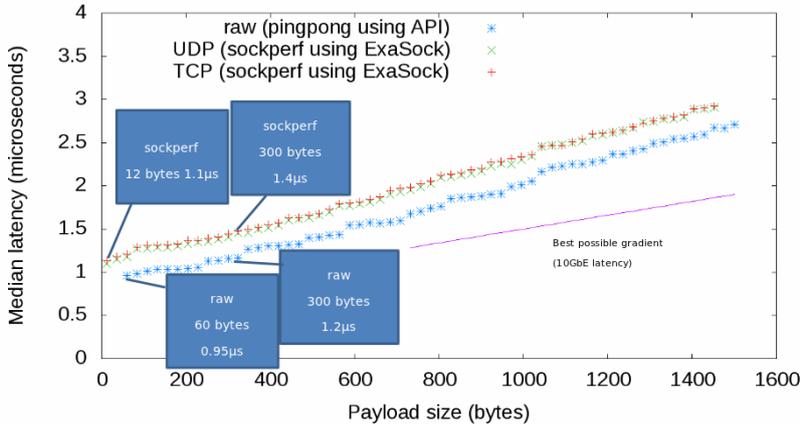


図 5 遅延に関するメーカーの公表値

図 6 に TCP の測定結果、図 7 に UDP の測定結果を示す。TCP の結果としては、64 バイトで 1.31 マイクロ秒、1518 バイトでは 3.06 マイクロ秒となり、図 5 に示すメーカーの公表値ともほぼ同等となった。UDP は 64 バイトで 1.17 マイクロ秒、1518 バイトでは 2.91 マイクロ秒となり、こちらも図 5 に示すメーカーの公表値ともほぼ同等となった。

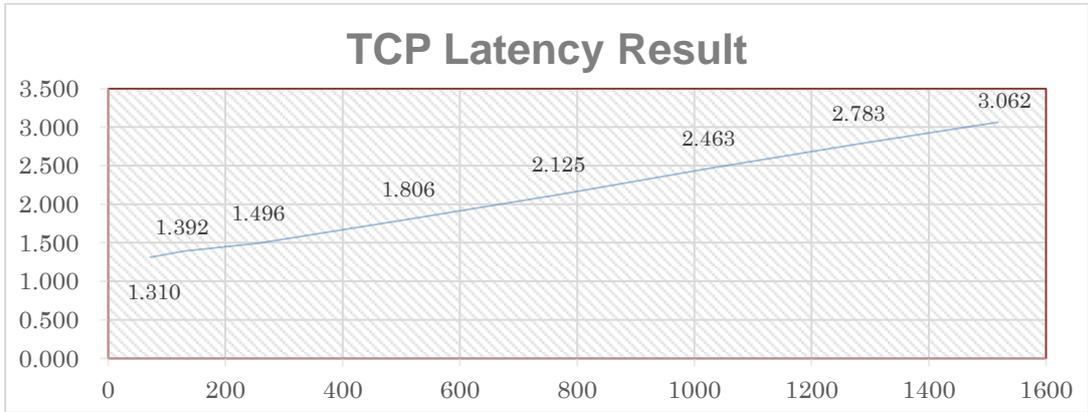


図 6 遅延に関する TCP の計測結果グラフ

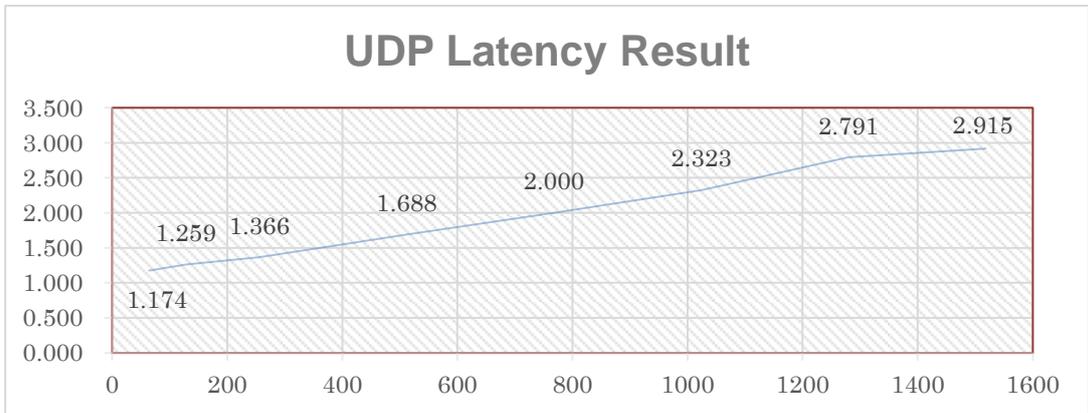


図 7 遅延に関する UDP の計測結果グラフ

図 6 の TCP の測定に用いたコマンドと結果を表 8 に、図 7 の UDP の測定結果に用いたコマンドと結果を表 9 に示す。

表 8 遅延に関する TCP の測定データ

#	Server 1	Server 2	Protocol	Packet Size(Byte)	Avg(μ s)
1	exasock taskset -c 1-7 sockperf ping-pong -i 192.168.10.2 -p 30100 -m 14 -t 5 --tcp	exasock taskset -c 1-7 sockperf server -i 192.168.10.2 -p 30100 --tcp	TCP	72	1.31
2	exasock taskset -c 1-7 sockperf ping-pong -i 192.168.10.2 -p 30100 -m 70 -t 5 --tcp	exasock taskset -c 1-7 sockperf server -i 192.168.10.2 -p 30100 --tcp	TCP	128	1.391667
3	exasock taskset -c 1-7 sockperf ping-pong -i 192.168.10.2 -p 30100 -m 198 -t 5 --tcp	exasock taskset -c 1-7 sockperf server -i 192.168.10.2 -p 30100 --tcp	TCP	256	1.495667
4	exasock taskset -c 1-7 sockperf ping-pong -i 192.168.10.2 -p 30100 -m 454 -t 5 --tcp	exasock taskset -c 1-7 sockperf server -i 192.168.10.2 -p 30100 --tcp	TCP	512	1.805667
5	exasock taskset -c 1-7 sockperf ping-pong -i 192.168.10.2 -p 30100 -m 710 -t 5 --tcp	exasock taskset -c 1-7 sockperf server -i 192.168.10.2 -p 30100 --tcp	TCP	768	2.124667
6	exasock taskset -c 1-7 sockperf ping-pong -i 192.168.10.2 -p 30100 -m 966 -t 5 --tcp	exasock taskset -c 1-7 sockperf server -i 192.168.10.2 -p 30100 --tcp	TCP	1024	2.462667
7	exasock taskset -c 1-7 sockperf ping-pong -i 192.168.10.2 -p 30100 -m 1222 -t 5 --tcp	exasock taskset -c 1-7 sockperf server -i 192.168.10.2 -p 30100 --tcp	TCP	1280	2.783333
8	exasock taskset -c 1-7 sockperf ping-pong -i 192.168.10.2 -p 30100 -m 1460 -t 5 --tcp	exasock taskset -c 1-7 sockperf server -i 192.168.10.2 -p 30100 --tcp	TCP	1518	3.061667

* sockperf の仕様により、TCP のパケットはパケットヘッダー含めて最少 72 バイトとなるためこれを最小なパケットとして計測を行った。

9 遅延に関する UDP の測定データ

#	Server 1	Server 2	Protocol	Packet Size(Byte)	Avg(μ s)
1	exasock taskset -c 1-7 sockperf ping-pong -i 192.168.101.2 -p 30100 -m 18 -t 5	exasock taskset -c 1-7 sockperf server -i 192.168.10.2 -p 30100	UDP	64	1.174333
2	exasock taskset -c 1-7 sockperf ping-pong -i 192.168.10.2 -p 30100 -m 82 -t 5	exasock taskset -c 1-7 sockperf server -i 192.168.10.2 -p 30100	UDP	128	1.259333
3	exasock taskset -c 1-7 sockperf ping-pong -i 192.168.10.2 -p 30100 -m 210 -t 5	exasock taskset -c 1-7 sockperf server -i 192.168.10.2 -p 30100	UDP	256	1.366333
4	exasock taskset -c 1-7 sockperf ping-pong -i 192.168.10.2 -p 30100 -m 466 -t 5	exasock taskset -c 1-7 sockperf server -i 192.168.10.2 -p 30100	UDP	512	1.687667
5	exasock taskset -c 1-7 sockperf ping-pong -i 192.168.10.2 -p 30100 -m 722 -t 5	exasock taskset -c 1-7 sockperf server -i 192.168.10.2 -p 30100	UDP	768	1.999667
6	exasock taskset -c 1-7 sockperf ping-pong -i 192.168.10.2 -p 30100 -m 978 -t 5	exasock taskset -c 1-7 sockperf server -i 192.168.10.2 -p 30100	UDP	1024	2.322667
7	exasock taskset -c 1-7 sockperf ping-pong -i 192.168.10.2 -p 30100 -m 1234 -t 5	exasock taskset -c 1-7 sockperf server -i 192.168.10.2 -p 30100	UDP	1280	2.791333
8	exasock taskset -c 1-7 sockperf ping-pong -i 192.168.10.2 -p 30100 -m 1472 -t 5	exasock taskset -c 1-7 sockperf server -i 192.168.10.2 -p 30100	UDP	1518	2.915

(エ) スループット確認

iperfコマンドで片方のサーバから、他方のサーバに10秒、60秒、100秒、3600秒間パケットを送信し、帯域幅を計測する。図8に測定結果を示す。
TCP/IPを介した実効データ転送性能が約7GB/秒と良好な結果となった。

図8のTCPの測定に用いたコマンドと結果を表10に示す。



図8 スループットに関する計測結果グラフ

表10 スループットに関する計測データ

#	Server 1	Server 2	Time (s)	Avg(Gbp/s)	Max	Min	Memo
1	exasock iperf3 -c 192.168.11.2 -w1024k -t 10	exasock iperf3 -s	10	7.075	7.3	6.82	
2	exasock iperf3 -c 192.168.11.2 -w1024k -t 60	exasock iperf3 -s	60	7.129	7.32	6.86	
3	exasock iperf3 -c 192.168.11.2 -w1024k -t 100	exasock iperf3 -s	100	7.0955	7.33	6.82	
4	exasock iperf3 -c 192.168.11.2 -w1024k -t 3600	exasock iperf3 -s	3600	6.96	7.02	6.9	

7. まとめ

今回の検証により、Exablaze 社製イーサネット・アダプター[ExaNIC]が正常にかつ期待通りの性能で富士通サーバ[RX2540 M1]で動作することが確認できた。性能についてはメーカーの公表値と同等のパフォーマンス(遅延)となり、スループットも TCP/IP を介した実効データ転送性能としては良好な結果となった。このことから、[ExaNIC]は富士通サーバ[RX2540 M1]との組み合わせにおいて、システムやアプリケーションの性能を向上に有効であることを確認した。

8. 参考情報、お問い合わせ先

ビットリーブ株式会社

〒151-0073
東京都渋谷区笹塚3-2-15 第2ベルビル6F
電話: 03-6276-0880
電子メール: sales@bitrieve.co.jp
担当: 営業部 中山、工藤

Exablaze 社の概要

メルボルンに本社を置く Exablaze 社は、革新的なエンジニアリングの洞察とプロセスにより、世界トップクラスの低遅延ネットワーク デバイスを設計、開発、製造。この企業の ExaLINK 50 レイヤー 1 スイッチ、ExaNIC X2/X4 イーサネット・アダプターおよび関連の FPGA 開発キットは、金融、ハイパフォーマンスコンピューティング、クラウド、ストレージ、データセンター、電気通信、エネルギー、防衛、教育などの幅広い業界の遅延重視の組織で普及が進んでいる。

ビットリーブ社の概要

2004 年に東京で設立されたビットリーブ株式会社は、評価のためのネットワークトラフィックの可視化およびネットワーク環境の保護に焦点を当てた、IP ネットワーク用のテストおよび監視機器の供給企業。日本の IT 市場に適した革新的な製品やツールを世界中から見つけ出し、紹介、提供することを目的とする。