

論文

非対称な通信路における TCP 伝送特性の評価

後藤 沙菜 藤井 哲郎

インターネットの通信を支えているのは、TCP/IP プロトコルである。特に TCP はコネクション型通信の代表であり、信頼性を保証したトランスポート層のプロトコルである。この TCP による通信は双方向であり、往復の通信経路が同じであれば、送信側と受信側の伝送特性は対称であると考えられる。ところが、根本らにより送信側と受信側の伝送速度が異なると伝送特性が必ずしも対称とならない場合があることが報告されている。本稿では、様々な伝送経路を網羅的に測定し、送受信の伝送特性が対称とならない通信が発生する条件の洗い出しを行った結果を報告する。送受信の NIC のインタフェース速度が異なる場合、NIC 間に設置される LAN スイッチにおいて伝送速度の変換が行われる。この速度変換の機構により伝送特性が異なってくる場合があることを明らかにしている。

キーワード：TCP, Cubic, 40GbE, 10GbE, LAN スイッチ, 非対称ネットワーク

1 まえがき

インターネット通信の代表的プロトコルといえば、TCP/IP である。特に TCP はコネクション型通信を代表する信頼性を保証したトランスポート層のプロトコルである。例えば、信頼性のある通信を必要とする Web の閲覧を行う http 通信、メールの転送を行う SMTP、ファイル転送を行う FTP などに幅広く用いられている。同時に TCP により、インターネットの輻輳も回避できるように設計されている。

今まで TCP には様々な改良が加えられてきた。歴代の代表的なアルゴリズムとして Tahoe, Reno, New Reno, Vegas, Compound TCP, Cubic などが広く知られている。特に Compound TCP は Windows Vista 以降に導入された高速 TCP である。これに対し、Cubic は Linux 2.6.1.9 以降のカーネルで実装され、最新の Windows 10 及び Windows Server 2019 でも採用されている^[1, 2]。これらはいずれも、通信が始めるときにはスロースタートと呼ばれている手順、即ちネットワークの帯域が解らないのでゆっくりと送信を開始する。回線帯域に余裕があればウィンドウサイズを迅速に増加させ、ネットワークが混雑してきたらウィンドウサイズを緩やかに減らす。そしてパケットロスを検知したら、一気にウィンドウサイズを減少させるという具合に動作する。この機能により、ネットワークの輻輳を回

避するという、重要な役割を担っている。

この TCP による通信は双方向であり、往復の通信経路の帯域が同じであれば、送信側と受信側の伝送特性に関して対称であると考えられる。ところが、根本らにより、40GbE と 10GbE の NIC 間の様に非対称な伝送環境において、送信側からのフローと受信側からのフローでは伝送速度が異なる場合が存在することが報告されている^[3]。一般に速度が異なる場合には LAN スイッチにおいてデータバッファを用いて速度変換が行われる。無限にバッファが存在する場合には、フローの方向により差は発生しないはずであるが、実際には有限であるために、フローの方向により速度が異なる場合が発生する。本稿では、どのような条件が揃うと TCP による伝送速度が対称でなくなるのか、網羅的かつ詳細に調査し、その結果を以下に報告する。なお、Windows Server 2019 と Ubuntu といった異なる OS では組み込まれている TCP プロトコルの種類が違うため、伝送速度が必ずしも対称とはならない、従って、異なる OS 間の TCP による伝送特性の評価は本稿では行わないこととする。

2 TCP 伝送特性の測定方法

PC の性能が通信に影響を与えないように、同じ種類のマザーボード (ASRock 社製 H77 Pro4-M) に同じ種類の CPU (Intel 製 Core i7 3770) を搭載する完全に同一スペックの PC を 2 台用意する。この PC のマザーボードに GbE インタフェースとして組み込まれているのがリアルテック製の RTL 8111E チップである。この 2 台の PC に用いる OS は、Windows Server 2019、

GOTO Sana
東京都市大学メディア情報学部情報システム学科 2020年度 4 年生
FUJII Tetsuro
東京都市大学メディア情報学部情報システム学科教授

Ubuntu 18.04, Ubuntu 20.04 の 3 種類である。3 台のハードディスクを入れ替えて立ち上げ、測定を行う。この 2 台の PC 間に Lan スイッチとネットワークエミュレータ netem を組み込んだ PC (HP 製 EliteDesk 800 G1 SFF, Ubuntu 18.04) を挿入して遅延時間を変化させながら、TCP の伝送特性の計測を行う。測定システムの構成を図 1 に示す。Ubuntu18.04 と Ubuntu 20.04 では TCP による通信速度の測定を nuttcp と iperf3 で行う。これに対し、Windows Server 2019 では iperf3 のみを用いて計測を行う。なお、ネットワークシミュレータ機能である netem において伝送遅延を両方向に関して発生させることにより、RTT (Round Trip Time) を変化させながら測定を行っている。

例えば非対称な 1GbE と 10GbE の NIC 間の通信では、1GbE から 10GbE 方向と逆の 10GbE から 1GbE 方向の双方向の伝送速度の計測を行う。これを起動する HDD を変えることによって OS も変更させ、3 パターン測定する。また、標準パケット (MTU 1500Byte) とジャンボパケット (MTU 9000Byte) に関しても比較の為に測定を行う。TCP による通信は 60 秒間連続し

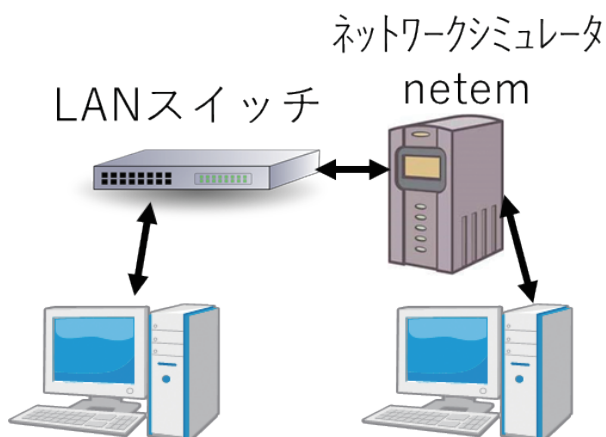


図 1 TCP 伝送特性測定システムの基本構成

表 1 測定の用いた LAN スイッチ及び NIC

LAN スイッチ	NEC 製 QX-S5524GT-4X2Q (40GbE, 10GbE, GbE ポート) Netgear 製 Prosafe XS708E (10GbE ポート) Buffalo 製 LXW-10G2/G4 (2.5GbE, 10GbE ポート)
NIC	Intel X710 QDA2 (40GbE) Intel X540-T1 (10GbE) QNAP QXG-5G1T-111C (5GbE) OKN PCE-2500 (リアルテックチップ) (2.5GbE) RTL 8111E (1GbE) Intel Pro/1000 PT Dual Port (1GbE)
トラン シーバ	10Gtek 製 ASF-10G-T (形状 SFP+) Multirate 対応 (10GbE/5GbE/2.5GbE/1GbE)

て計測を行い、平均速度を算出する。この時、netem を用いて RTT を 0ms から 100ms まで変化をさせて測定を行う。用いた 40GbE, 10GbE, 2.5GbE, GbE NIC と LAN スイッチの種類、用いたトランシーバを表 1 に示しておく。

3 対称な通信路での基本特性

3.1 1GbE NIC 間での伝送特性

まず、基本特性を把握するために、伝送速度が対称な NIC 間の TCP による伝送特性の測定を行う。Intel 製の NIC (Pro/1000 PT Dual Port) を用いた時の伝送特性の測定結果を図 2 に示す。LAN スイッチは NetGear 製 Prosafe XS708E 1 台を 1Gbps モードで利用した。同図より、Windows Server 2019 の TCP ドライバーよりも Ubuntu の TCP ドライバーの方が遅延に対して強く、RTT が 20msec まで伝送速度が全く遅くなっていないことが明らかである。RTT 20msec は、東京・鹿児島間程度であり、長距離伝送の特性が大幅に改善されていることが明らかである。最新の Ubuntu を用いれば、TCP による通信において、日本国内の通信では遅延に関する問題が殆ど発生しないことが解る。なお、同図において、Ubuntu 18.04 と 20.04 は完全に特性が一致し、グラフ

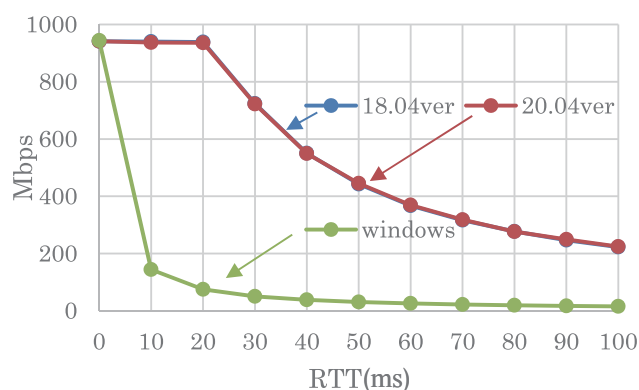


図 2 Intel 製 1GbE NIC 間の TCP 伝送特性 (MTU 1500Byte)

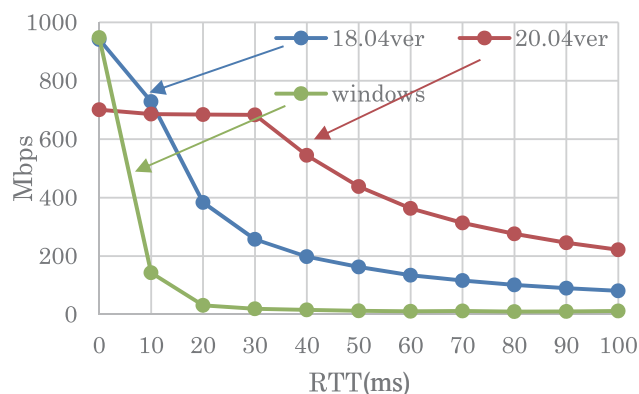


図 3 RTL8111E 1GbE 間の TCP 伝送特性 (MTU 1500Byte)

は完全に重なっている。Ubuntu のバージョンによる差が全く無いことより、以下 Intel 製の NIC に関しては、20.04 での測定結果のみを掲載することにする。

マザーボードに組み込まれたリアルテック製チップ RTL 8111E を用いた時の TCP による伝送特性の測定結果を図 3 に示す。同図より、Windows Server 2019 の TCP ドライバーよりも Ubuntu の TCP ドライバーの方が遅延に対して強いことが解る。但し、Ubuntu 20.04 は 18.04 と比較して、遅延に対しては RTT 30msec まで劣化しないものの、RTT が 0 msec における伝送特性が 700Mbps しかないことも明らかである。Ubuntu 20.04 と Ubuntu 18.04 は、どちらも sysctl コマンドでチェックすると cubic が用いられていると表示される。従って、20.04 に関しては、リアルテック製チップに関して十分なチューニングができていないことが推察される。このように TCP 伝送特性は、ドライバーのチューニングにも大きく影響を受ける場合があることもありえることを考慮しておく必要がある。

3.2 10GbE NIC 間での伝送特性

より高速で、伝送速度が対称な NIC 間伝送特性を把握するために、Intel 製の 10GbE NIC (X540-T1) を用いた時の Windows Server 2019 と Ubuntu 20.04 の TCP 伝送特性の測定結果を図 4 に示す。LAN スイッチは NetGear 製 Prosafe XS708E を利用し、ジャンボパケットである MTU を 9000Byte に設定して測定した結果である。同図より、Windows Server 2019 の TCP ドライバーよりも Ubuntu 20.04 の TCP ドライバーの方が伝送速度の点で優れていることが解る。RTT 0msec で、Ubuntu が 9.9Gbps の速度が出ているのに対して、Windows Server 2019 では約 1/3 の 3.3Gbps しか伝送速度が出ていない。また、RTT が 20msec でも Ubuntu では 1.2Gbps の伝送速度を保っているが、Windows Server 2019 では 76Mbps にまで劣化してしまっている。TCP の制御アルゴリズムとして同じ Cubic を採用しているが、インプリメント及びチューニングの差が大きいことが明らかである。

3.3 40GbE NIC 間での伝送特性

さらに高速で、伝送速度が対称な NIC 間伝送特性を把握するために、Intel 製の 40GbE NIC (GT100) を用いた時の Windows Server 2019 と Ubuntu 20.04 の TCP 伝送特性の測定結果を図 5 に示す。LAN スイッチは NEC 製 QX-S5524GT-4X2Q を利用した。Ubuntu 20.04 では、ほぼフルスピードの 39.4Gbps をひきだせているが、Windows Server 2019 ではその 1/4 の 8.0Gbps にとどまっている。10GbE と比較して、Windows Server 2019 と Ubuntu 20.04 の TCP の伝送

速度の差がさらに大きくなってしまっていることが明らかである。但し、40GbE で Windows Server 2019 で 8Gbps 出ているということは、10GbE においても、チューニングをしっかりと行えば、RTT 0msec での TCP の伝送速度を 8Gbps まで改善できるのではないかと推察される。

4 非対称な通信路での特性評価

4.1 40GbE LAN スイッチを用いたケースの解析

装置前面に 10/100/1000Base-T(1GbE)を 24 ポート、SFP+ (10GbE) を 4 ポート、装置背面に QSFP+ (40GbE) を 2 ポート装備する NEC 製 LAN スイッチ QX-S5524GT-4X2Q を用いて、40GbE と 10GbE NIC 間の非対称な通信を行う。この LAN スイッチは、速度変換の為に大容量の内部データバッファを有している。Windows Server 2019 間の TCP による伝送の特性を測定した結果を図 6 に示す。このとき、10Gbps と 40Gbps の速度変換は LAN スイッチの内部で行われている。同図より明らかなように、RTT 0msec において、40GbE から 10GbE の方向には 5.5Gbps 伝送できているのに対し、逆方向は 3.4Gbps しか伝送速度が出ていない。このように、フローの方向により伝送速度が異なる。

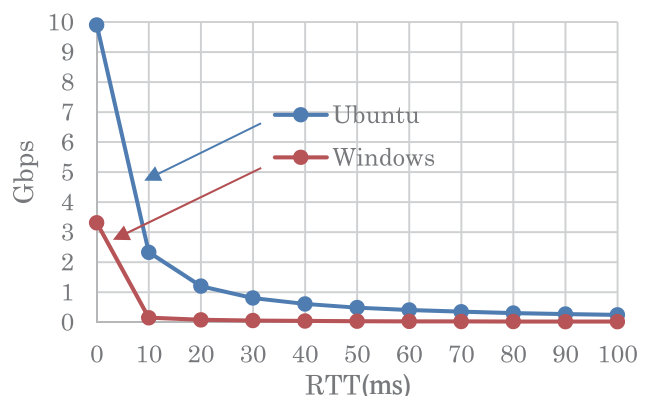


図 4 10GbE NIC 間の TCP 伝送特性 (MTU9000Byte) Windows Server 2019 と Ubuntu 20.04 の比較

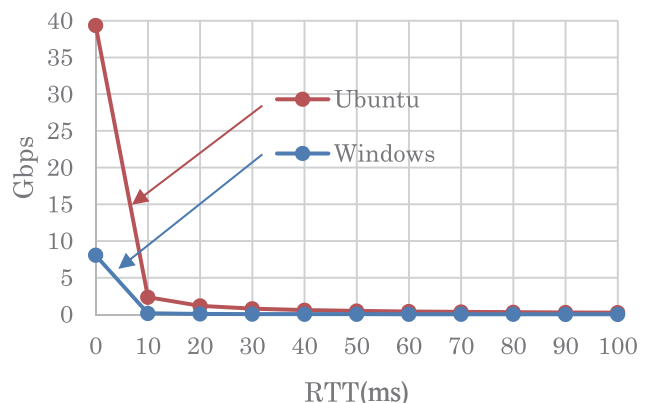


図 5 40GbE NIC 間の TCP 伝送特性 (MTU9000Byte) Windows Server 2019 と Ubuntu 20.04 の比較

るという現象が確認でき、根本による報告を追試できた^[1]。これはTCPのWindow制御機能とLANスイッチ内に位置するデータバッファの制御に伴い発生すると考えられる。また、スピードの速い40GbEから10GbEの方向に向けたフローの方が早いことより、LANスイッチの内部データバッファがオーバーフローを起こしてはいないと考えられる。詳細な動作の解析のためには、輻輳Windowの制御から検討を進める必要がありそうである。

図6の構成において10GbE NICをIntel製の1GbE NICに交換して同様な測定を行った結果を図7に示す。同図より、データ伝送速度が1Gbpsと1/10の低い伝送速度であるためにLANスイッチ内のデータバッファ量が無限とみなせる理想的な状態になり、フロー方向による伝送レートの差が表れていないことが解る。

4.2 2.5G-10GbE LANスイッチを用いたケースの解析

2.5GbEを4ポート、10Gbase-Tを2ポート有するBuffalo社製のLANスイッチを用いた速度変換でのフ

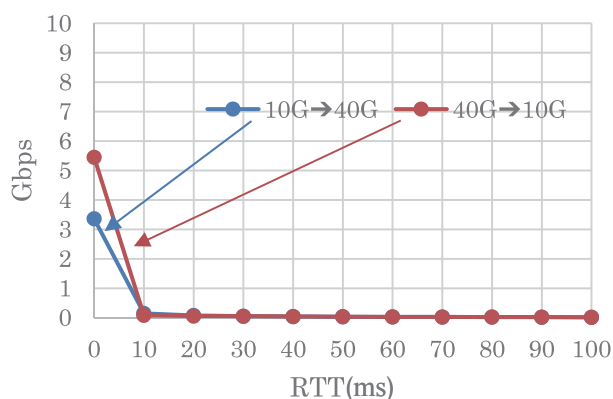


図6 40GbEと10GbE NIC間のWindows Server 2019におけるTCP伝送特性 (QX-S5524GT-4X2Q, MTU 1500Byte)

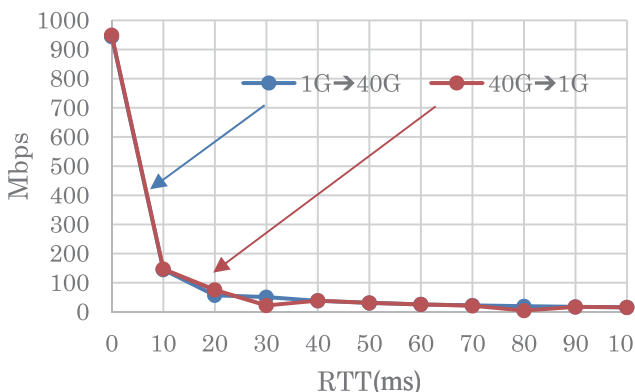


図7 40GbEと1GbE NIC間のWindows Server 2019におけるTCP伝送特性 (QX-S5524GT-4X2Q, MTU1500Byte)

ロー方向で伝送特性の評価を行う。LANスイッチの機種による差を検証するためである。測定では、2.5GbE用NICとしてOKN製PCE-2500を用いる。Ubuntu 20.04における測定結果を図8に示す。同図より明らかに、RTT 0msecにおいて、2.5GbEから10GbEの方向には2.35Gbps伝送できているのに対し、逆方向は2.26Gbpsと若干下がっている。同様に、Windows Server 2019における測定結果を図9に示す。同図より明らかに、RTT 0msecにおいて、2.5GbEから10GbEの方向には2.2Gbps伝送できているのに対し、逆方向は1.88Gbpsと同様に若干下がっている。両者で共通するのは、前節とは異なり、スピードの遅い2.5GbEから10GbEの方向に向けたフローの方が伝送速度が早い。これは、LANスイッチの内部データバッファが若干ではあるがオーバーフローを起こしている可能性を示唆している。

4.3 マルチレートSFP+における速度変換における特性

さらにLANスイッチ本体ではなく、トランシーバであるSFP+ポートにおける速度変換を利用した場合のTCPによる伝送特性を評価した結果を示す。トラン

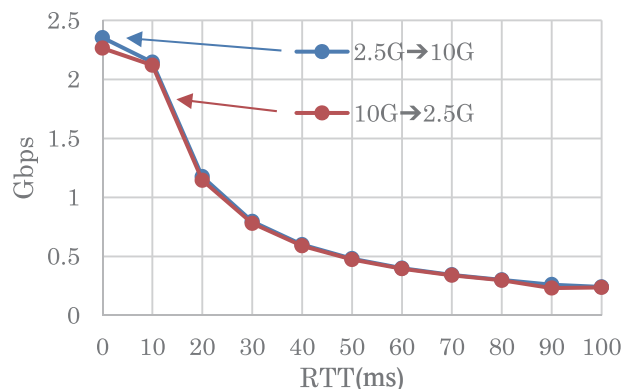


図8 10GbEと2.5GbE NIC間のUbuntu 20.04におけるTCP伝送特性 (MTU 1500Byte)

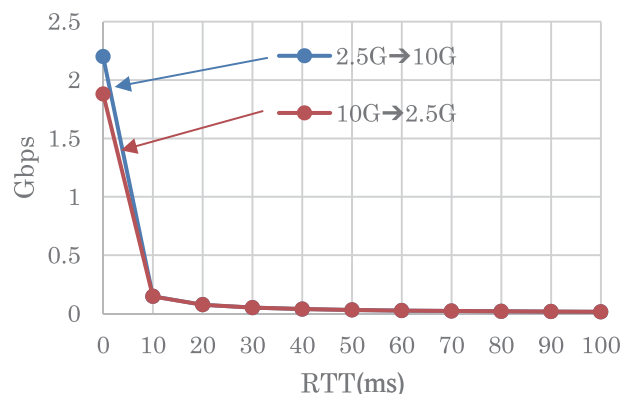


図9 10GbEと2.5GbE NIC間のWindows Server 2019におけるTCP伝送特性 (MTU 1500Byte)

シーバにおける内部データバッファ量は非常に小さいことが推察され、十分なバッファ容量を有する LAN スイッチにおける特性とは大きく異なることが予想される。用いたのは 10GTek 製のトランシーバ ASF-10G-T (形状 SFP+) であり、SFP+ において 10GbE, 5GbE, 2.5GbE, 1GbE の 4 種類の伝送速度に対応している優れモノである。

LAN スイッチとして NEC 製 QX-S5524GT-4X2Q を用い、この 10Gbps のポートに SFP+ を挿入して計測を行った。5GbE の NIC として QNAP QXG-5G1T-111C を用い、10GbE の NIC として Intel X540-T1 を用いて、Ubuntu 20.04 における TCP の伝送速度を計測した結果を図 10 に示す。同図より明らかなように、5GbE から 10GbE の方向では 4.7Gbps の伝送速度が出ているのに対し、逆方向の 10GbE から 5GbE 方向には 1/3 以下の 1.3Gbps しか伝送速度が出ていない。非対称性が明確に表れる結果となった。速度の速い NIC から速度が遅い NIC に対するフローの速度が大幅に劣化している。

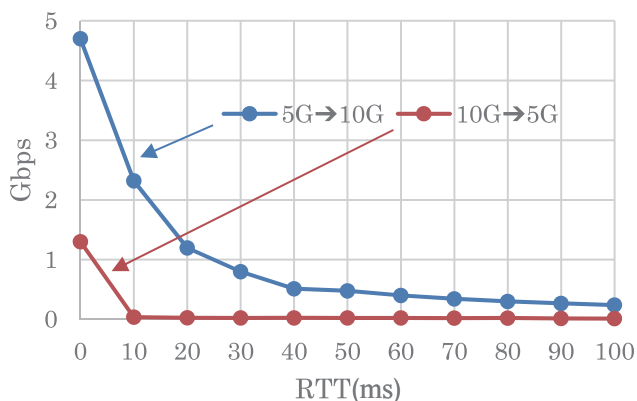


図 10 Intel 製 10GbE と TP-Link 製 5GbE NIC 間の Ubuntu 20.04 における TCP 伝送特性, SFP+ で 10Gbps を 5Gbps に速度変換 (MTU 1500Byte)

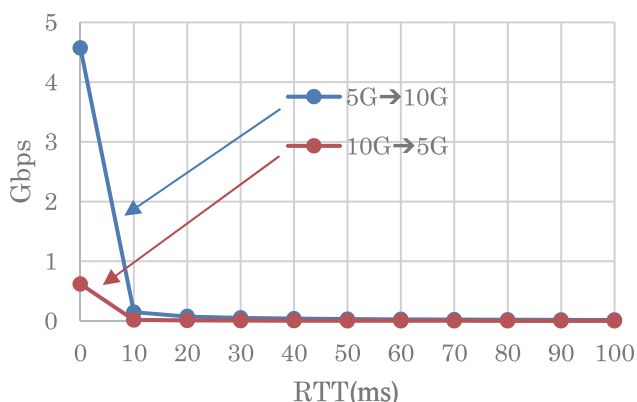


図 11 Intel 製 10GbE と TP-Link 製 5GbE NIC 間の Windows Server 2019 における TCP 伝送特性, SFP+ で 10Gbps を 5Gbps に速度変換 (MTU 1500Byte)

これは、明らかに伝送速度を変換するポイントに存在するデータバッファにおいて大量のオーバーフローが発生していることを示唆している。また、TCP の Window 制御の制御間隔の時間が長すぎる結果であるとも推察される。同様の測定を Windows Server 2019 に関して行った結果を図 11 に示す。同図より明らかなように、5GbE から 10GbE の方向では 4.6Gbps の伝送速度が出ているのに対し、逆方向の 10GbE から 5GbE 方向には 1/7 以下の 620Mbps しか伝送速度が出ていない。Ubuntu と同様な傾向が確認できた。

図 10 及び図 11 から明らかなように、トランシーバにおける伝送速度の変換は、データバッファの容量が小さいためにオーバーフローが頻発し、パケットロスが多発していると推定できる。このような利用方法はできるだけ避けるべきと考えられる。現在、ギガビットイーサの高速化を模索して従来の 10GbE ではなく、中間的な速度である 2.5GbE 或いは 5GbE がその手軽さから普及を始めようとしている。この場合、ネットワークの構成として安易にトランシーバで伝送速度の変換を行うべきではなく、LAN スイッチ内の大容量のデータバッファを利用した速度変換が必須であると考えられる。ネットワーク全体の設計方針をしっかりと持つ必要があることが示されたともいえる。

5 まとめ

非対称に接続された NIC 間の TCP による伝送において、その伝送速度が非対称となる現象が様々な条件下で確認できた。これらすべてに共通する発生現象があるわけではなく、ケースバイケースでその発生要因が異なっていると思われる。特に、高速な NIC から低速な NIC に向けた伝送のみならず、逆のフローにおいても伝送速度の劣化が発生することが確認できた。LAN を構築する場合に必要なことが明らかとなった。また、高速に NIC (10GbE と 40GbE) 間の伝送を行う場合、Windows Server 2019 の TCP プロトコルが Cubic であるにも関わらず、チューニングが不十分と思われ、Ubuntu と比較して十分な能力を発揮できていないことも付加的に明らかとなった。

2002 年に 1Gbps ギガイーサの次世代として 10GbE が開発されたが、これまでデータセンターのみで用いられ、一般には全く普及してこなかった。そこで、中間的な速度である 2.5GbE が新たに登場し、これから普及が始まろうとしている。ところが本稿で明らかとなったように、速度変換をトランシーバで行うネットワーク構成の場合、伝送フローの方向によってはパケットロスが発生して、その能力が全く発揮できない場合があることが示された。2.5GbE 或いは 5GbE の普及に向けて注意点を喚起できたとも考えている。

参考文献

- [1] 水野「コンピュータネットワーク概論」共立出版
- [2] 伊藤「高速 TCP& 次世代トランスポートプロトコル」ASCII.Technologies, pp.46-61, 2010 年 10 月
- [3] 根本「Windows Server 2019 Hyper-V を用いた非対称 TCP 伝送特性の評価」東京都市大学, メディア情報学部, 卒業論文, 2019 年度