Vol.2013-IOT-22 No.1
2013/8/1

x86 サーバにおける 40Gigabit Ethernet 性能測定と課題

松本直人[†1]

本稿では，40Gigabit Ethernet の性能評価と課題について紹介する．40Gitbit Ethernet は x86 サーバの OS 対応は発展途上の部分が多く残っており，導入を前提とした際に問題となる知見を共有するために，高速ネットワークでの具体的な性能測定ツールの紹介および測定段階での課題について検証する．

The 40 Gigabit Ethernet Performance Analysis on x86 Server

NAOTO MATSUMOTO[†1]

This paper is a introduce to analysis a 40 Gigabit Ethernet performance on x86 server environment. The 40 Gigabit Ethernet is still emerging technology in current industry, it has some problem relevant operating system and network driver yet. This paper is shared for you some kind of bad know how and technical issue.

1. はじめに

40 Gigabit Ethernet は IEEE 802.3 ETHERNET WORKING GROUP により標準化された新たな広帯域ネットワーク通信規格です．[1] 標準化は 2008 年から行われており，現在 40 Gigabit Ethernet に対応したネットワークインターフェイスカード(以下 NIC)およびスイッチ(以下 Switch)が商品化されています．しかしながら当該機器の情報は少なく，Windows や Linux に代表される標準的な OS への対応も始まったばかりです．本稿は x86 サーバ環境における 40 Gigabit Ethernet の利用と課題についての情報共有を目的としています．

x86 サーバにおける 40 Gigabit Ethernet NIC の動作には最新のファームウェア，ドライバおよび関連ツールが必須となっています．現在入手可能な最新の Linux カーネル[2]環境であれば, 既に 40 Gigabit Ethernet NIC ドライバおよび関連ツールは最新バージョンに保たれているため問題に遭遇することはありませんが，旧来からの OS 環境で 40 Gigabit Ethernet NIC を動作させる場合に必要となります．これら更新作業を行わない場合には，設定情報が正しく表示されない場合があり，ネットワークシステム運用上の問題点を内在することになるます．

40Gigabit Ethernet 性能測定を行う場合，トラフィック生成を行う機器が必要とされます．しかし専用ハードウェアを用いた機器は高価であり，容易に入手できるものではありません. 本稿では x86 サーバ上で動作する標準的な Linux 環境において 40 Gigabit Ethernet NIC を用いたトラフィック生成および測定評価の手法について情報共有します．

2. ドライバ関連ツールの問題点

Linux 環境において 40 Gigabit Ethernet NIC の動作確認に必要とされるツールに ethtool，lspci があります．ethtool は query or control network driver and hardware settings とされており，40 Gigabit Ethernet NIC のドライバおよびハードウェアの確認および設定に用いられます．[3] lspci は list all PCI devices とされており, 40 Gigabit Ethernet NIC の PCI Express 上での接続状況の確認に用いられます．[4] いずれもハードウェアに近い部分の設定および動作確認に用いられるため，正確な情報を得ることが極めて重要になります．

特に pcutils に含まれる lspci の場合，動作する x86 サーバ環境が PCI Express 3.0 [5]に対応していた場合に有効に働きます．

40 Gigabit Ethernet NIC 動作において，当該ツールのバージョンが古い場合，正確な情報を得ることが出来ず，事実誤認による運用障害を引き起こすことも想定されるため注意が必要です．問題点への対処としては，現在入手可能な最新の Linux カーネル環境を用意するか，当該ツールの最新版をレポジトリ[3][4]よりダウンロードした上でインストールする必要があります．

†1 さくらインターネット株式会社 さくらインターネット研究所
SAKURA Internet Research Center, SAKURA Internet, Inc.

3. トラフィック生成ツールの特性と理解

ネットワーク性能測定の代表的なツールとして，iperf [6], netperf [7]があります．いずれも TCP や UDP によるトラフィック生成と受信をクライアント・サーバに分かれて動作します．40 Gigabit Ethernet 環境でも従来通りにネットワーク性能測定は可能ですが，さらに低レイヤでの性能測定を行うには不十分です．Linux 環境で IP 層以下の低レイヤの性能測定を行う場合には，pktgen [8] が有効です．

pktgen は IP アドレス，MAC アドレス，VLAN ID と IP パケットサイズを指定範囲で組み合わせてトラフィック生成できるツールです．IP パケット送信タイミングをナノ秒単位で調節でき，動作は割り当てた CPU を占有して動作するため極めて高速に動作します．

pktgen を利用する理由として，Linux Kernel 2.6.35 からパケットやフロー単位でマルチコア CPU を円滑利用する RPS(Receive Packet Steering)と RFS(Receive flow steering)が導入されてたこと，NIC に TCP/UDP ハードウェアオフロード機能を有していることが上げられます．実環境に近いトラフィック生成と性能測定を行うことは，40 Gigabit Ethernet 本来の性能を俯瞰的に確認することができ，上位アプリケーションでのボトルネックをより明確に切り分けることが可能となります．(図 1)

```
#!/bin/sh
echo "rem_device_all"           > /proc/net/pktgen/kpktgend_0
echo "add_device eth0"          > /proc/net/pktgen/kpktgend_0
echo "count 0"                  > /proc/net/pktgen/eth0
echo "clone_skb 1"              > /proc/net/pktgen/eth0
echo "pkt_size 64"              > /proc/net/pktgen/eth0
echo "delay 0"                  > /proc/net/pktgen/eth0
echo "dst 10.10.11.2"           > /proc/net/pktgen/eth0
echo "dst_mac 00:04:23:08:91:dc" > /proc/net/pktgen/eth0
echo "start"                    > /proc/net/pktgen/pgctrl
```

図 1 pktgen トラフィック生成スクリプト例
Figure 1 pktgen Traffic generate script example.

Linux 環境では 64 バイトなど小さいパケット処理性能が低いと言われており，その比較も極めて重要な測定です．

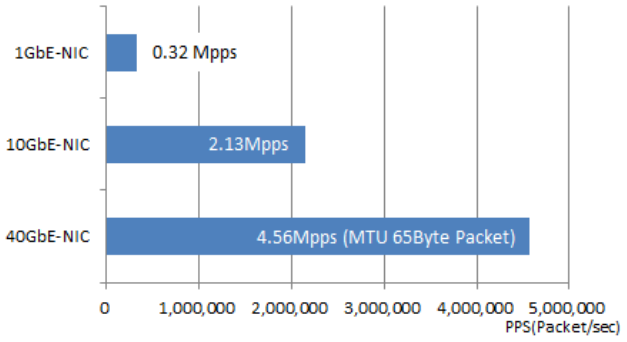

図 2 1/10/40 Gigabit Ethernet パケット受信処理性能
Figure 2 1/10/40 Gigabit Ethernet RX Packet Process result

図 2 では，Realtek RTL8168B 1GbE-NIC，Intel 82599EB 10GbE-NIC，Mellanox ConnectX3 40GbE-NIC を搭載する 2 台の x86 サーバをケーブルで直結した環境 (図 3) で評価したものであり，その測定結果からも 40 Gigabit Ethernet とその他の違いが見てとれます．

送信側 x86 サーバ環境
CPU: Intel Core i7-3930K 3.20GHz, 32GB-DRAM
OS: Linux 3.7-rc7
Bus: PCI Express 2.0

受信側 x86 サーバ環境
CPU: Intel Core i7-3930K 3.20GHz, 32GB-DRAM
OS: Linux 3.7-rc7
Bus: PCI Express 2.0

図 3 トラフィック測定環境
Figure 3 Traffic Analysis Environment

4. トラフィック測定ツールについて

トラフィック測定をリアルタイムに視認しながら行うツールとして vnstat が有効です．[9] vnstat は a console-based network traffic monitor とされており，各 NIC 単位でのトラフィック測定を可能とします．(図 4)

```
# vnstat -l -i eth1
Monitoring eth1... (press CTRL-C to stop)

 rx:  29.56 Gbit/s 61243 p/s          tx:  0 Mbit/s 0 p/s
```

図 4 vnstat トラフィック測定の例
Figure 4 vnstat Traffic Real-time Monitoring

vnstat は転送レートおよびパケット処理性能を秒単位で計測可能なツールであり，トラフィック受信および生成の状況をリアルタイムで確認しながら計測を行うことを可能とします．pktgen と vnstat を組み合わせることで，高価なハードウェア機器によるトラフィック測定環境を組むことなく簡易に 40 Gigabit Ethernet 測定環境が構築できます．

5. 測定結果

pktgen と vnstat を用いた 40 Gigabit Ethernet 環境のトラフィック測定結果を以下に記します．(図 5) トラフィック測定環境は Realtek RTL8168B 1GbE-NIC，Intel 82599EB 10GbE-NIC，Mellanox ConnectX3 40GbE-NIC を搭載する 2 台の x86 サーバをケーブルで直結した環境 (図 3) となっており，特に小さいパケット処理性能差について着目しました．

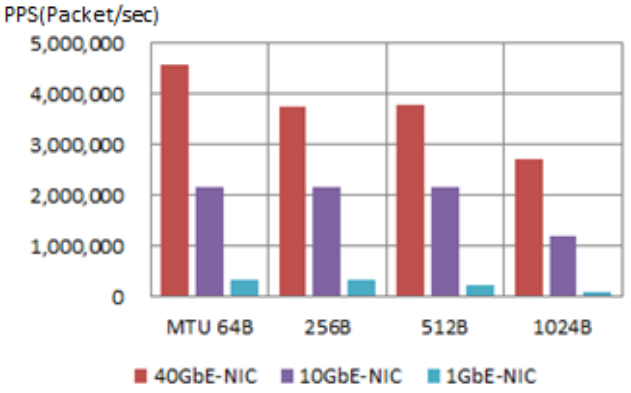

図 5 1/10/40 Gigabit Ethernet パケット受信処理性能(2)
Figure 5 1/10/40 Gigabit Ethernet RX Packet Process result'(2)

図 5 から 10Gigabit Ethernet に比べて 40Gigabit Ethernet でのパケット処理性能は全体的に向上している点が確認できます．ハードウェア機器を用いた性能評価と同じく pktgen と vnstat を用いたトラフィック測定でも性能比較が行えることが確認できました．

6. まとめ

40 Gigabit Ethernet という広帯域ネットワーク通信規格の性能測定を考えた場合，トラフィック生成に用いるハードウェア機器が高価であるという課題がありました．本稿の実験において，既存の x86 サーバ環境とオープンソースソフトウェアのみでもトラフィック測定と性能比較が行えることが確認できました．しかしながら 40 Gigabit Ethernet は新しい技術であり日進月歩でドライバおよび関連ツールの修整が行われています．40 Gigabit Ethernet を用いたシステム構築を考えた場合，常に最新の安定したドライバおよび関連ツールの利用を強く推奨するとともに，性能測定の前提となる pktgen および vnstat, ethtool, lspci など関連ツールの利用に習熟しておく必要があります．本稿を通じて新たな 40 Gigabit Ethernet のネットワークシステム構築の理解が深まりましたら幸いです．

参考文献

1) IEEE 802.3 ETHERNET WORKING GROUP
http://www.ieee802.org/3/
2) Linux kernel
https://www.kernel.org/
3) ethtool
http://git.kernel.org/cgit/network/ethtool/ethtool.git/
4) lspci
https://www.kernel.org/pub/software/utils/pciutils/
5) PSI-SIG PCI Express Base 3.0 specification.
http://www.pcisig.com/specifications/pciexpress/base3/
6) iperf
http://code.google.com/p/iperf/
7) netperf
http://www.netperf.org
8) pktgen
http://www.linuxfoundation.org/collaborate/workgroups/networking/pktgen
9) vnstat
http://humdi.net/vnstat/