

## 長距離TCP通信高速化への挑戦

平木 敬<sup>\*1</sup> 稲葉真理<sup>\*1</sup> 菅原 豊<sup>\*1</sup>  
吉野剛史<sup>\*1</sup> 玉造潤史<sup>\*2</sup> 加藤 朗<sup>\*3</sup>

<sup>\*1</sup> 東京大学情報理工学系研究科 <sup>\*2</sup> 東京大学理学系研究科  
<sup>\*3</sup> 東京大学情報基盤センター

## はじめに

近年のネットワーク技術の進歩に伴い、ネットワークは現在の社会基盤として欠くべからざるものになってきている。1990年代 World Wide Web の提案に始まる情報革命は、それまでの通信を質・量ともに変化させ社会基盤の変化をもたらした。動画・音楽配信など高いネットワーク帯域を必要とするアプリケーションが普及し、これと併行してネットワークの広帯域化が急速に進みつつある。

たとえば、1983年に10Mbpsの10BASE-5 (IEEE802.3 CSMA/CD) が標準化されてから1995年にFast Ethernet (IEEE 802.3u)が標準化されるまで12年の歳月が流れたが、ギガビットイーサネット (以下GbE) 1000BASE-SX (IEEE802.3z) が標準化されたのは1998年、そして10ギガビットイーサネット (以下10GbE) 10Gbps Ethernet over fiber (IEEE802.3ae) が標準化されたのは2003年であった。

しかしながら、インターネットにおけるデータ通信で中心的な役割を果たすTCP通信は、ネットワークの高速化、遠距離化に伴い困難が増してきた。たとえば、2002年に米国SC02 (SuperComputing)のバンド幅チャレンジに初参加した私たちのData Reservoirプロジェクト<sup>6)</sup>は12,000km、622Mbpsボトルネック APAN/TransPAC OC-12の日米間通信でTCP/IP並列ストリームを利用したディスク間データ転送を行い585Mbps、ネットワークバンド幅を平均9割以上利用した。しかし、そのためには26台のサーバを対向させることが必要であり、並列TCPストリームではTCPストリーム1本あたりのバンド幅が低下すること、ストリーム間の性能のばらつきが大きく、並列ストリーム全体の性能が大きく低下するという問題点が確認された。さらに、SC04のバンド幅チャレンジでは、パケットロスによるTCPの性能低下と、性能低下からの非常に遅い復活が、

問題点であることが確認された。

これらは、古くからTCP通信を遠距離・高バンド幅ネットワークで使用するときの問題点として指摘されていたことであり、Internet2 Land Speed Record (インターネット速度記録、以後LSRと略記する)は超遠距離・超高速TCP通信の効率向上を目的として設定されたものである。本稿ではまずLSRの概要を述べ、私たちが過去4年間に行ってきた挑戦と性能向上を示す。次に、超遠距離・超高速ネットワークにおけるTCP性能向上技術、LSR挑戦において経験した、非技術的困難点について論じる。

## Internet2 Land Speed Record

Internet2 Land Speed Record<sup>1)</sup>は、遠距離・高速TCP通信技術の向上を目的として、2000年に創設された賞である。LSRの特徴は、評価をバンド幅ではなく、TCP通信の距離・バンド幅積で行うこと、評価が、ネットワークではなくエンドホスト・エンドホスト間であることを特徴としている。使用するネットワークは科学技術用のパブリックネットワークを区間として含み、2カ所以上ルーティングされたネットワークであり、距離はルーティングポイント間の大圏距離の和として計算し、30,000kmを上限とすることが規定されている。LSR規則の概要を付録Aに示す。

賞はIPv4、IPv6の2つのカテゴリにおいて、TCPの単一ストリームと複数ストリームの4クラスに分けられている<sup>☆1</sup>。

2000年に認定された最初のLSRは、5,626kmのネットワーク上で751Mbpsでデータ転送し、4,278Tbit m/sの記録であった(付録B)。現在私たちが保持してい

☆1 これは、複数ストリームのほうが単一ストリームより容易と考えられていたからであるが、実際には、単一ストリームのほうが容易であることが経験的に分かっている。

る記録と比較すると、バンド幅で 12 倍、距離・バンド幅積で 64 倍の性能向上が 6 年間で得られたことになる。また、現在の記録が、10Gbps ネットワーク帯域の 99 % をシングル TCP ストリームで実現していることから、LSR の目的であった高効率かつ実用的な遠距離 TCP 通信の実現は、達成されたといえる。

それでは、このような性能向上はどのように実現したのであるか。TCP 通信では、通信データへの信頼性の保証と、自律的な通信バンド幅制御を、バッファリングと輻輳ウィンドウサイズを調節することを基本的な方法として実現している。信頼性は、受信側から ACK が返ってくるまで、送信データを保持し、必要な場合に再送することにより実現される。自律的な通信バンド幅制御は、輻輳をパケットロスまたは通信遅延時間 (Round Trip Time, 以下では RTT と記す) の増加で検出し、輻輳ウィンドウサイズを調節して流量調整をすることにより実現される。通信においては、1RTT ごとに、基本的には輻輳ウィンドウサイズ分のデータを通信する。一般的な TCP プロトコルでは、パケットロスにより、輻輳ウィンドウサイズを半減させ、以後、1 回の ACK ごとに線形に輻輳ウィンドウサイズを増加させる AIMD (Additive Increase, Multiplicative Decrease) 制御により流量調整を実現している。

遠距離・高バンド幅ネットワークでは、RTT が大きいいため、単純な AIMD 制御では遠距離ネットワークが不利となり、通信バンド幅が出にくい。また、エンドホスト間の通信だけを情報として制御するため、偶発的パケットロスと輻輳によるパケットロスが区別できず、本質的でない性能低下が発生する。

これらの問題点を克服するため、数多くの改良された TCP プロトコルが提案され、過去の LSR 実験で使用された。たとえば、私たちの直接の競合相手であった、カリフォルニア工科大学の Harvey Newman のグループは、遅延時間の増加から輻輳を検出する FAST TCP<sup>2)</sup> を提案し、使用した。High-Speed TCP<sup>3)</sup>、Scalable TCP<sup>4)</sup> も、長距離ネットワーク用 TCP プロトコルとして知られている。しかしながら、本質的には AIMD 制御の係数が変わる以上のメリットを出すことは困難であり、遠距離・高バンド幅ネットワークでは不安定な振舞いを示し、高い性能を得ることができず、またバッファサイズ、キュー長などのパラメータ最適化が困難を極めた。私たちが参加する直前の状況は、FAST TCP を用いるカリフォルニア工科大学のグループとヨーロッパ・米国間の長距離ネットワークを活用するスウェーデンの SUNET のグループが激しく競い合っているという状況であった。

## LSR への挑戦

本章では私たちの Data Reservoir プロジェクトの LSR への挑戦について述べる。Data Reservoir システムは理学研究の実験・観測施設で生成される巨大データを観測実験施設と大学・研究機関との間で共有することを目的としたネットワーク基盤である。

データアクセスには TCP/IP 上の iSCSI (internet SCSI) プロトコルを用い、2 段階階層的データストライピングを行っている。近距離通信と遠距離通信を分離し、遠距離データ転送時はブロックレベルで複数 TCP ストリームによる並列転送を行うことでファイルシステム透過性を持たせ、ソフトウェアオーバーヘッドを減らすことでピーク値を持続し高効率を実現している。

私たちの LSR への出発点は、SC02 における日米通信実験であった。我々の Data Reservoir プロジェクトは 12,000km の日米間通信で 26 本の TCP/IP 並列ストリームを利用したディスク間データ転送を行い 585Mbps を実現した。これは、TCP ストリーム 1 本あたりに換算すると 23Mbps である。性能がここまで低かったのには、さまざまな原因が混在していたが、現時点での理解では、(1) 不十分な最適化、経験が不十分であったため、TCP の各種パラメータが最適な値にセットされなかったこと、(2) 多数のストリームを使ったため、並列ストリーム内の TCP ストリーム同士で性能を殺したこと、(3) その結果、非常に低速のストリームが発生し、ストリーム間のバンド幅不平衡が発生し、より多くのストリーム数が必要であったこと、(4) ディスク間転送であったことから、多くのバッファ間コピーが発生し、当時の遅い CPU が目一杯になったことと考えられる。

2002 年の反省から、私たちは、並列ストリーム間で衝突しあうことを避け、バランスした性能を出すために「レイヤ間協調最適化技術<sup>5)</sup>」を開発し、2003 年には、日米一往復半の 24,000km 上で、ソフトウェア TCP を 16 ストリーム用い TCP ストリーム 1 本あたり 438Mbps のデータ転送を行った。また、長距離通信部分に UDP を用いた TCP 加速ハードウェアを用いて、1 ストリームあたり 875Mbps を実現し、非 TCP プロトコルが超遠距離通信で TCP プロトコルを置き換える可能性を示した<sup>☆2</sup>。1 年間の最適化努力により性能を約 20 倍にすることはできたが、当時の LSR バンド幅である 5.4Gbps まではほど遠い性能であった。

私たちが、LSR に挑戦する直接的な契機は、2004 年

☆2 私たちは、富士通研究所が開発した COMET ネットワークプロセッサを用い、遠距離部分で UDP を用いた。それ以外では、UDT などが比較的広く用いられてきたが、TCP 技術の進展、UDT における CPU オーバーヘッドの高さから、TCP に対する優位性は失われつつある。

秋に実施した東京からアメリカ・カナダ・オランダを經由しスイスの欧州原子核研究機構(以下 CERN)に至る 18,500km, 17 のタイムゾーンをまたぐ WAN PHY 実験ネットワークを利用したデータ転送実験である。私たちは TOE 機能を持つインテリジェント 10GbE NIC T110 を採用し NIC 上の FPGA (Field Programmable Gate Array) を変更することでペーシング機能を TOE 機能に付加し RTT ごとのバーストを防ぎ、シングルストリームのスループット 7Gbps を達成した。

このとき WAN PHY 構築にかかわった共同研究者らから LSR への挑戦を薦められた。7Gbps のバンド幅が、20,000km を超えるネットワークで実現すれば、LSR の記録を更新できることから、同年の 11 月に実施された SC04 バンド幅チャレンジにおいて記録に挑戦した。

日本から米国を經由してオランダを接続し 20,645km 上でシングルストリーム 7.2Gbps のメモリ間転送を実現し、私たちとして最初の LSR を得た。会場のあるピッツバーグからシカゴまでは Abilene の OC-192 を、シカゴ・東京間は APAN/JGN2 の OC-192 を、そして東京・CERN 間は予備実験の WAN PHY をほぼそのまま利用し、ピッツバーグから東京・アメリカ経由スイスという約 10Gbps 31,248km, RTT433ms の経路で実験を行った。MTU (Maximum Transmission Unit) 1,500 バイトのスタンダード TCP で 7.57Gbps を達成した。

ここでのボトルネックはネットワークでも NIC でもなくホスト内 PCI-X 1.0 のバス約 8Gbps であったが、私たちはこのボトルネックの約 9 割の通信を達成した。

実は、上記記録を Internet2 に申請したところ、東京・CERN 間の WAN PHY にルータが入っていないため、日本・ヨーロッパが直線距離として距離計算され、LSR のレギュレーションにのっとると総距離が 31,248km ではなく 20,645km として算定することとなり記録は 148,850 Tbit m/s となってしまった。

一方 Caltech のグループが我々の 1 日前、すなわち SC での計測前日にマルチストリームで 26,950km 6.86Gbps 184,877 Tbit m/s という記録を出し LSR の記録を更新していたため、私たちの記録はシングルストリームの記録としてのみ認められることとなった。

30,000km を超える距離で記録更新に十分なスループットは達成されているため再実験を計画し、ネットワークの流量が 1 年間で最も少なくなるクリスマス休暇中に実験を行うこととなり、12 月 24 日から 1 月 4 日にかけて実験を行った。

図-1 にクリスマス実験のネットワークを示す。SC04 のネットワークを、アムステルダムからニューヨークへ折り返し、ニューヨーク・シカゴ間は Abilene を経

由することで東京・シアトル・シカゴ・アムステルダム・ニューヨーク・シカゴ・東京という 33,979km<sup>☆3</sup>, RTT 498ms のネットワークを構成し 7.21Gbps の記録を達成し、LSR の IPv4 single and multiple stream 記録を更新した。

これらの実験では 10GbE NIC の TOE 機能を使うことにより、Packet Pacing を実現し、ホスト CPU による限界を避けることができたが、IPv4 のみの対応であった。IPv6 のためのハードウェア手法として 10GbE を 2 ポート、FPGA を 2 個搭載したプログラマブルなネットワークブリッジ TGNLE-1 を開発、ホストとスイッチの中間でパケット間隔の平滑化およびその解析を行った。

IPv6 の TCP 通信を実現するためには、すべてのプロトコル処理をソフトウェアで実現する必要がある。しかしながら、ソフトウェア実現には、CPU 能力の限界により比較的低いバンド幅に受信能力の限界があり、パケットロスの原因となる。一方、ソフトウェアによる TCP 通信は、TOE を用いる場合と異なり、パケット流量が揺れ動き、受信限界を一時的に超過してパケットロスが発生すること、TCP 性能の上限を決定することが、シアトルで行われる SC05 の予備実験により明らかとなった。

予備実験では、東京からシアトル折り返しおよびシカゴ折り返しの日米往復・日米 2 往復となる IPv6 実験を行い、各々の経路で LSR を更新した。しかしながら、送信側に TGNLE-1 を入れて、Packet Pacing するだけでは、通信のジッタの問題が解決しないことが明らかとなった。通信の途中経路において、かなり大きく送信バンド幅が微視的に変動するからである。この問題を解決するため、新たに、受信側サーバの直前に Packet Pacing をする回路と、バッファを置き、受信レートが決して受信限界を超さないようにした(受信側 Pacing と呼ぶ)。このときは、流量が揺れ動く原因がよくつかめていなかったが、現象論的には問題が解決し、さらに 2 回、IPv6 の LSR を更新した。

IPv4 の LSR については、主要なボトルネックが PCI-X 1.0 の使用にあったため、バス的高速化が課題であった。ちょうど、PCI-X 2.0 の 10GbE NIC の販売が開始されたためシアトル市内にある Pacific Northwest Gigapop (以下 PNW Gigapop) にラックを借り、SC05 に 2 週間ほど先立ち PCI-X 2.0 対応の Windows マシン 2 台を持ち込み米国 Microsoft Research と共同実験を行った。OS として Windows Server を用い、シアトル→東京→シカゴ→アムステルダム→シアトルの 32,372km (LSR 距離で 30,000km) のネットワークを使用した。実験の結果、7.994 Gbps のバンド幅を得て LSR を更新したが、

☆3 LSR 的距離は 30,000km.

## Single stream TCP – Tokyo – Chicago – Amsterdam round trip

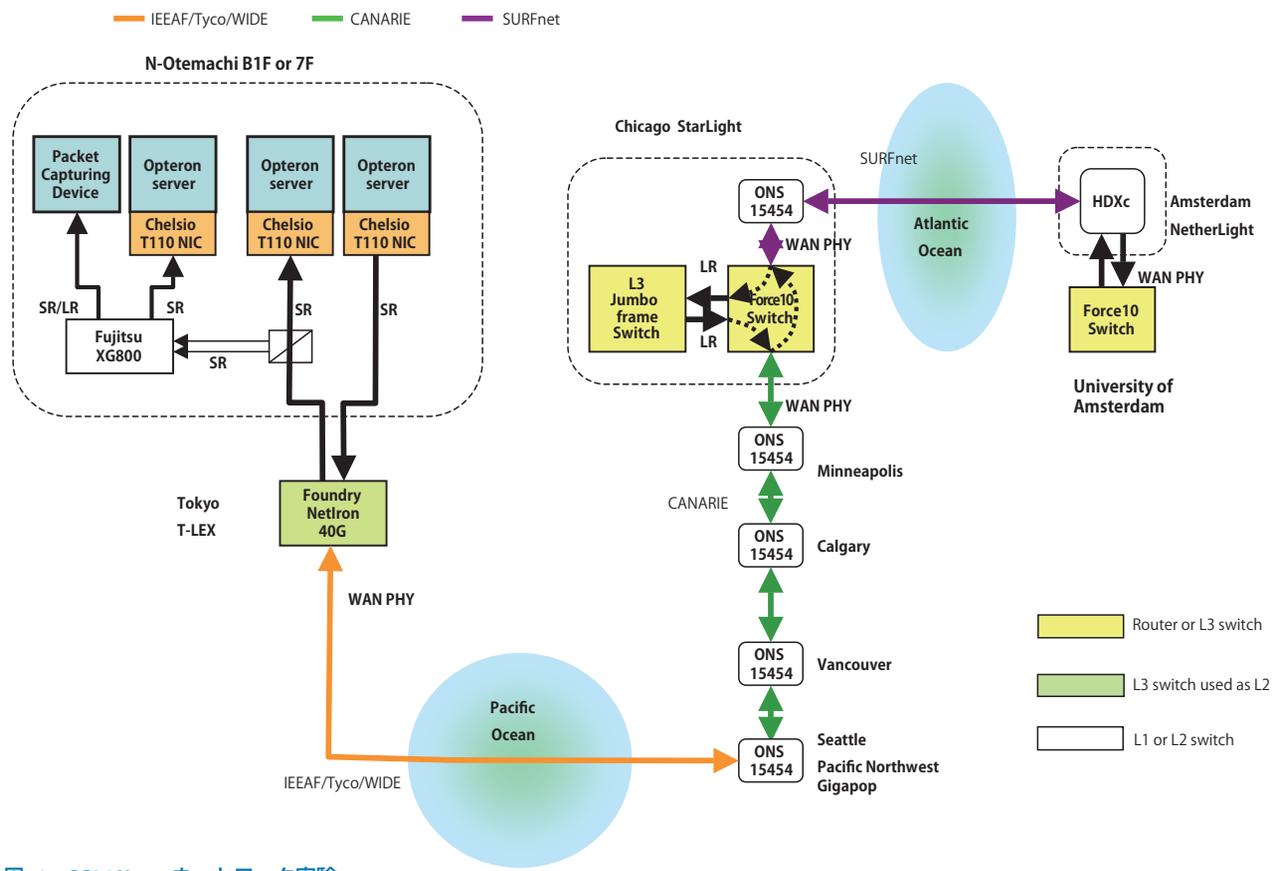


図-1 SC04 Xmas ネットワーク実験

送信側の CPU 能力が 100% になり、8Gbps を超すことは不可能であった。

SC05 の時期から、ワイヤレートで使える疑似遅延装置を使い始め、私たちが開発した TGNLE-1 を用いたパケット発生器、パケット解析器を用いて、遠距離・高バンド幅ネットワークで TCP 通信の性能が低下する原因の詳細な解析を開始した。

図-2 に SC05 のネットワーク経路図を示す。

Microsoft Research と共同実験である Windows NT を用いた実験は IPv4 single and multiple stream の LSR を更新した。

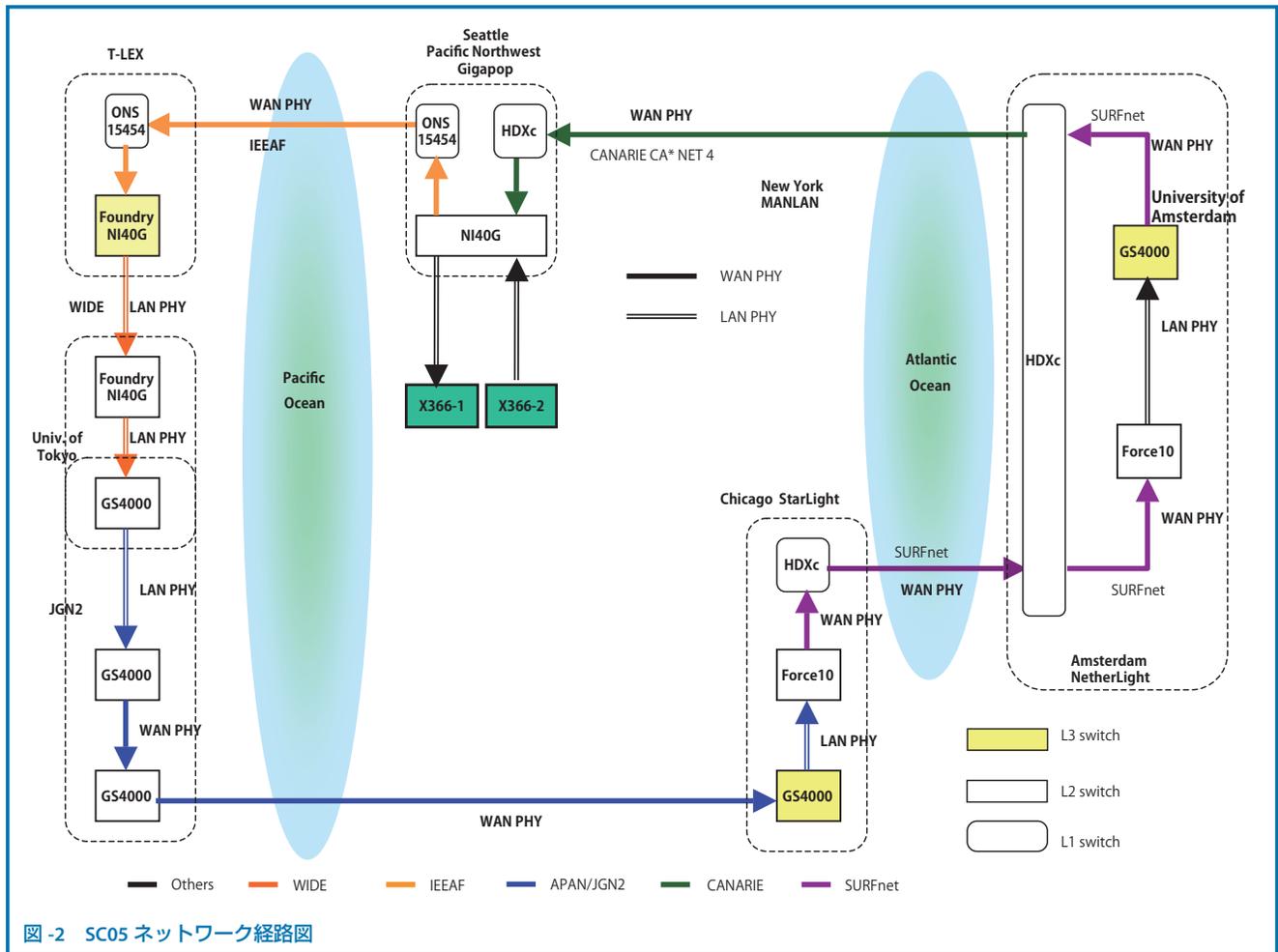
### 遠距離・高バンド幅 TCP

10Gbps ネットワークの TCP による完全利用が実現した現在の視点で遠距離・高バンド幅 TCP 通信の効率化方法を考察する。

すでに述べたように、遠距離・高バンド幅 TCP 通信性能が低下する原因は：

(1) Inflight のデータ量は、 $RTT \cdot BW$  であり、非常に大きい量となること。いったん輻輳が発生すると多量のデータが失われること、

- (2) Inflight のデータ量が大きいため、中間のスイッチ・ルータのバッファ量が不十分であり、TCP 特有のバースト的通信によりパケットロスが生じること、
- (3) TCP プロトコルを処理する CPU 資源が不足すること。とくに、受信側で不足するとパケットロスが発生すること、
- (4) End-to-End の通信プロトコルであるため、輻輳によるパケットロス、中間スイッチ・ルータのバッファサイズが不十分であるため発生するパケットロス、受信側の CPU 資源の不足によるパケットロス、ネットワーク本来のランダムなエラーに基づくパケットロスを見分けることが困難であり、不必要に通信バンド幅が低下すること、
- (5) 通信バンド幅が、微視的に見ると変動し、平均バンド幅がパケットロスを起こす最大バンド幅より小さくなること。変動原因は、中間スイッチに置けるブロッキング、WAN PHY と LAN PHY の変換時における ACK パケットのバッキングが原因であること、
- (7) I/O パスの最大バンド幅がネットワークバンド幅以下である場合に、受信側 NIC のバッファ量が不十分で、パケットロスが発生すること、
- (8) 通信に伴う送信側、受信側のホスト上のバッファメ



メモリがキャッシュサイズを超し、CPUの実行速度が著しく低下し、プロトコル処理速度が低下すること、特に、ユーザ空間からカーネル空間へのコピー操作が多大なCPUオーバーヘッドを生じること。

これらの問題点が明らかになったきっかけは、長距離ネットワーク利用で①10GbpsのNICより、1GbpsのNIC、さらには100MbpsのNICのほうが、実現するバンド幅が高いことが多いこと、②クロックサイクルレベルで正確な疑似遅延装置を用いると、適切にバッファサイズを増加させれば、ローカル通信とほぼ同じ性能が出ることに對し、実ネットワークでは性能が低下することが観察されたことである。つまり、一般的に知られているTCP通信パラメータの最適化は、本質を突いてなく、それ以外の要因が重要であるということに気がついたからである。

上記要因は、ネットワークパケットの振舞いを精密に観察することにより、初めて明らかになったものである。これらの要因を、さまざまな方法で回避することにより、安定した超遠距離・高バンド幅TCP通信を実現することに成功した。

### LSR 挑戦の裏話

ここでは、科学論文としては書けないLSR挑戦で経験したことを述べる。

最初の話題は、競争相手との戦いである。LSRのルールの制約から、私たちも、競争相手もほぼ同じネットワーク機器、NICやサーバを使うため、ソフトウェアの相違を除けば、時間との戦いの側面がある。2005年までは、性能が低下する原因が明確でなかったこともあり、常に相手より早く記録を出す必要があった。実際、私たちの最初のシングルストリーム記録の直後に、競争相手もシングルストリーム記録を出そうとしていた。その後も、数日の差で記録を守ったことが何回もあった。TCP効率化の基本的な方法が明確化された後は、時間的競争はかなり楽になった。

10Gbpsネットワーク、ネットワーク機器の不安定性は常に実験の障害であった。実験中に常に悩まされたことは、10Gbpsのネットワーク機器、特にスイッチと光機器の「疲れ」であった。LSR実験の準備期間中には、最大バンド幅の90%を超すトラフィックを、何時間も通して通信路のエラー、安定性をチェックするが、実験期間が1週間目くらいになると、なぜか突然ポートが次々

と壊れはじめ、使えるポートが減少したり、遠距離ネットワークのエラーレートがいきなり上がったたり、ひどい場合には切断したりした。この結果、スイッチのポート数は、必要数よりかなり余裕を持って準備すること、長距離ネットワークのエラーカウンタは常にチェックするということが習慣化した。

いまだに原因が分からないこととして、衆目の中でスイッチはどんどん動作不良になるのだが、実験終了後にメーカーに保守に出すと、不良が発見されないことであった。つまり、あまり 10Gbps 目いっぱい押し過ぎると、これらのデバイスが疲れて、動作を嫌がるようになっていくとしか思えない状況であった。

### おわりに

LSR への挑戦に限らず、世界一への挑戦は、新しい技術を生み出す良い機会である。4 年間にわたる TCP 性能向上は、単一の技術で実現したものではなく、TCP 通信にかかわる多くの技術開発の集大成といえるものと感じている。さらに、TCP 通信技術だけでなく、ネットワーク構築・デバッグ技術、国際的なネットワーク機関間の連携、ストリーム間協調技術からネットワークインタフェース・メーカーやスイッチ・メーカーとの技術連携など、今後のネットワーク技術研究開発の基礎を固める結果となった。

本稿では LSR への挑戦を軸に、ここ数年のネットワーク基盤の変化に触れながら超遠距離・超高速ネットワークにおける TCP 通信技術の進展について触れた。

**謝辞** LSR の達成にあたり、昼夜を問わずご協力をいただいた長谷部克幸、山本成一、関谷勇司、稲上克史、亀澤寛之、青嶋奈緒、西村亮、中村誠、来栖竜太郎、坂元眞和、古川裕希、生田祐吉、田中仁、服部佳隆、JGN2 NOC メンバ、村上満雄、根尾美由紀、小林克志に感謝します。10GbE NIC 上の実装について Chelsio Communications 特に Felix Marti 氏、Wael Nouredine 氏に感謝します。

遠距離実験を行うにあたってサポートを行っていただいた WIDE, IEEAF, APAN, JGN, SURFnet, SCinet, PNW Gigapop, SARA, UvA, FLA, CERN 諸機関に感謝します。ヨーロッパ接続について Kees Neggers, Cees de Laat, 米国接続について Don Reilly, Jan Eveleth, Bill Mar, Linda Winkler, カナダ接続について Bill St. Arnaud, Rene Hatem, Thomas Tam, Herve Guy 諸氏に感謝します。

なお、本実験は科学技術振興調整費先導的研究基盤整備「科学技術研究向け超高速ネットワーク基盤整備」の一環として研究開発が行われ、基板研究 B (2)

15300014「アプリケーショントランスペアレントな大域データインテンシブ計算機構」によって補助され、科学技術振興事業団 CREST による研究領域「情報社会を支える新しい高性能情報処理技術」研究課題「ディペンダブル情報処理基盤」の補助により実験を実施しました。

### 参考文献

- 1) <http://lsr.internet2.edu/>
- 2) Jin, C. Wei, D. X. and Low, S. H. : FAST TCP for Highspeed Long-distance Networks, Internet Draft, <http://netlab.caltech.edu/FAST/> (June 2003).
- 3) Floyd, S. : HighSpeed TCP for Large Congestion Windows, RFC 3649 (Dec. 2003).
- 4) Kelly, T. : Scalable TCP : Improving Performance in HighSpeed Wide Area Networks, First International Workshop on Protocols for Fast Long-Distance Networks (2003).
- 5) Kamezawa, H., Nakamura, M., Tamatsukuri, J., Aoshima, N., Inaba, M. and Hiraki, K. : Inter-layer Coordination for Parallel TCP Streams on Long Fat Pipe Networks, SC2004 (Nov. 2004).
- 6) Hiraki, K., Inaba, M., Tamatsukuri, J., Kurusu, R., Ikuta, Y., Koga, H. and Zinzaki, A. : Data Reservoir : Utilization of Multi-Gigabit Backbone Network for Data-Intensive Research, SC2002 (Nov. 2002). <http://www.sc-2002.org/paperpdfs/pap.pap327.pdf> (平成 19 年 12 月 26 日受付)

#### 平木 敬(正会員)

hiraki@is.s.u-tokyo.ac.jp

東京大学情報理工学系研究科教授、同大理工学系研究科物理学専門課程博士課程中退。理学博士。計算やネットワークの高速化に関する研究に興味を持つ。最近、超並列スーパーコンピュータ GRAPE-DR の研究に従事している。

#### 稲葉真理(正会員)

mary@is.s.u-tokyo.ac.jp

東京大学工学部建築学科卒業。(株)リコーに勤務した後、東京大学大学院理学系研究科博士課程中退。理学系研究科助手、講師、情報理工学系研究科特任助教授、准教授、博士(理学)。計算幾何、超高速インターネットの研究に従事。

#### 菅原 豊(正会員)

sugawara@is.s.u-tokyo.ac.jp

東京大学情報理工学系研究科助教、同大情報理工学系研究科コンピュータ科学専攻博士課程修了。博士(情報理工学)。再構成ハードウェアを用いた超高速ネットワーク処理、侵入検知システムなどの研究に従事。

#### 吉野剛史(正会員)

ysn@is.s.u-tokyo.ac.jp

東京大学大学院情報理工学系研究科コンピュータ科学専攻修士課程在学中。超高速長距離ネットワーク上での TCP/IP 通信最適化、ネットワーク上の問題分析に適した通信測定手法に関する研究に従事。

#### 玉造潤史(正会員)

junji@adm.s.u-tokyo.ac.jp

東京大学理学系研究科講師、同大理工学系研究科情報科学専攻博士課程中退。理学修士。東京大学の情報システムの管理運営に従事するとともに、超高速ネットワークに関する実験的研究に従事している。

#### 加藤 朗(正会員)

kato@wide.ad.jp

東京大学情報基盤センター准教授、東京工業大学理工学系研究科情報工学専攻博士課程中退。工学(政策・メディア)。インターネットの基礎技術、ラムダネットワークの運用技術などの研究開発に従事している。

【付録 A】 LSR の規則 (概要)

Internet2 Land Speed Record では、11 のルールを定め、記録を認定している。以下にルールの要約を示す：

• ルール 1

データ転送は、100Km 以上距離が離れ、少なくとも 2 個のルータが中間にあるネットワーク上を、10 分以上の期間実施することが必要である。また、距離の最大値は 30,000km である。

• ルール 2

データ転送を行うネットワークは、1 つ以上の平常運用されている研究教育ネットワーク上を通過することが必要である。Abilene, ESnet, CA\*net3, NREN, GEANT はその例である。

• ルール 3

データ転送は、一対の IP アドレス間を、RFC 793 と RFC 791 を満たす TCP/IP プロトコルで実施する必要がある。

• ルール 4

使用する機器、ソフトウェアは Internet2 の米国メンバが商品として入手可能か、またはオープンソースソフトウェアであることが必要である。

• ルール 5

申請には、認定に必要なすべての情報を添付することが必要である。

• ルール 6

Internet Land Speed Record には、単一ストリームの TCP/IPv4、複数ストリームの TCP/IPv4、単一ストリームの TCP/IPv6、複数ストリームの TCP/IPv6 のクラスがある。単一ストリームの記録が複数ストリームの従来記録より大きい場合には、単一ストリームは複数ストリームと解釈する。

• ルール 7

データは、ユーザ空間上のバッファ間で行われ、バッファごとにデータの内容が異なることが必要である。サムチェックなどの方法でデータが正しく伝達されたことを確認することが必要である。

• ルール 8

記録は、バンド幅と距離の積で評価する。距離は、端点と、途中のルータ間を、最短距離で結んだ距離の和とする。

• ルール 9 省略

• ルール 10

新たな記録は、それまでの記録を少なくとも 10% 凌駕することが必要である。

• ルール 11 省略

【付録 B】

過去の LSR(IPv4)

年月日	バンド幅 距離積 Tbit m/s	研究グループ	カテゴリ・クラス
2000年3月21日	4278	Microsoft, Qwest Communications, University of Washington, USC Information Sciences Institute	IPv4 Single Stream
2000年3月29日	5384	Microsoft, Qwest Communications, University of Washington, USC Information Sciences Institute	IPv4 Multiple Stream
2002年4月9日	4933	University of Alaska at Fairbanks, Faculty of Science of the University of Amsterdam, SURFnet	IPv4 Single Stream
2002年11月19日	10136	Nationaal Instituut voor Kernfysica en Hoge-Energiefysica (NIKHEF), Universiteit van Amsterdam (UvA), Stanford Linear Accelerator Center (SLAC), California Institute of Technology (Caltech)	IPv4 Single and Multiple stream
2003年2月23日	23888	California Institute of Technology (Caltech), CERN, Los Alamos National Laboratory (LANL), Stanford Linear Accelerator Center (SLAC)	IPv4 Single and Multiple stream
2003年10月1日	38420	California Institute of Technology (Caltech), CERN	IPv4 Single and Multiple stream
2003年11月11日	61752	California Institute of Technology (Caltech), CERN	IPv4 Single and Multiple stream
2004年2月22日	68431	California Institute of Technology (Caltech), CERN	IPv4 Multiple stream
2004年4月14日	69073	SUNET, Sprint	IPv4 Single stream
2004年5月6日	77699	California Institute of Technology (Caltech), CERN	IPv4 Multiple stream
2004年6月25日	104529	California Institute of Technology (Caltech), CERN	IPv4 Multiple stream

2004年6月28日	103583	California Institute of Technology (Caltech), CERN	IPv4 Single stream
2004年9月12日	124935	SUNET, Sprint	IPv4 Single and Multiple stream
2004年11月8日	184877	California Institute of Technology (Caltech), CERN, CENEC	IPv4 Multiple stream
2004年11月9日	148850	University of Tokyo, Fujitsu Computer Technologies, WIDE, Chelsio Communications, NTT Communications, CERN, SURFnet, Universiteit van Amsterdam, SCinet, CANARIE, IEEAF	IPv4 Single stream
2004年12月25日	216300	University of Tokyo, WIDE, NTT Communications, IEEAF, Chelsio Communications, CANARIE, StarLight, SURFnet, Universiteit van Amsterdam	IPv4 Single and Multiple stream
2005年11月10日	239820	University of Tokyo, Fujitsu Computer Technologies, WIDE, Microsoft, NTT Communications, JGN2, Pacific Northwest Gigapop, IEEAF, CANARIE, SURFnet StarLight, Universiteit van Amsterdam, SARA	IPv4 Single and Multiple stream
2006年2月20日	264147	University of Tokyo, WIDE, Chelsio Communications, JGN2, NTT Communications, Pacific Northwest Gigapop IEEAF, SURFnet, CANARIE, StarLight, Universiteit van Amsterdam, SARA	IPv4 Single and Multiple stream

過去のLSR(IPv6)

年月日	バンド幅 距離積 Tbit m/s	研究グループ	カテゴリ・クラス
2002年8月22日	40	Oregon Gigapop, NYSERNet, University of Oregon	IPv6 Multiple Stream
2002年9月27日	2517	ARNES, DANTE, RedIRIS	IPv6 Single Stream
2002年10月9日	5154	ARNES, DANTE, RedIRIS	IPv6 Single Stream
2003年5月6日	6947	California Institute of Technology (Caltech), CERN	IPv6 Single and Multiple stream
2003年11月18日	46156	California Institute of Technology (Caltech), CERN	IPv6 Single and Multiple stream
2005年1月19日	72225	California Institute of Technology(Caltech), CERN	IPv6 Single and Multiple stream
2005年10月28日	91838	University of Tokyo, WIDE, Chelsio Communications, NTT Communications, JGN2, IEEAF, Pacific Northwest Gigapop	IPv6 Single and Multiple stream
2005年10月29日	167400	University of Tokyo, WIDE, Chelsio Communications, NTT Communications, JGN2, IEEAF, Pacific Northwest Gigapop	IPv6 Single and Multiple stream
2005年11月12日	185400	University of Tokyo, Fujitsu Computer Technologies, WIDE, Chelsio Communications, JGN2, NTT Communications, Pacific Northwest Gigapop, IEEAF, SCinet	IPv6 Single and Multiple stream
2005年11月14日	208800	University of Tokyo, Fujitsu Computer Technologies, WIDE, Chelsio Communications, JGN2, NTT Communications, Pacific Northwest Gigapop, IEEAF, SURFnet, CANARIE, StarLight, Universiteit van Amsterdam, SARA, SCinet	IPv6 Single and Multiple stream
2006年12月30日	230100	University of Tokyo, WIDE, Chelsio Communications JGN2, NTT Communications, Pacific Northwest Gigapop, IEEAF, TransLight, StarLight, CANARIE, SURFnet, Universiteit van Amsterdam, SARA, MANLAN	IPv6 Single and Multiple stream
2006年12月31日	272400	University of Tokyo, WIDE, Chelsio Communications JGN2, NTT Communications, Pacific Northwest Gigapop, IEEAF, TransLight, StarLight, CANARIE, SURFnet, Universiteit van Amsterdam, SARA, MANLAN	IPv6 Single and Multiple stream

