

卒業論文 2012年度(平成24年度)

High Precision Performance
Measurement of Network Device

慶應義塾大学 総合政策学部

氏名：澁田 拓也

担当教員

慶應義塾大学 環境情報学部

村井 純

徳田 英幸

楠本 博之

中村 修

高汐 一紀

Rodney D. Van Meter III

植原 啓介

三次 仁

中澤 仁

武田 圭史

平成24年7月19日

High Precision Performance
Measurement of Network Device

コンピュータネットワークが発展し、多くの国と地域と世界中の人口に普及した事により、ユーザはインターネットサービスプロバイダ (ISP) 等に運用されている夥しい数のネットワーク機器を通じて世界中のコンピュータ間で通信を行うようになった。

Gigabit Ethernet 等に見られる様に、ネットワーク機器の高速化と帯域の大容量化は著しく、コンピュータのペリフェラルバスと比較出来る程度の遅延で高スループットな通信を行えるようになった。そのため、ネットワークに記憶装置を設置し、ネットワークを介してアクセスを行う iSCSI のような規格が策定された。このようなネットワークを利用した新たなアプリケーション領域は、ネットワークを介して発生する遅延による影響が大きく、高い遅延はそれらのサービスや技術の利便性を激しく損ねる事につながる。

更に近年クラウドコンピューティングサービスの分野は著しく成長したが、多数の利用者からのアクセスに耐えうるデータセンター内のネットワークの運用を行うにはネットワーク内で発生する遅延を最小に抑えなければならない。そのためサービス提供者は、運用するルータやスイッチ等のネットワーク機器の性能を計測し、ネットワークの遅延を低くする必要がある。しかし、既存のネットワーク機器の計測手法は、精度と正確さ、アクセシビリティと要求に応える柔軟性を持っておらず、サービス提供者がネットワーク機器の性能を要求に応じて計測を行えない。

本論文では、FPGA を用いた計測器が高精度且つ正確で、サービス提供者に対するアクセシビリティが高い事を校正と実験により示した。また、FPGA を用いた計測手法により実験を行った結果、家庭用ネットワーク機器が Ethernet プロトコルに基づいたフレーム転送を業務用ネットワーク機器よりも低遅延で行える場合がある事、ネットワーク機器によってバッファリング処理時の特性が異なる為、フレーム転送時に発生する遅延に大きな隔たりが存在する事、そしてネットワーク機器によっては通信する端末の接続するポートによっては同じネットワーク機器でも性能と特性が異なる事が明らかになった。これらの成果に基づき、FPGA を用いた計測器がサービス提供者の求める柔軟性を持っている事を示した。

キーワード:

1. ネットワーク計測, 2. ネットワーク機器, 3. 内部遅延, 4. ネットワーク遅延,

慶應義塾大学 総合政策学部

澁田 拓也

High Precision Performance Measurement of Network Device

As computer networks have covered most countries and population globally, users exchange data with various computers around the world, through numerous network devices managed by internet service providers.

Internet users started to exchange multimedia data, such as online gaming, voice messaging application, video streaming, remote control tool, and so on.

However these application fields are very sensitive to network latency and high latency will cause these services to be deserted by the services' users. Furthermore, with recent developments with virtual machine and virtual network came a rapid growth of cloud computing services, where the latency that occurs in its service providers' networks are critical to the evaluation of the services.

In order to ensure that the value of the data exchanged by users and service providers, is not degraded by network latency, there arises a need to measure the internal latency caused by network devices, such as router and switches. This thesis shows measured internal latency of various network devices, measured with several types of data, and clarifies their behavior according to the types of data used. The types of protocols which are measured are Ethernet IPv4, and measurements with series of data length at smallest Inter Frame Gap(IFG) were performed.

Based on the evaluation of the measurement data in this research, it was concluded that FPGA device is a very good tool for use in measurement of network devices' performance, and several tendency and characteristics of network devices were discovered. Performance of some network devices were measured with the FPGA implementation of measurement tool and some tendency and characteristics of network devices are discovered. The tendency and characteristics are that, home-use network devices performs better than some corporate-use networking devices in Ethernet switching, performance of network devices differs due to choice of ports used for connection and data transmission, and difference in buffering mechanism causes unique pattern in change of latency values of corresponding network devices and affects performance drastically.

Keywords :

1. Network Measurement, 2. Network Hardware, 3. Internal Delay, 4. Network Delay

Keio University, Faculty of Policy Management

Takuya Shibuta

目次

| | | |
|------------|--------------------------------|-----------|
| 第1章 | 序論 | 1 |
| 1.1 | 近年のネットワーク利用の動向 | 1 |
| 1.1.1 | コンピュータシステムのバスとしてのネットワーク | 1 |
| 1.1.2 | クラウドコンピューティングの普及 | 2 |
| 1.2 | 遅延により発生する問題 | 2 |
| 1.3 | 仮説 | 3 |
| 1.4 | 本論文の貢献 | 3 |
| 1.5 | 本論文の構成 | 4 |
| 第2章 | 既存の計測手法 | 5 |
| 2.1 | ネットワーク機器の性能の計測手法 | 5 |
| 2.1.1 | ネットワーク機器の性能の計測時の構成 | 5 |
| 2.1.2 | ネットワーク機器の性能の計測にあたっての問題 | 6 |
| 2.2 | ソフトウェア実装を用いたネットワーク計測 | 6 |
| 2.2.1 | コンテキストスイッチによる処理にかかる時間の揺らぎ | 6 |
| 2.2.2 | 入出力待ちによる遅延の発生 | 8 |
| 2.2.3 | ペリフェラルバスを介したI/Oでの遅延の揺らぎ | 8 |
| 2.2.4 | 非一般的なソフトウェア実装による計測 | 8 |
| 2.3 | ハードウェアを用いたネットワーク機器のパフォーマンス計測 | 9 |
| 2.3.1 | 専用ハードウェア機のアクセシビリティの問題 | 9 |
| 2.3.2 | 内部ハードウェアの柔軟性の欠如 | 9 |
| 2.3.3 | 性能の問題 | 9 |
| 2.4 | まとめ | 10 |
| 第3章 | FPGAを用いた計測手法 | 11 |
| 3.1 | FPGAを用いたネットワーク機器のパフォーマンス計測の優位性 | 11 |
| 3.2 | 本研究で用いる実装 | 11 |
| 3.2.1 | 実験環境 | 12 |
| 3.2.2 | 精度 | 12 |
| 3.2.3 | 実装物の校正 | 12 |
| 3.3 | 計測対象の限定 | 14 |
| 3.4 | 計測に用いるネットワーク通信プロトコル | 14 |
| 3.5 | 計測時の情報の送受信の詳細 | 15 |

| | | |
|-------|--------------------------------------|----|
| 第4章 | 実験結果・考察 | 16 |
| 4.1 | 計測対象 | 16 |
| 4.2 | Ethernet フレームの転送性能の違い | 16 |
| 4.2.1 | Ethernet フレームの転送性能の優劣の考察 | 17 |
| 4.3 | バッファリングの仕組みの違いによる IPv4 パケット転送時の遅延の違い | 18 |
| 4.3.1 | IPv4 でのパケット転送時の遅延の値 | 18 |
| 4.3.2 | 遅延の値から見られる 3750G-24PS-E のバッファリング傾向 | 20 |
| 4.3.3 | 遅延の値から見られる AX3630S-24T のバッファリング傾向 | 21 |
| 4.3.4 | バッファリングの挙動の分析からの考察 | 22 |
| 4.4 | ポート毎の性能の違い | 23 |
| 4.4.1 | ポート毎の遅延とフレームドロップレートの際の考察 | 23 |
| 第5章 | 結論 | 26 |
| 5.1 | まとめ | 26 |
| 5.2 | 今後の展望 | 27 |
| | 謝辞 | 28 |

目次

| | | |
|-----|--|----|
| 2.1 | 計測時のネットワーク機器と測定器の構成 | 5 |
| 2.2 | ソフトウェア実装による遅延の計測 | 7 |
| 2.3 | OSによるプリエンプション時の挙動 | 7 |
| 2.4 | PCIバスで専有が発生した際の挙動 | 8 |
| 3.1 | ネットワークケーブルを用いた実装物の校正 | 13 |
| 3.2 | 複数のケーブル長での計測による校正 | 13 |
| 4.1 | 複数のネットワーク機器の Ethernet でのフレームサイズ毎の遅延 | 17 |
| 4.2 | 3750G-24PS-E と AX3630S-24T のバッファリングによる IPv4 パケット転送時の遅延の違い | 19 |
| 4.3 | CISCO Catalyst 3750G-24PS-E のフレームサイズ 1998byte 時の IPv4 による転送時の遅延の変動 | 20 |
| 4.4 | AlaxalA 社の AX3630S-24T のフレームサイズ 2048byte 時の IPv4 による転送時の遅延の変動 | 21 |
| 4.5 | AlaxalA 社の AX3630S-24T のフレームサイズ 4096byte 時の IPv4 による転送時の遅延の変動 | 22 |
| 4.6 | LSW4-GT-8EP のポート毎の Ethernet フレームの転送に要する遅延の違い | 23 |
| 4.7 | LSW4-GT-8EP のポート毎のフレームドロップレート | 24 |

目 次

| | | |
|-----|--|----|
| 4.1 | Ethernet フレームの転送性能を一次関数に最小二乗法を用いて当てはめて表された性能 | 18 |
| 4.2 | IPv4 パケットの転送性能を一次関数に最小二乗法を用いて当てはめて表された性能 | 19 |
| 4.3 | LSW4-GT-8EP の Ethernet フレームの転送性能を一次関数に最小二乗法を用いて当てはめて表された性能 | 25 |

第1章 序論

本章では、ネットワーク機器により生じる遅延のサービスへの影響と問題に関して述べ、本研究の目的、研究内容、本論文の貢献、本論文の構成を記す。

1.1 近年のネットワーク利用の動向

近年のコンピュータの普及、そしてインターネット利用者数の増加には目覚ましいものがある。内閣府の消費動向調査によると 2010 年 3 月末には、一般世帯へのパーソナルコンピュータの普及率は 77.3%にまでのぼり、2005 年には一時低下したものの大方においてなおも普及率は上昇傾向にある [1]。また総務省によるとインターネット利用率は、2010 年度では単身世帯を含んだ全世帯におけるインターネット利用率は 93.8%に及び [2]、コンピュータやインターネットが多くの世帯で利用されていることが明らかとなっており、2010 年度に自宅からブロードバンド回線を利用している世帯が 77.9%、高速な光回線を利用する世帯は 52.2%に及ぶ。この事から近年より多くのインターネット利用者が大帯域で高速なネットワークにアクセスするようになった事が伺える。

1.1.1 コンピュータシステムのバスとしてのネットワーク

ネットワークの高速化、大容量化により、ネットワークを通じたデジタル情報の通信速度がコンピュータのペリフェラルバスと比較しうる性能を持った。それにより、ネットワークを介した補助記憶装置やペリフェラルへのアクセス、他のコンピュータのメモリ等の計算資源へのアクセスの実現を目的とした研究と技術の開発が取り組まれている。

iSCSI(internet Small Computer Interface) と FCoE(Fibre Channel over Ethernet) は、ネットワークを介した HDD 等の補助記憶装置へのアクセス手法の規格として制定された、SAN(Storage Area Network) を構成する技術である。iSCSI は、SCSI 規格を TCP/IP、FCoE の場合は FC(Fibre Channel) プロトコルに対し Ethernet フレームでカプセル化を行い、ネットワークに接続した HDD 等の補助記憶装置に対してブロック I/O による読み書きを行うことを可能にする。従来ネットワークでのファイル共有を可能にしてきた NAS(Network Attached Storage) は、読み書きをファイル操作単位になっている為、OS によるファイルへのアクセス要求の処理を行う必要があり、データの読み書きに対してファイル共有を提供する端末内部で処理に要するオーバーヘッドにより大きな遅延が発生する。一方 SAN 技術は、ファイルシステムレイヤーで動作せず、アクセスを行う端末の

ローカルディスクとして振る舞うため、低遅延および高スループットを発揮する事が可能である [3]。

また、周辺機器との通信を行う規格である USB プロトコルの IO 命令を IP パケットとしてカプセル化を行う USB/IP [4] の開発が行われており、マウスやキーボード等の周辺機器のネットワークを介して利用が可能になっている。

主記憶装置であるメモリ等の計算資源のネットワークを介した共有を行う IP バスコンピュータアーキテクチャの実現を目した研究等が行われている [5]。

このようにネットワークに接続されている様々なデバイスや計算資源をユーザの操作するコンピュータシステムの一部として扱う技術の開発と普及が著しい。その為、ネットワークを介した通信の遅延がコンピュータシステムに与える影響の割合は大きくなりつつある。

1.1.2 クラウドコンピューティングの普及

近年インターネットを介してコンピュータ、ネットワーク、ストレージ等の計算資源を自由に構成し拡張を行える仮想的なコンピュータ環境をユーザに提供するクラウドコンピューティングのサービスが急速に普及している。クラウドコンピューティングは、仮想マシン技術 [6] を用いて実現されており、仮想マシンの稼働するコンピュータを、負荷等に応じてその仮想マシンのデータをネットワークを介して他のコンピュータへ移行し、変更するマイグレーション技術等を利用している。そのため、クラウドコンピューティングを構成するネットワークは高速なネットワークでなければならない。

1.2 遅延により発生する問題

ネットワーク通信において遅延は、ユーザ体験の悪化、システムの稼働効率の低下等の様々な問題の原因となる。

Skype [7] などの音声通信アプリケーションでは、遅延による影響で音声が遅れて届く事により会話のテンポがずれて違和感が発生する事が指摘されている [8]。

iSCSI [9] や Fibre Channel over Ethernet (FCoE) [10] 等のネットワークを介した補助記憶装置へアクセスする技術の運用に当たってスループットの大幅な低下を引き起こす事 [11] 等が一般的に知られ、明らかにもされている。またネットワーク機器の内部遅延が大きく、トラフィックが集中した場合、ネットワーク機器内部のバッファにオーバーフローが発生し、パケットを破棄する事がある。そのような事態が発生した際、破棄されたパケットが TCP に基づくものであった場合、破棄されたパケットの再送信が求められるためユーザは更に大きな遅延を被る事となる。

この様に、遅延がユーザやサービス提供者に対して不利益を発生させる以上、ネットワークの遅延の排除にサービス提供者や ISP は尽力しなければならない。その為にも彼らは、運用しているネットワーク機器で発生するネットワーク機器の内部遅延を計測し、

性能を把握し適切な形でその機器を運用するとともに、より性能の高い機器を利用しなければならない。

1.3 仮説

本論文では、既存のソフトウェア実装によるネットワークの計測手法には正確さと精確さに問題があり、専用ハードウェアを用いたネットワーク機器の性能の計測手法には、アクセシビリティの問題と様々なネットワークの運用環境に則したネットワーク機器の性能の計測を行いたいというユーザの要求を満たせないという問題が存在する事を指摘する。本論文では、FPGA を用いたネットワーク機器の性能の計測手法は、高い正確さと精確さを持ち、ユーザの抱えるネットワーク機器の様々な性能を計測したいというニーズを満たせるという仮説に基づき、FPGA によるネットワーク機器の性能測定器に対して校正を行い、FPGA によるネットワーク機器による計測手法の精度と正確さを立証する。また、本研究では、既存の計測手法では明らかに出来無いネットワーク機器の特性と性質があると仮定し、

- Ethernet プロトコルによるフレームの転送では、安価で簡潔なハードウェア構成となっている家庭用ネットワーク機器の方が高価で複雑な業務用ネットワーク機器より低遅延で処理を行える場合がある
- 業務用ネットワーク機器では、バッファリング機構の挙動と性質の違いによって遅延とフレームドロップレートに大きな差異が発生し、本研究で用いるネットワーク機器を用いる事によるそのバッファに溜まったフレームの処理時の傾向を明らかに出来る
- ネットワーク機器によっては、ポート毎に機器の内部で接続している ASIC が異なり、通信に用いるポートによっては転送性能に差異が生じる

という三つの仮説に則り実験と考察を行なう。

1.4 本論文の貢献

本論文では、FPGA により実装された多くのエンジニアにとって安価で利用しやすい計測器が精度と正確さを要求されるネットワーク機器の性能計測において有意である事を実証した。また、前節で述べた仮説に則り実験を行った結果、安価な家庭用ネットワーク機器が高等な業務用ネットワーク機器よりも低遅延で Ethernet フレームの転送を行う事、通信に用いるポートによっては同じネットワーク機器でも転送性能が異なる事、そしてバッファリング機構の挙動と性質の違いによっては大きな遅延とフレームロスの変動が起こり、フレーム転送性能に大きな差異が生じるという仮説が証明された。

1.5 本論文の構成

本論文は全 5 章から構成される。第 2 章では、ネットワーク機器に対する既存の計測手法とその問題点を明らかにする。第 3 章では、本論文で計測対象に対して用いた計測手法に関して述べる。第 4 章では、第 3 章で述べた計測手法を用いて複数のネットワーク機器の性能を計測して得た結果を表し、その計測結果よりネットワーク機器の性能と特性の考察を行う。最後に第 5 章では、本論文の結論と、今後の展望を述べる。

第2章 既存の計測手法

本章では、ネットワーク機器の性能の計測手法を記し、それらの手法の特徴や問題点を明らかにする。

2.1 ネットワーク機器の性能の計測手法

特定のネットワークに発生する遅延を削減するには、そのネットワークを構成するネットワーク機器にて生じる内部遅延を把握し、より高性能な機器の導入、機器の特性に適した形で運用等を行わなければならない。そのためにもネットワーク機器の性能の計測は不可欠である。

2.1.1 ネットワーク機器の性能の計測時の構成

ネットワーク機器の性能の計測は、その機器と計測に用いる実装により図 2.1 に示されている様な構成のもと行われる。

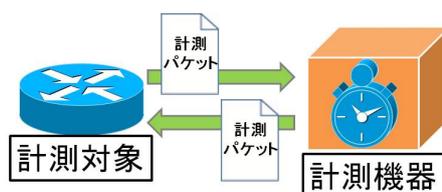


図 2.1: 計測時のネットワーク機器と測定器の構成

ネットワーク機器の性能計測機を 2 本のネットワークケーブルを用いて計測対象であるネットワーク機器に接続し、性能を計測したい通信規格を用いたフレームを計測対象へ一方のネットワークインターフェイスから送信する。計測対象であるネットワーク機器は、受信したフレームをもう一方のネットワークインターフェイスへ転送する。計測フレーム

を受信した計測器は、送信したフレームに埋め込まれていた送信時のタイムスタンプと受信時のタイムスタンプの差分を計算し受信するまでに経過した時間を取得する。そして計測器自体の内部にて発生する遅延を加味して、減算を行なって得られた値をネットワーク機器内部にて発生する内部遅延として扱う。

2.1.2 ネットワーク機器の性能の計測にあたっての問題

ネットワーク機器は、マイクロ秒単位でパケットの伝送を行う処理能力を兼ね備えている。そのためネットワーク機器の性能の計測に用いられる計測器は高精度且つ正確でなくてはならない。また、ネットワークを介した通信には、Ethernet、IPv4、IPv6 など様々な規格が存在する。また、それらの規格によってネットワーク機器が情報の伝送に必要とする処理内容が変わり、その伝送処理に要する時間も変化する。そのため、計測に用いられる機器は、高精度で正確であり、ユーザのニーズに応える柔軟性を兼ね備えていなければならない。

ネットワーク機器の性能の計測機の実装には、ソフトウェアによる実装とハードウェアによる実装が存在する。以下の節にてそれらの実装の特徴や問題点を記す。

2.2 ソフトウェア実装を用いたネットワーク計測

ネットワーク機器で発生する内部遅延の計測は、重要な課題である。しかしながら、多くの技術者が保有する汎用コンピュータ上で稼働するソフトウェアによる計測手段の実装では、精度の問題からこの問題に対する問題解決手法にはなりえない。その問題の一端を図 2.2 にて明らかにする。図に表されている物は、汎用コンピュータからネットワークを介した特定ホストとの通信において発生する遅延の計測に使われている ping[12] というツールをローカルホストに対して利用した際の出力結果だ。つまり外部環境による遅延の影響に左右されずにオペレーティングシステムやデバイスドライバを介して自機と通信した際に発生した遅延のみを表しているものだが、最大で 0.26 ミリ秒もの遅延のゆらぎが生じている事が示されている。専用ハードウェア機器を用いた 100Mbps に対応したネットワーク機器の内部遅延の計測結果が平均で 3 から 5 μ s、最大でも 9.18 μ s である事が明らかになっている [13] ため、遅延の計測結果に 260 マイクロ秒の揺らぎが発生するソフトウェア実装ではネットワーク機器の性能計測には適さない。

ソフトウェアによる遅延計測の実装がネットワーク機器の内部遅延に適さない原因は複数存在する。

2.2.1 コンテキストスイッチによる処理にかかる時間の揺らぎ

ソフトウェアによる計測器の実装は、その実装が動作する下のレイヤーの OS により生じる揺らぎに影響される事に繋がる。OS の代表的な動作としてコンテキストスイッチが挙げられる。現代のコンピュータのアーキテクチャでは、複数のプロセスが 1 つか複数の

```

File Edit View Search Terminal Help
[bhangra@bhangra tmp]$ ping localhost
PING localhost (127.0.0.1) 56(84) bytes of data.
64 bytes from localhost (127.0.0.1): icmp_req=1 ttl=64 time=0.055 ms
64 bytes from localhost (127.0.0.1): icmp_req=2 ttl=64 time=0.057 ms
64 bytes from localhost (127.0.0.1): icmp_req=3 ttl=64 time=0.057 ms
64 bytes from localhost (127.0.0.1): icmp_req=4 ttl=64 time=0.060 ms
64 bytes from localhost (127.0.0.1): icmp_req=5 ttl=64 time=0.065 ms
64 bytes from localhost (127.0.0.1): icmp_req=6 ttl=64 time=0.057 ms
64 bytes from localhost (127.0.0.1): icmp_req=7 ttl=64 time=0.043 ms
64 bytes from localhost (127.0.0.1): icmp_req=8 ttl=64 time=0.053 ms
64 bytes from localhost (127.0.0.1): icmp_req=9 ttl=64 time=0.051 ms
64 bytes from localhost (127.0.0.1): icmp_req=10 ttl=64 time=0.044 ms
64 bytes from localhost (127.0.0.1): icmp_req=11 ttl=64 time=0.069 ms
64 bytes from localhost (127.0.0.1): icmp_req=12 ttl=64 time=0.061 ms
64 bytes from localhost (127.0.0.1): icmp_req=13 ttl=64 time=0.056 ms
64 bytes from localhost (127.0.0.1): icmp_req=14 ttl=64 time=0.055 ms
64 bytes from localhost (127.0.0.1): icmp_req=15 ttl=64 time=0.048 ms
64 bytes from localhost (127.0.0.1): icmp_req=16 ttl=64 time=0.048 ms
^C

```

図 2.2: ソフトウェア実装による遅延の計測

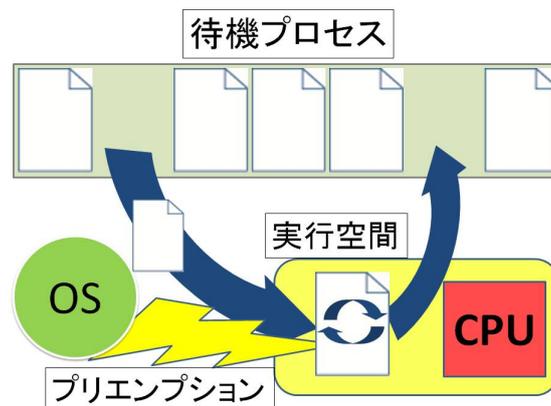


図 2.3: OS によるプリエンプション時の挙動

CPU を共有し、それらのプロセスの処理を実行している状態を停止して保存したり、再実行する事により、プロセスが CPU を利用するマルチタスクオペレーティングシステムが一般的である。そのためソフトウェア実装による計測手法は、計測に用いるパケットの送信時受信時のタイムスタンプ作成時までの不定な処理時間の揺らぎが発生する。このような処理時間の揺らぎが、ネットワーク機器の性能の計測に悪影響を起こすのは明らかである。

2.2.2 入出力待ちによる遅延の発生

ソフトウェア実装の動作する環境である OS 上でネットワーク機器との通信を行うプロセスには、計測の為の実装以外にも、ネットワークに関する設定を行うデーモンやネットワークアプリケーションなども存在する。それらのプロセスが計測対象であるネットワーク機器に対してパケットを送受信すれば NIC 等のハードウェアとネットワーク機器の帯域の専有を行う事により、計測用のパケットの送受信が滞り、ソフトウェア実装による計測結果に揺らぎが発生する。このような入出力待ちに起因する計測結果への揺らぎもソフトウェア実装の問題点である。

2.2.3 ペリフェラルバスを介した I/O での遅延の揺らぎ

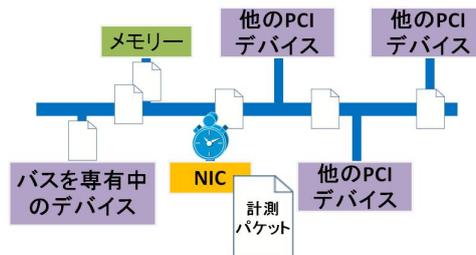


図 2.4: PCIバスで専有が発生した際の挙動

ソフトウェアで遅延計測の実装を行うという事は、そのソフトウェアが動作するハードウェアで発生する不確定要素をも内包してしまうという事である。ネットワークにて情報を送受信する時、ペリフェラルバスを介して情報伝達する際に遅延の揺らぎが発生する。ソフトウェア実装では、他のハードウェアがバスを専有して通信を行っていた際に、ネットワークインターフェースカード (NIC) 等と情報の送受信を行う必要性があっても他のハードウェアの通信が終了するのを待つ必要がある。すなわち計測の為のパケット等を送信する時のタイムスタンプを発行した後に一定ではない待ち時間が発生するという事である。このような計測結果に悪影響を及ぼす遅延の揺らぎは、マイクロ秒単位での計測を必要とするネットワーク機器の性能の計測には望ましくないものである。

2.2.4 非一般的なソフトウェア実装による計測

Orosz らによる研究では、既存の libpcap ではマイクロ秒単位でのタイムスタンプの取得しか行えなかったものに拡張を加え、ナノ秒単位のタイムスタンプの取得を可能にした

ものを用いてソフトウェアによる計測の精度の評価を行なっている [14]。この研究では、計測結果が OS によるプリエンブション、すなわちアプリケーションやソフトウェア実装の動作を OS のスケジューラにより割り込みされないように一般的ではないノンプリエンブティブな Linux カーネルを用いている。しかしながら結果としては、既存のマイクロ秒単位での計測しか行えなかった通常のソフトウェア実装よりも高い精密さを示したものの、タイムスタンプの生成に要する処理時間の揺らぎが計測結果に有意な変動をもたらしており、ソフトウェア実装による計測の正確さと精密さの限界を示している。

2.3 ハードウェアを用いたネットワーク機器のパフォーマンス計測

SPIRENT Communications 社のハードウェアによるネットワーク機器の性能計測器は、精度と正確さの高さからネットワーク機器の性能の計測手法のデファクトスタンダードとなっており、ネットワーク機器の製造販売を行う企業や ISP 等のネットワークの運用に携わる企業などで活用されている [15]。しかしながら、ハードウェアによる実装は、アクセシビリティと柔軟性の観点から問題を抱えている。

2.3.1 専用ハードウェア機のアクセシビリティの問題

SPIRENT Communications 社等のネットワーク機器の性能計測器は、高価で多くのネットワーク機器の運用構築に携わる人物にとってアクセシビリティが悪く、ネットワーク機器のカタログスペックと実機の性能の比較を行なったり、特定の運用用途での性能を計測するのが困難である。

2.3.2 内部ハードウェアの柔軟性の欠如

また、専用ハードウェアとして実装されたネットワーク機器の性能計測器は、コンソールから IFG 値やフレームサイズの設定を変更出来るものの、内部回路に ASIC を用いられていたり、ユーザによる改造を想定していないため、内部の論理回路などを変更し、遅延、スループット、フレームロスレート等の計測手法や精度等にユーザの手で変更をくわえる事が不可能である。

2.3.3 性能の問題

前節で述べたソフトウェア実装によるネットワーク計測手法と比較してハードウェアによる実装では高い精度と正確さで計測を行えるものの、その性能にも限りがある。計測に用いる事が出来るジャンボフレームは、10,000byte までしか対応しておらず、計測出来る遅延の値も ± 100 ナノ秒の誤差で 100 ナノ秒の精度までとなっており、精度と正確さに

限りがある [16]。本研究で計測対象とした機器では、BUFFALO 社の LSW4-GT-8EP は、16,000byte のジャンボフレームまで転送可能であり、SPIRENT 社の機器では 16,000byte での性能を計測する事は不可能である。

2.4 まとめ

本章では、ソフトウェアによる実装とハードウェアによるネットワーク機器の性能計測機の実装の問題点を明らかにした。次章にて本研究で用いる FPGA によるネットワーク機器の性能測定器の優位性を示し、具体的な計測器のハードウェア構成と校正を行なって明らかとなった本研究で用いる実装物の精度と正確さを示し、実際にネットワーク機器に対して実施した計測手法、計測対象であるネットワーク機器、計測に用いた通信規格を記す。

第3章 FPGAを用いた計測手法

本章では、本研究で用いる FPGA によるネットワーク機器の計測の優位性の実証、並びにネットワーク機器の計測の実験に用いる実装と計測手法の詳細に関して述べる。

3.1 FPGAを用いたネットワーク機器のパフォーマンス計測の優位性

2章では、安価で多くの技術者に手の届くソフトウェアによる実装と高価なハードウェアによる実装の問題点を示した。本研究で着目したのは、ハードウェアの論理構成の変更が可能な FPGA(Field Programmable Gateway Array) による性能計測器の実装である。FPGA を用いた実装では、HDL(Hardware Description Language) 言語で記述された論理構成をコンパイラと CAD ツールにより論理合成を行い、FPGA ハードウェアの挙動を、工場で製造し出荷された後もユーザの手により可能である。そのため、ソフトウェア実装の仕様が変更可能であるというよい特性を持ちつつも、精確さと正確さを求められるネットワーク機器の計測時に割り込みや不確定な処理時間に害されず、プリント基板に実装されたクロックのサイクルに合わせて処理を行えるというハードウェアの優位性を兼ね備えている。この事から高速で高精度、且つ正確さが要求されるネットワーク機器の性能計測器には FPGA を用いた実装に優位性があると仮定される。以下の節にて本研究で利用する FPGA によるネットワーク機器の性能計測器の校正手法と結果を示し、FPGA を用いたネットワーク機器の性能計測器の優位性を検証する。

3.2 本研究で用いる実装

本研究では、慶應義塾大学村井研究室に所属する松谷健史氏によるネットワーク機器の性能計測機の実装を用いる。本節では本研究で用いる計測器の実装の概要を説明し、複数のケーブルを用いて実装の校正を行なった結果を示し、FPGA を用いたネットワーク機器の性能計測器の優位性を検証する。

本研究で用いる計測器は、Ethernet フレームが転送される際に経過した遅延、受信した秒間フレーム数等が取得出来る。本研究で用いる計測器による遅延の計測は、一方の Ethernet ポートから送信時のタイムスタンプを埋めこまれた Ethernet フレームを送信する。その後もう一方の Ethernet ポートにて受信した際に、Ethernet フレームに埋めこまれたタイムスタンプと受信時のタイムスタンプの差分から、計測に用いられた Ethernet

フレームが転送されるのに要した遅延を取得する。計測して遅延の値は、PCIバス経由で計測結果の閲覧に用いる PC のメモリに記録される。計測に用いる Ethernet フレームのフレームサイズ、送信時のフレーム間のアイドル期間である IFG(Inter Frame Gap)、ネットワーク機器が転送にどのプロトコルを利用するのか等の設定は PC のレジスタへ記入し、PCIバスから計測器が動作する FPGA ボードがそれらの情報を取得し、反映される。

計測器による受信した秒間フレーム数の計測は、計測フレームの受信に用いられる Ethernet ポートで受信した Ethernet フレームの総数を 1 秒おきに PCI バスから PC のメモリに記録を行う。そのため、送信したフレームの内、どの Ethernet フレームが計測対象のネットワーク機器により破棄されたか、受信した Ethernet フレームの内いくつかは計測器により送信されたものでないかは不明である。

どの計測に用いられる Ethernet フレームがドロップされたか、受信したフレームの内いくつかは計測器に発されていないかを記録出来るように実装の変更を行う事は、今後の課題とする。

3.2.1 実験環境

本研究で用いる実験環境は、スタンフォード大学の研究グループによりリリースされた NetFPGA 1G[17] と NetFPGA を PCI 接続により搭載した IBM 社の ThinkCentre S50(8183-64J) により構成される [18]。NetFPGA は、Xilinx 社の FPGA である Virtex-II Pro 50 と Spartan-II、4 つの 1000base-T インターフェース、125MHz クロック、PCI 端子、SRAM 等のハードウェアを備えた FPGA ボードである。Thinkcentre S50 のハードウェア構成は、CPU は Intel 社の Celeron D 325 2.53GHz で、メモリは 512MB 搭載している。本研究で用いた計測器の実装は、NetFPGA に搭載されている FPGA である Virtex-II Pro 50 と Spartan-II を JTAG ケーブルを用いて HDL で記述された論理回路から生成されたバイナリファイルで再構成する事により動作する。

3.2.2 精度

理論的に NetFPGA に施された本研究で用いる計測器は、NetFPGA に搭載された 125MHz のクロックに基づき、8ns を分解能とした高精度な計測が可能である。次節にて実装物の校正を行なった結果を示し、計測器の精密さと正確さを明らかにする。

3.2.3 実装物の校正

本節では、本研究で用いるネットワーク機器の性能計測器の実装の校正をどのように行なったか、得られた結果、そして考察を記す。校正の過程において、計測器を図 3.1 の様に複数の異なるケーブルで接続し、逐次計測を行い、得られた値の分布から精度と正確さを検証し、実装物内部でどれほどの遅延が発生し、Ethernet ケーブルを Ethernet フレー

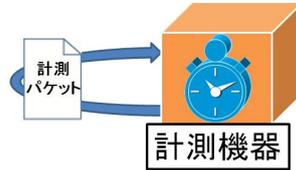


図 3.1: ネットワークケーブルを用いた実装物の校正

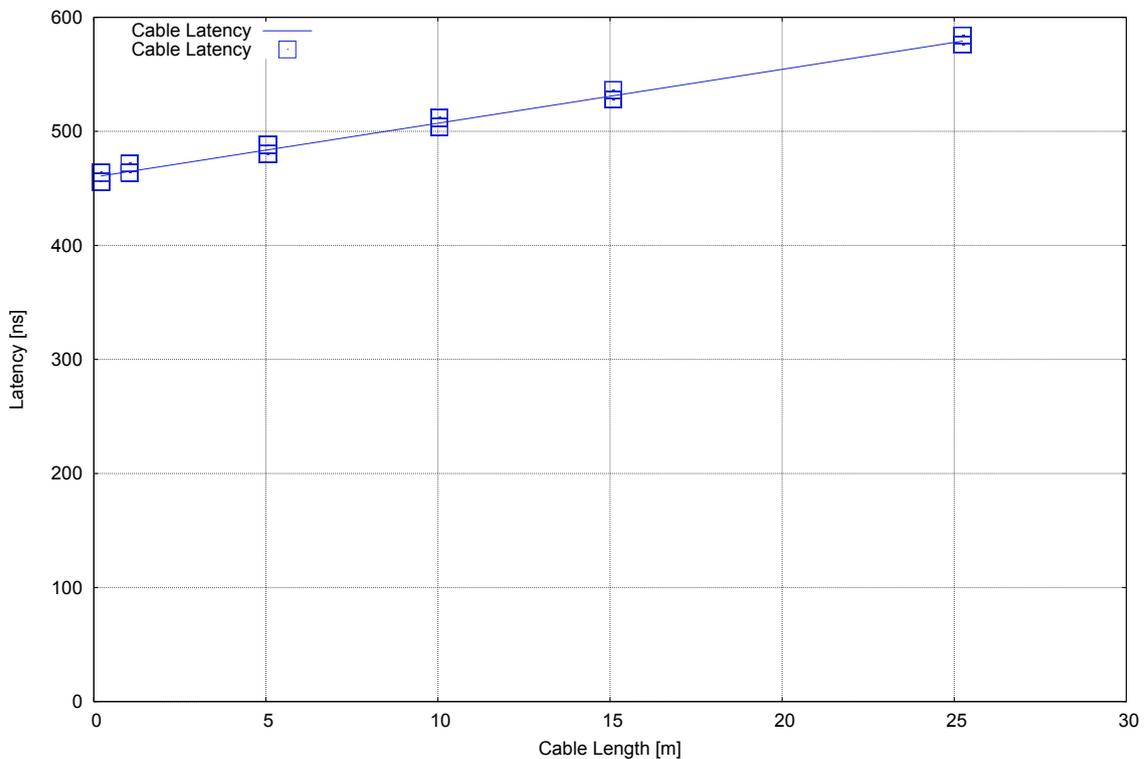


図 3.2: 複数のケーブル長での計測による校正

ムというデジタル情報がどれほどの速度で伝送されるかを確認した。計測して得られた遅延の数値をグラフに記載したものを図 3.2 に示す。

図 3.2 には、実装した計測器に対して計測に用いる 2 つの Ethernet インターフェースに対して複数の長さ (20 センチメートル、1.04 メートル、5.05 メートル、10.04 メートル、15.09 メートル、25.25 メートル) の UTP ケーブルで接続して計測を行い、計測時のケー

ブルの長さをメートル単位で横軸に、計測結果の遅延をナノ秒単位で縦軸に表したグラフである。青の点で表示されている値がそれぞれのケーブル長での遅延の計測結果であり、計測結果から最小二乗法を用いて一次関数の最良の当てはめ線を描画したものが青い線である。計測結果から最小二乗法を用いて得られた、ケーブル長に応じた遅延の増加を表す一次関数の方程式は、以下に示す。

$$t(\mathcal{L})ns = 4.72284 \frac{ns}{m} \times \mathcal{L}m + 459.913ns.$$

そして UTP ケーブルで情報が伝達される速度が

$$211737006 \frac{m}{s}.$$

で、真空での光速の 71%程であるとの数値が導きだされ、真空での光速に対する UTP ケーブル上で信号が伝達する速度の比率としてワイエスソリューションズが示す 66%という数値に近い値を確認された [19]。

この校正で得られた結果から、本研究で用いる FPGA に実装された計測器の精度と正確さが高く、ネットワーク機器の性能の計測に用いると判明した。また、計測時に計測器内部で 459.913 ナノ秒程の遅延が発生し、計測対象と接続に用いるケーブルにより 1 メートル毎に 4.722843 ナノ秒の遅延が発生するため、以降ネットワーク機器に対して計測を行い得られた遅延のデータからはそれらの値から導き出される計測対象のネットワーク機器に起因しない遅延の値を削減したものを利用して計測対象の性能を示す。

3.3 計測対象の限定

本研究での計測対象は、現在高速で多くの企業や一般家庭でも利用されている、1000base-T に対応したネットワーク機器に限る [20]。1000base-T は、IEEE により 802.3ab として制定された Ethernet 通信の規格である。1000base-T では、1 ギガビット毎秒で信号を送受信を行う。そのため高速で大帯域なネットワークの構築が 1000base-T 対応のネットワーク機器により可能である。

3.4 計測に用いるネットワーク通信プロトコル

本研究で計測を行なったネットワーク機器が転送に用いる通信規格は、Ethernet と IPv4 である。計測対象のネットワーク機器の Ethernet を利用した通信での転送性能の計測には、計測器の受信に用いられるポートに割り当てられた MAC アドレスを宛先アドレスに設定した Ethernet フレームヘッダで IPv4 パケットをカプセル化したものを用いた。計測対象に対する IPv4 での通信の転送性能の計測では、計測対象の計測に用いるポートにコンソールから IP アドレスの割り当てを行い、ARP 解決により計測対象のポートの MAC アドレスを取得する。その後、宛先 IP アドレスを計測器の受信ポートに宛てたものを、計測対象で受信に用いるポートの MAC アドレスを宛先に設定した Ethernet ヘッダでカプセル化を行ったフレームで計測を行う。

3.5 計測時の情報の送受信の詳細

ネットワーク機器の性能の計測時は、コンソールである PC から、転送に用いられるプロトコル、計測に用いる情報の一通信単位である Ethernet フレームのサイズ、Ethernet フレーム間に必要とされる回線のアイドル期間の Inter Frame Gap(IFG) の長さ等の設定を PCI バスを通じて行う。基本的に全ての計測シーケンスは、先ず計測対象の Ethernet が IPv4 転送性能を図るかを選択し、最短 IFG である 96 ナノ秒おきに計測の為のフレームを連続で送信し続け、100 秒おきに計測に用いるフレームサイズの変更を行う。そしてその間 1 秒おきに直近に受信したフレームが計測対象を介して送受信されるのに要した時間、1 秒間に計測器が受信した総フレーム数とペイロードをコンソールである PC で記録する。

その為、ネットワーク機器によっては、最短 IFG で受信され続けるフレーム数に転送処理が追いつかず、バッファにフレームが溜まった状態で別のフレームサイズでのデータの収集に移行し、記録した遅延と受信フレーム数に影響が見られたケースが存在した。適切なフレームドロップ数の計測を行えなかったため、本研究では、フレームドロップレートをフレームサイズ毎にサイズ、IFG 値、プリアンブル部を元に割り出された、そのサイズで転送可能な最大秒間フレーム数と比較して受信したフレーム数の少なさを % 表記した値を用いて表した。計測器の発したフレームに対する正確なフレームドロップ率の記録は、今後の課題とする。

第4章 実験結果・考察

本研究では、複数の仮説に基づき複数のネットワーク機器に対して3章に記載した計測手法で実験を行い、実験結果から考察を行ない仮説の検証を行なった。本研究で実験と検証を行なった仮説は、

- Ethernet プロトコルによるフレームの転送では、安価で簡潔なハードウェア構成となっている家庭用ネットワーク機器の方が高価で複雑な業務用ネットワーク機器より低遅延で処理を行える場合がある
- 業務用ネットワーク機器では、バッファリング機構の挙動と性質の違いによって遅延とフレームドロップレートに大きな差異が発生し、本研究で用いる計測器によりそのバッファに溜まったフレームの処理時の傾向を明らかに出来る
- ネットワーク機器によっては、ポート毎に機器の内部で接続している ASIC が異なり、通信に用いるポートによっては転送性能に差異が生じる

の三つであり、本研究では、実験からこれらの仮説を証明する結果が得られた。

4.1 計測対象

本研究で計測を行なったネットワーク機器は以下のとおりである。業務用ネットワークを構成するのに用いられているネットワークの機器のサンプルとして、Cisco 社の Catalyst2960G-8TC-L[21]、Calayst3750G -24PS-E[22] と Catalyst3750E-24PD、Alaxala 社の AX3630S-24T [23] に対して計測を行なった。また家庭用ネットワーク機器のサンプルとして BUFFALO 社のギガビットスイッチングハブである LSW4-GT-8EP[24] と LSW3-GT-5EP [25]、そして Allied Telesis 社の CentreCOM GS908E [26] に対しても計測をおこなった。

4.2 Ethernet フレームの転送性能の違い

本節では、業務用ネットワークの構築運用に用いられる CISCO 社の Catalyst2960G -8TC-L、Catalyst3750E-24PD、AlaxalA 社の AX3630S-24T、家庭用ネットワーク機器として用いられる安価な BUFFALO 社の LSW4-GT-8EP、LSW3-GT-5EP、そして Allied Telesis 社の CentreCOM GS908E の Ethernet でのフレーム転送性能の差異をグラフに表し、家庭用ネットワーク機器の方が低い遅延でフレームの転送を行えるという仮説の検証

を行う。以下にこれらのネットワーク機器の Ethernet フレームの転送性能をフレームサイズ毎に遅延を計測して得られたグラフを図 4.1 に示す。

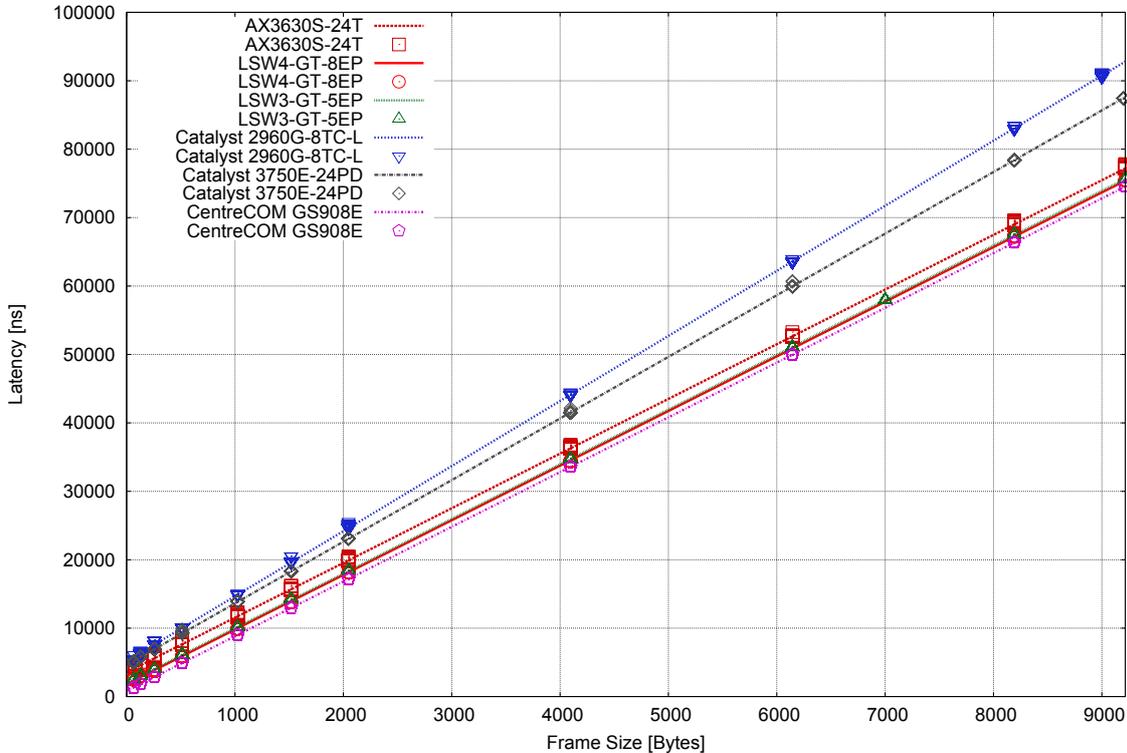


図 4.1: 複数のネットワーク機器の Ethernet でのフレームサイズ毎の遅延

4.2.1 Ethernet フレームの転送性能の優劣の考察

図 4.1 の点は、それぞれのフレームサイズ毎に百回計測して得られた遅延の値を示したものだ。そしてこのグラフに描かれた線は、それらの値を元に最小二乗法を用いてフレームサイズと遅延が比例する一次関数を導き出し、グラフに描画したものである。グラフの横軸はフレームサイズを byte 単位で示しており、縦軸はフレームが転送された際の遅延を ns 単位で示している。つまりグラフの下部に特定の機体での遅延を計測して得られた値を示した点とそれを元に描かれた最良の当てはめ線があればその機体はフレームに転送する際に発生する遅延が低く、Ethernet フレームの転送性能が高いと結論づける事が出来る。表 4.1 には、計測対象毎に得られた計測結果を元に最小二乗法から導きだされたフレームサイズの増加に伴う遅延の増加傾向を示す一次関数の傾きと y-切片、そしてそれらの値の計測結果に対する漸近的標準誤差が記されている。計測結果を元に得られた遅延の増加傾向を分析した結果、Ethernet プロトコルによるフレーム転送において、本研究での計測対象では家庭用ネットワーク機器の方が業務用ネットワーク機器より低遅延でフレーム転送を行なっている事が判明した。

表 4.1: Ethernet フレームの転送性能を一次関数に最小二乗法を用いて当てはめて表された性能

| \ 計測対象 項目 \ | AX3630S -24T | LSW4-GT -8EP | LSW3-GT -5EP | C2960G -8TC-L | C3750E -24PD | CentreCOM GS908E |
|-------------------------------------|-----------------|-----------------|-----------------|------------------|-----------------|---------------------|
| メーカー | AlaxalA | BUFFALO | BUFFALO | CISCO | CISCO | Allied Telesis |
| 販売開始 | 2005 年 | 2009 年 | 2007 年 | 2007 年 | 2007 年 | 2006 年 |
| 用途 | 業務用 | 家庭用 | 家庭用 | 業務用 | 業務用 | 家庭用 |
| フレーム サイズ毎の 遅延の増加 [ns/byte] | 7.99 | 7.98 | 7.99 | 9.51 | 9.01 | 8.00 |
| 転送に要 する遅延 の定数 [ns] | 3,556 | 1,810 | 1,998 | 5,155 | 4,624 | 811 |
| 最大フレーム サイズ [byte] | 9,216 | 16,000 | 16,367 | 9,000 | 9198 | 9887 |
| 傾斜に対する 漸近的標準誤差 [ns/byte] | 0.001006 | 0.001139 | 0.0003565 | 0.001563 | 0.0008529 | 0.00002949 |
| 傾斜に対する 漸近的標準誤差 [%] | 0.01258 | 0.01426 | 0.00446 | 0.01643 | 0.009172 | 0.0003687 |
| 遅延の定数に 対する漸近的 標準誤差 [ns] | 4.44 | 4.787 | 1.748 | 6.838 | 3.644 | 0.1302 |
| 遅延の定数に 対する漸近的 標準誤差 [%] | 0.1249 | 0.2645 | 0.08745 | 0.1327 | 0.07881 | 0.01605 |

4.3 バッファリングの仕組みの違いによる IPv4 パケット 転送時の遅延の違い

本節では、業務用ネットワークの構築運用に用いられる CISCO 社の Catalyst3750G - 24PS-E と AlaxalA 社の AX3630S-24T のレイヤー 3 プロトコルの IPv4 でのパケット転送性能の差異をグラフに表し、機器ごとにバッファリングの行い方の違いによりパケットに発生する遅延が変わりうるという仮説の検証を行う。これらのネットワーク機器の IPv4 によるパケットの転送性能をフレームサイズを変更しながら遅延を計測して得られたグラフを図 4.2 に示す。

4.3.1 IPv4 でのパケット 転送時の遅延の値

図 4.1 の点は、それぞれのフレームサイズ毎に百回一秒おきに遅延を計測し、64byte から順にフレームサイズを大きくしながら計測を続けて得られた遅延の値を羅列したものだ。そしてこのグラフに描かれた線は、それらの値を元に一次関数の最良の当てはめ線を

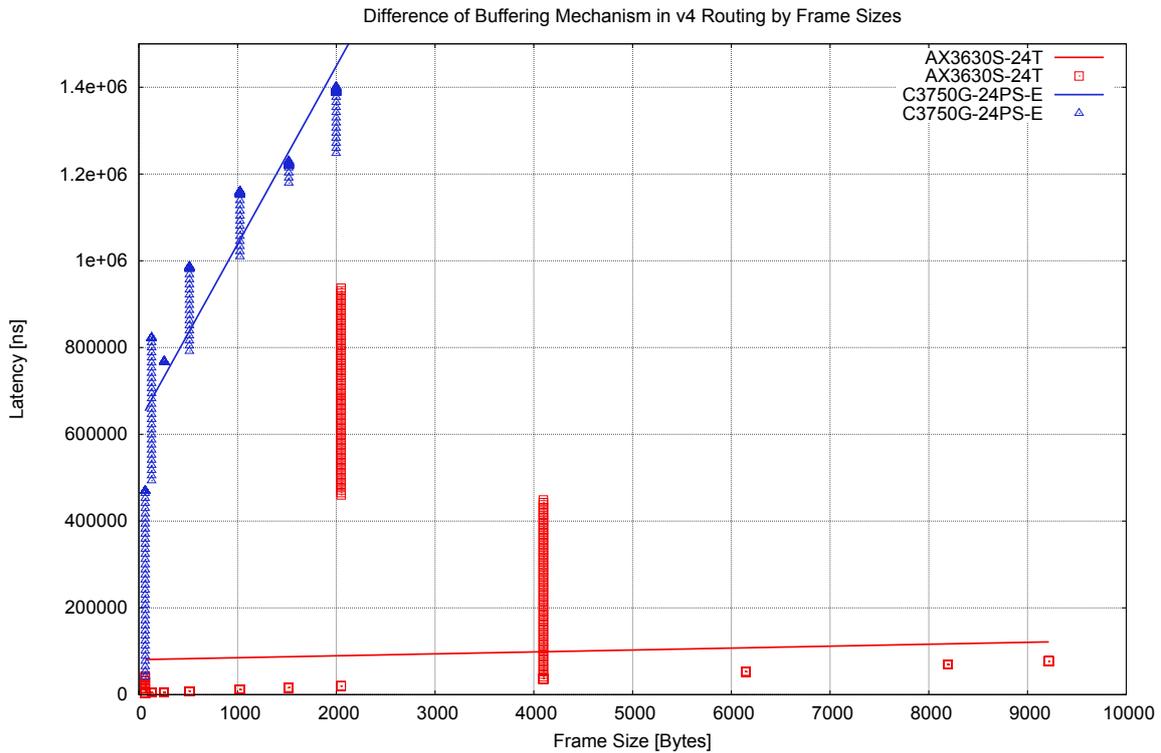


図 4.2: 3750G-24PS-E と AX3630S-24T のバッファリングによる IPv4 パケット転送時の遅延の違い

表 4.2: IPv4 パケットの転送性能を一次関数に最小二乗法を用いて当てはめて表された性能

| ＼計測対象 項目＼ | AX3630S -24T | C3750G -24PS-E |
|-------------------------------------|----------------------|-------------------|
| フレーム サイズ毎の 遅延の増加 [ns/byte] | 4.4346 | 410.711 |
| 転送に要 する遅延 の定数 [ns] | 80,616.1 | 627,770 |
| 最大フレーム サイズ [byte] | 9,216 | 1998 |
| 傾斜に対する 漸近的標準誤差 [ns/byte](%) | 1.783ns/byte(40.21%) | 7.727ns(1.881%) |
| 遅延の定数に 対する漸近的 標準誤差 [ns](%) | 7874ns(9.67%) | 8101(1.29%) |

描画したものである。グラフの横軸はフレームサイズを byte 単位で示しており、縦軸はフレームが転送された際の遅延を ns 単位で示している。つまり、先ほどの図 4.1 と同じように、いずれかの機体の計測データを示す点と線で低いほうが IPv4 プロトコルによるパケット転送性能が高いと結論づける事が出来る。全体的にパケット転送に要する遅延は AX3630S-24T の方が低い為、パケット転送性能は、3750G-24PS-E よりも AX3630S-24T の方が高いと結論づける事が出来る。

機体別に見ると、Catalyst3750G-24PS-E は、計測開始時より急速に遅延が上昇しており、その上昇傾向はフレームサイズが変更されてもほぼ継続している。その一方で、AX3630S-24T は、フレームサイズが 2048byte と 4098byte での計測時に計測される遅延の値が大きく変動している事が伺える。以下の節では、機体別にそのバッファリングの傾向の詳細を確認する。

4.3.2 遅延の値から見られる 3750G-24PS-E のバッファリング傾向

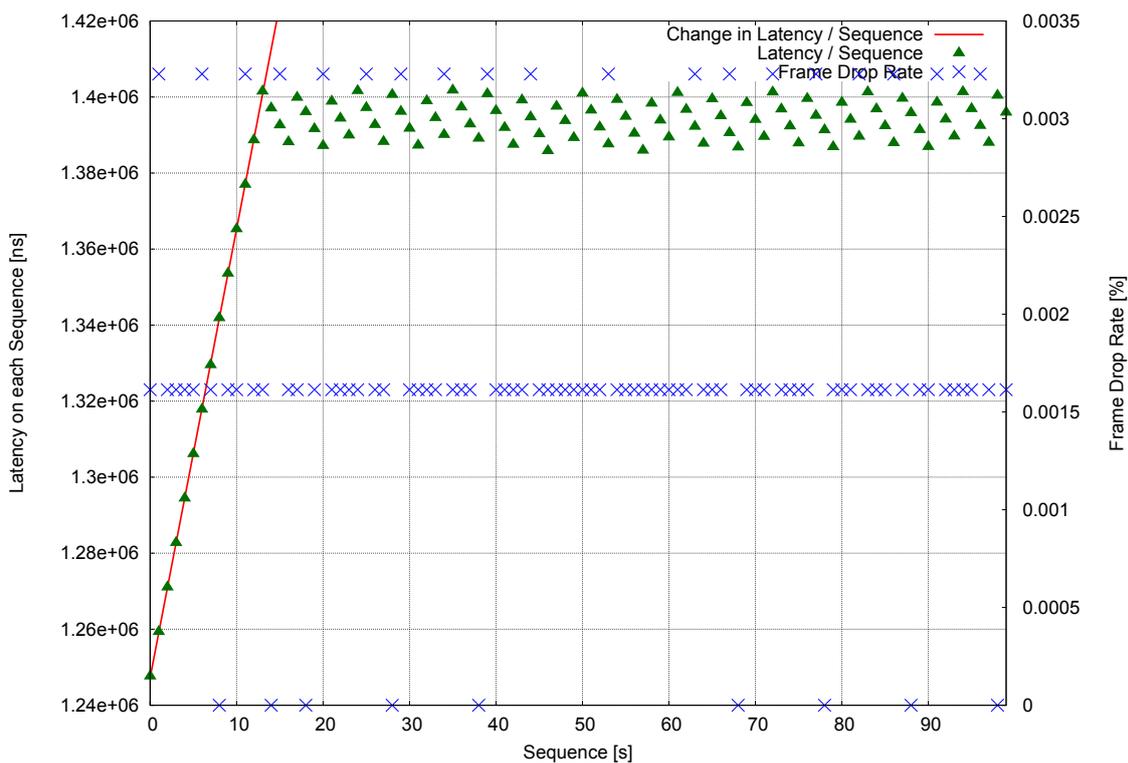


図 4.3: CISCO Catalyst 3750G-24PS-E のフレームサイズ 1998byte 時の IPv4 による転送時の遅延の変動

図 4.3 は、複数のフレームサイズを用いた Catalyst 3750G-24PS-E の性能計測時の内、フレームサイズ 1998byte 時に得られた遅延の値とフレームドロップレートをグラフに示したものだ。グラフの緑の点は、一秒おきに計測して得られた遅延の値を示しており、青い

点は、同時に計測して得られたフレームドロップレートを示している。図 4.3 では、シーケンス 0 から 13 まで遅延が上昇しており、最小二乗法によりその遅延の増加傾向を計算した所、 $\pm 20.7\text{ns}$ (%表記で 0.1754%) の漸近的標準誤差で毎秒 11800.2 ナノ秒のレートで遅延が増加している事が判明し、その遅延の上昇傾向は図 4.3 上に赤い線で示している。それ以降 1385000 ナノ秒から 1402000 ナノ秒の範囲で遅延の値が減少と急な増加を繰り返している事とフレームがドロップされている事が伺える。この事から Catalyst 3750G-24PS-E は、IPv4 でのパケット転送時に受信したパケットをバッファに記憶し、転送処理を行い続ける為、遅延が高くなると結論づけられる。

4.3.3 遅延の値から見られる AX3630S-24T のバッファリング傾向

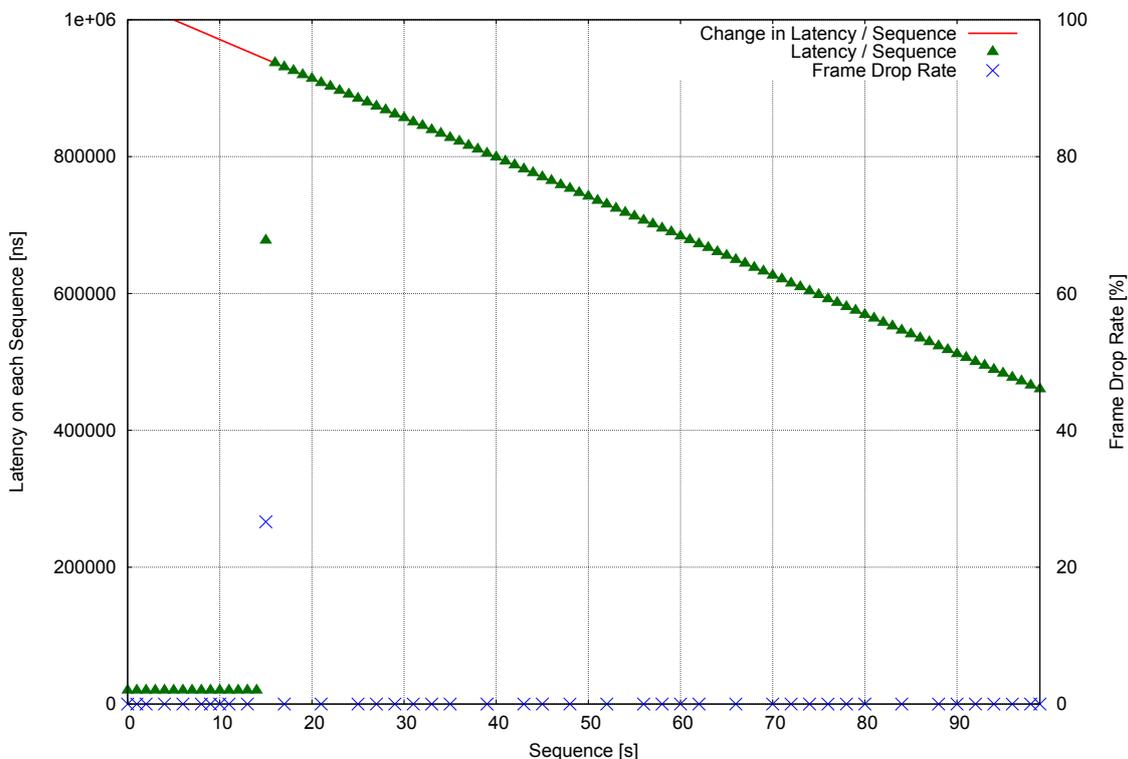


図 4.4: AlaxalA 社の AX3630S-24T のフレームサイズ 2048byte 時の IPv4 による転送時の遅延の変動

図 4.4 は、フレームサイズ 2048byte で 1 秒おきに計測して得られた遅延の値を緑の点で示し、フレームドロップレートを %表記した値を青の点で表したものである。計測シーケンス 15 の時点で急激に遅延が増加し、シーケンス 16 移行は一定のレートで遅延が低下しており、最小二乗法によりその遅延の減少傾向を計算した所、 $\pm 0.9199\text{ns}$ (%表記で 0.01601%) の漸近的標準誤差で毎秒 5744.49 ナノ秒のレートで遅延が減少しており、それ以前ではナノ秒からナノ秒の範囲で遅延の値を示している事が伺える。

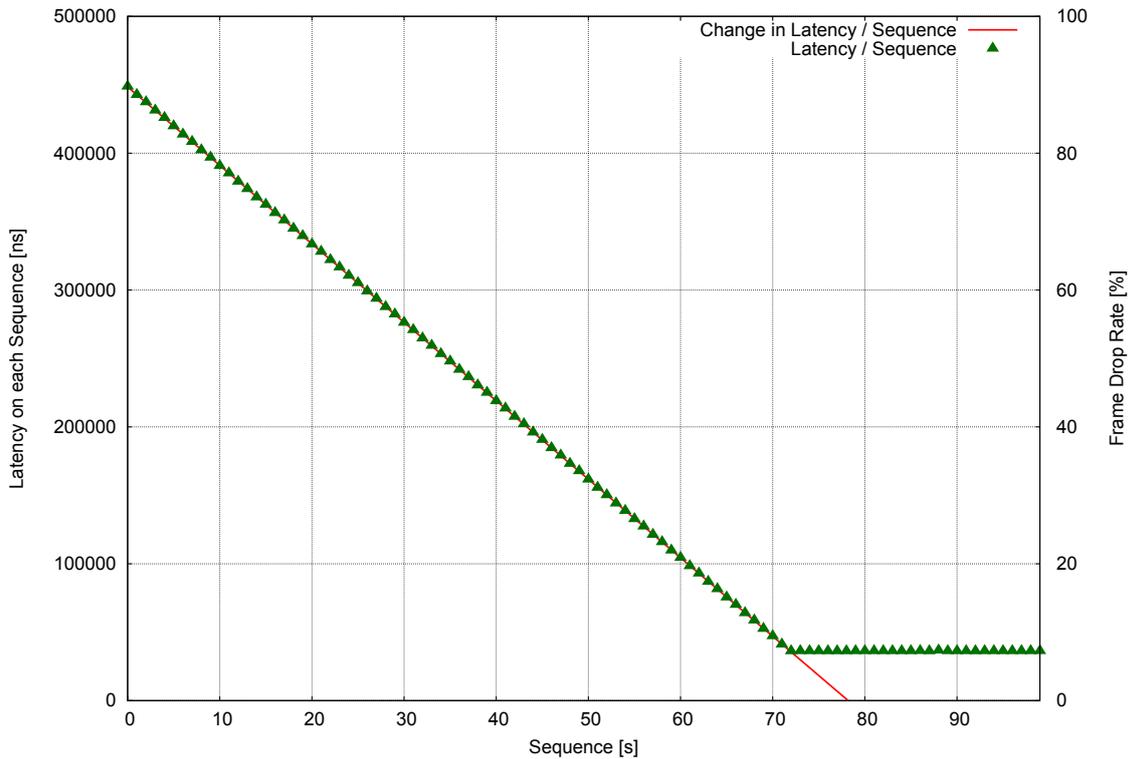


図 4.5: AlaxalA 社の AX3630S-24T のフレームサイズ 4096byte 時の IPv4 による転送時の遅延の変動

図 4.5 は、フレームサイズ 2048 での計測終了後に 1 秒間経った後にフレームサイズ 4096byte で 1 秒おきに計測を行なって得られた遅延を緑の点で示したものだ。フレームサイズ 4096byte 時の計測シーケンス 0 から 72 まで遅延が減少しており、最小二乗法によりその遅延の減少傾向を計算した所、 $\pm 1.233\text{ns}$ (%表記で 0.0215%) の漸近的標準誤差で毎秒 5734.48 ナノ秒のレートで遅延が減少しており、それ以降ナノ秒からナノ秒の範囲で遅延の値を示している事が伺える。この事から AX3630S-24T は、IPv4 での転送時に何らかの理由で徹m総処理が滞り、バッファにパケットが貯まり遅延が上昇したものの、バッファの処理を一定のレートで行いパケットの転送に要する遅延を低下させ続ける事が判明した。

4.3.4 バッファリングの挙動の分析からの考察

3750G は、バッファに IPv4 パケットが溜まってもバッファを減らそうとしない為、IPv4 ルーティングに要する遅延が高い値のまま動作する。AX3630S は、何らかの問題で IPv4 パケットの転送が瞬間的に滞り、バッファにフレームが貯まり遅延が突発的に大きくなるが、バッファに溜まったフレームを少なくし、遅延を減らす動作を行う為、遅延は低下する。その為、IPv4 での転送性能は、AX3630S の方が 3750G よりも高く、低遅延でパケッ

トの転送を行えるような設計になっていると結論づけられる。

4.4 ポート毎の性能の違い

本節では、特定のネットワーク機器では通信に用いるポートによって、機器の内部でフレーム転送に用いられる ASIC が異なっており、位置の離れたポート間の通信では内部チップ間での通信が必要となり、Ethernet フレームの転送速度が隣接ポート間の通信より遅延が高くなるという仮説のもとに計測対象の機材に対し、それぞれ 3 つのポートで計測を行なって得られたデータをグラフに統合し比較を行なった。その結果、BUFFALO 社の LSW4-GT-8EP の計測データからポートによって Ethernet フレームの転送性能が異なる事が判明したため、詳細を示す。

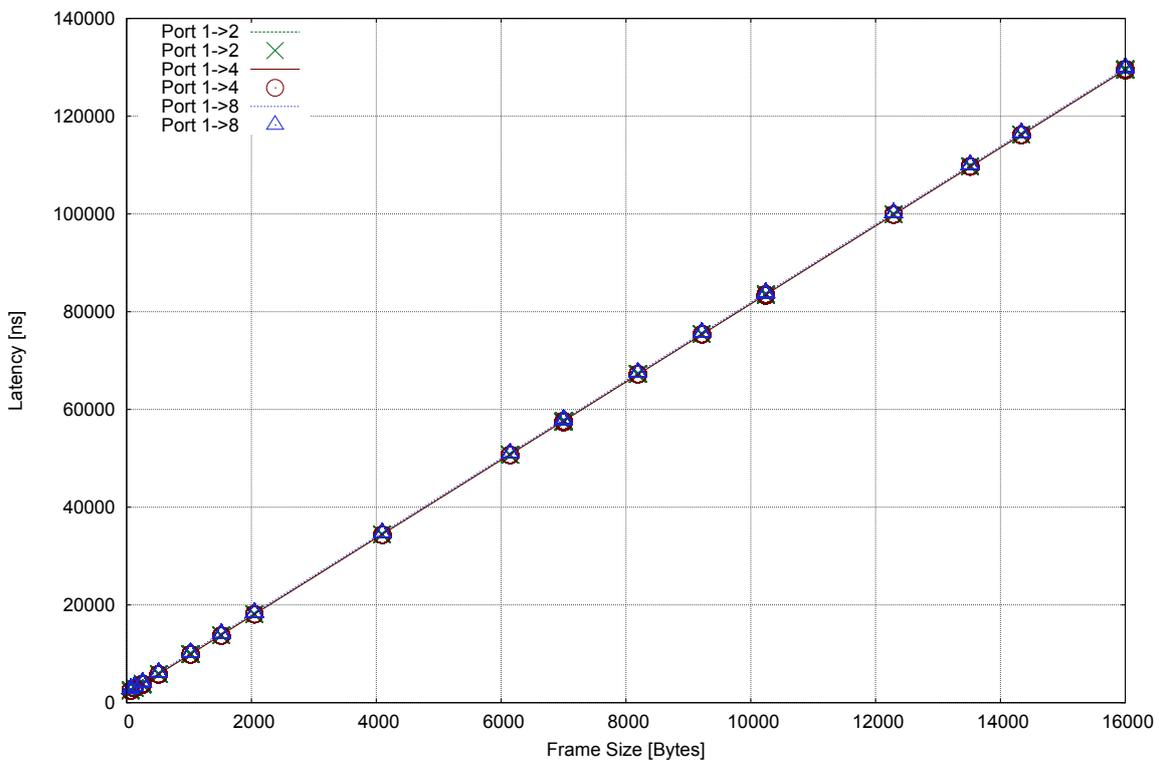


図 4.6: LSW4-GT-8EP のポート毎の Ethernet フレームの転送に要する遅延の違い

4.4.1 ポート毎の遅延とフレームドロップレートの際の考察

図 4.6 は LSW4-GT-8EP に対してポート 1 から Ethernet フレームをポート 2、4、8 へ送信し、計測を行なった結果得られた遅延の値を点で示し、それらの遅延の値から最小二乗法で得られたフレームサイズに対して発生する遅延を示す一次関数を線で示したものだ。図 4.6 の計測して得られた異なるポートの遅延を示す点と線はほぼ重なりあっている

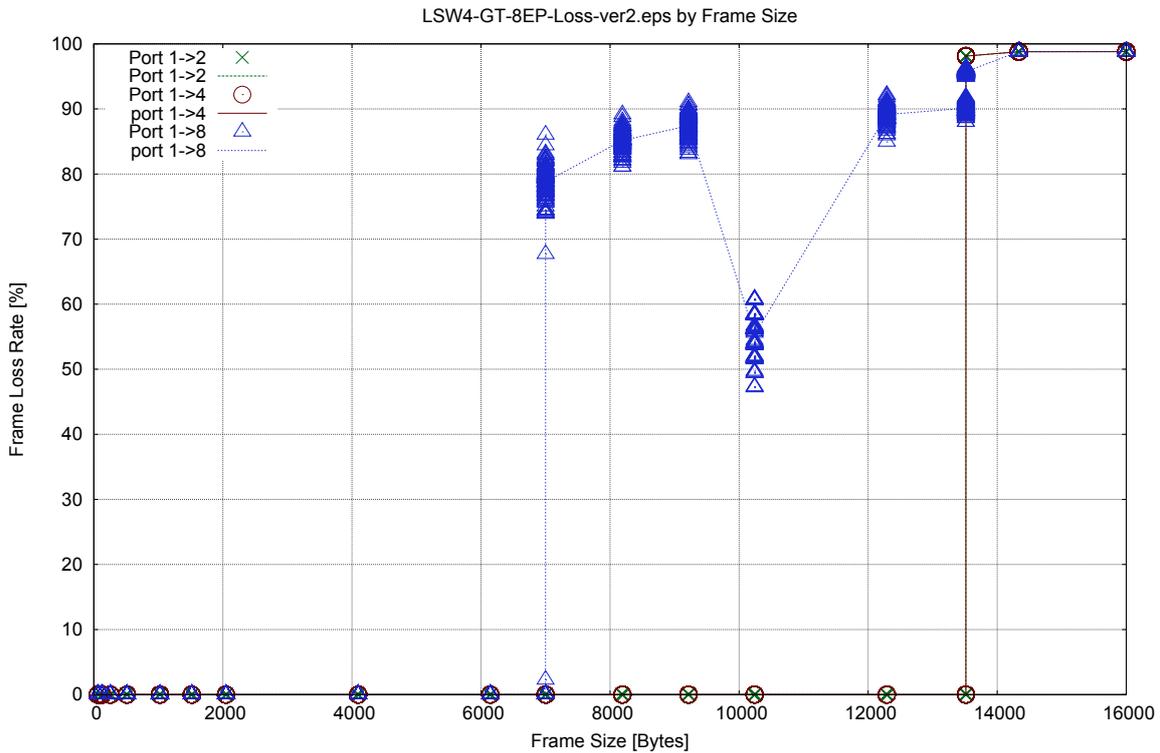


図 4.7: LSW4-GT-8EP のポート毎のフレームドロップレート

ことから、計測結果がいずれも近い値を示しており、ポート 2,4,8 のいずれにおいても 4.2 節にて示されたように Ethernet プロトコルによるフレームの転送性能は極めて高く、低遅延での転送を行なっている事が伺える。

計測を行なったポートでは、それぞれ計測に用いられるフレームサイズが 1byte 増加するのに応じておよそ 8 ナノ秒遅延が上昇している事から、計測対象である LSW4-GT-8EP がカットスルーを行わず、Ethernet フレームがバッファに記憶された後に転送を行うストアアンドフォワード型の特性を示している。また、Ethernet フレームの転送に要する転送速度を示す定数は、ポート 2 では 1796 ナノ秒、ポート 4 では 1739 ナノ秒経過している事が最小二乗法を用いたデータの解析から判明した。この事からポート 2 とポート 4 の間には転送性能に差異は殆ど無い事が明らかである。しかし、ポート 8 では転送処理に要する遅延を示す定数が 2067 ナノ秒となっており、ポート 2、4 と比較して、300ns 程 Ethernet フレームの転送性能に差異が発生している事が判明した。

図 4.7 は、LSW4-GT-8EP に対してポート 1 から Ethernet フレームをポート 2、4、8 へ送信し、計測を行なった結果得られた論理的に送信可能なフレーム数に対受信フレーム数との差異からフレームドロップレートを % 表記で示したものである。フレームドロップレートを確認した所、ポート 2 と 4 で計測に用いられる Ethernet フレームのサイズが 13514byte の時点までは、フレームドロップが発生していなかったが、計測に用いられるフレームサイズが 13515byte に上昇した時点から、突如フレームドロップレートが 98% 程に大きく飛躍した。ポート 8 の計測の際には、計測に用いるフレームサイズが 7000byte の

表 4.3: LSW4-GT-8EP の Ethernet フレームの転送性能を一次関数に最小二乗法を用いて当てはめて表された性能

| 項目 \ 計測対象 | Port 2 | Port 4 | Port 8 |
|-------------------------------------|-----------|-----------|----------|
| フレーム サイズ毎の 遅延の増加 [ns/byte] | 7.98324 | 7.98308 | 7.98417 |
| 転送に要 する遅延 の定数 [ns] | 1796.28 | 1739.55 | 2067.42 |
| 傾斜に対する 漸近的標準誤差 [ns/byte] | 0.0004848 | 0.0004961 | 0.000478 |
| 傾斜に対する 漸近的標準誤差 [%] | 0.006073 | 0.006214 | 0.005987 |
| 遅延の定数に 対する漸近的 標準誤差 [ns](%) | 4.155 | 4.251 | 4.096 |
| 遅延の定数に 対する漸近的 標準誤差 [%] | 0.2313 | 0.2444 | 0.1981 |

時点までは、フレームドロップが発生していなかったが、計測に用いるフレームサイズが 7001byte に上昇してからフレームドロップレートが 80%程に大きく飛躍した。このポート 2、4 と 8 の間の性能の隔たりは、ポート 1 から 4 までは同一 ASIC に接続されており、ポート 5 から 8 は別の ASIC に接続されており、ポート 1 とポート 8 間で計測を行なった際に ASIC 間での通信が発生したためだと推測される。

BUFFALO 社の製品仕様の説明によると、[27] ワイヤレートでの Ethernet フレームの転送性能、つまり最短 IFG で最大レートでフレームを転送処理を行えるのは、9216byte までと記載されているがそれはポート 1 から 4 の間に送信と受信を行う端末が共に接続された場合のみである事、またポート 1 から 4 までの間であればカタログスペックである 9216byte までではなく、13514 byte までワイヤレートでのジャンボフレームの転送が可能である事が本研究での計測から判明した。

第5章 結論

本章では、本論文の成果をまとめ、FPGA を用いて実装された計測器を適切に用いることにより、従来のソフトウェアや ASIC 等を用いた専用計測機器によるネットワーク機器の性能計測手法より精確で正確な計測を行える事を示した。また、FPGA に実装された計測器を用いてネットワーク機器の特性を明らかにし、ネットワーク機器の運用構築実装を行う人物にとって有益な知見をもたらすという目的の中で達成された部分を述べ、達成されていない部位に関して今後の展望として述べる。

5.1 まとめ

本論文では、松谷氏による FPGA を用いたネットワーク機器の性能計測器の実装の校正を行い、その精度と正確さを示した。Ethernet プロトコルに基づいたフレームの転送は安価な家庭用ネットワーク機器の方が低遅延で行えて、バッファリング機構の違いにより転送速度によって大きな違いが発生し、機器によっては離れたポートが内部で別の ASIC に接続されており転送性能に違いがあるという仮説を元にネットワーク機器のフレーム転送性能の計測、並びにデータの解析を行なった。

Ethernet フレームの転送性能の計測結果から第 4.2 節にて、Allied Telesis 社の CentreCOM GS908 や BUFFALO 社の LSW4-GT-8EP の性能の優位性から、家庭用ネットワーク機器の方が業務用ネットワーク機器よりも低遅延で Ethernet フレームの転送を行える場合が存在する事を示した。

第 4.3 節では、バッファリングを行なっている形跡が見受けられる CISCO 社の 3750G-24PS-E と Alaxala 社の 3630S-24T を比較し、バッファリング機構の違いにより、IPv4 パケットの転送に要する遅延が大きく影響を受ける事を示した。

第 4.4 節では、BUFFALO 社の LSW4-GT-8EP で実験結果からポートにより Ethernet フレームの転送時に生ずる遅延に違いがあり、ワイヤレートでの転送処理を行えるフレームサイズに隔たりがある事から、ネットワーク機器によっては、ポートが機器内部にて接続している転送処理に用いられる内部チップが異なり、転送性能に差異がある事が示した。

これらの計測と考察の結果から本研究の仮説の正しさが証明され、有益な知見が示されたと結論づける事が出来る。

5.2 今後の展望

本論文では、ネットワーク機器の性能と性質の一部を示したに過ぎない。ネットワーク機器には多種多様な種類が存在しており、デジタル情報の転送に用いられるプロトコルにも新たな規格が生まれつつある。その為本研究で計測を行わなかったネットワーク機器やプロトコルに対して対応するため、本研究で用いた計測器の実装の拡張を行い、より多くのネットワーク機器の計測を行う必要がある。

本研究で用いた計測器では、計測の為に送信した Ethernet フレームのうち、どのフレームが転送されなかったかの確認を行う事が出来ず、秒間受信フレーム数として示された値に計測対象が接続されている端末に対して生成して送信したキープアライブ信号が混在している恐れがある。また計測器の現段階では、IPv6 でのネットワーク機器の転送性能の計測を行う事が出来なかった。どの Ethernet フレームがネットワーク機器によりドロップされたか、受信したフレームの内どれだけ計測器が発していないフレームが混在しているかを確認出来るように実装の拡張、そしてこれまでインターネットに接続する端末間での通信に重く用いられていた IPv4 のアドレススペースが昨年枯渇した事から [28]、ネットワーク機器の IPv6 でのパケット転送性能の計測への対応は急務である。

現在ネットワークのトポロジーの動的変更をソフトウェアにより実現する、Software Defined Network というネットワークアーキテクチャの実現を目的として OpenFlow という標準の規格策定が行われている。Google、Facebook、Yahoo!、Verizon、Deutsche Telekom、NTT コミュニケーションズ等著名なネットワークを介してサービスを提供している企業により Open Network Foundation という非営利コンソーシアムが組織されており、注目を浴びている [29]。Holzle の発表によると Google では既に OpenFlow によるネットワークの運用構築が行われており、OpenFlow による管理に対応したネットワーク機器の実装も行われている [30]。今後のネットワークの運用構築では、OpenFlow 等の Software Defined Network を実現する技術が利用される局面が増えるものと予測される。その為、OpenFlow に対応したハードウェア実装によるネットワーク機器とソフトウェアルータの性能計測を行い、既存のネットワーク機器との性能の差異の計測を行う必要がある。

また本研究では、Ethernet ヘッダか IPv4 を用いたパケットのルーティング性能のみの計測を行なったが、ラベルと呼ばれる固定長の識別子を用いた MPLS とや LISP (Locator/ID Separation Protocol) 等のパケット転送技術が大手 ISP 等のインターネットへの接続性やクラウドコンピューティングサービスを提供している企業のバックボーンネットワークでのパケットの転送に用いられている [31][32]。これらのパケット転送技術利用時のネットワーク機器の転送性能の計測も今後の課題である。

謝辞

本論文の作成にあたり、ご指導頂いた慶應義塾大学環境情報学部教授村井 純博士、同学部教授 徳田 英幸博士、同学部教授 中村 修博士、同学部准教授 楠本 博之博士、同学部准教授 高汐 一紀博士、同学部准教授 三次 仁博士、同学部准教授 植原 啓介博士、同学部准教授 Rodney D. Van Meter III 博士、同学部専任講師 中澤 仁博士、同学部教授 武田 圭史博士、同大学特任講師 吉藤 英明博士、同大学特任講師 斉藤 賢爾博士に感謝致します。特に Rodney D. Van Meter III 博士は、研究で行き詰まる私に対して非常に力強い応援の言葉をかけていただき、何度も私にサイエンスへの取り組みかたを教えて下さいました。本当にありがとうございました。

そして、本研究を進めていく上で様々な励ましと助言をいただきました、村井研究室卒業生である黒宮 祐介氏、六田 佳祐氏、峯木 巖氏、村上 滋希氏に感謝致します。

慶應義塾大学大学院政策・メディア研究科後期博士課程 岡田 耕司氏、堀場勝広氏、工藤 紀篤氏、松谷 健史氏、空閑 洋平氏、同研究科修士課程、三部 剛義氏、佐藤 弘崇氏に感謝致します。特に松谷 健史氏は、博士論文の執筆や実装で多忙な身にも関わらず、親身に相談に乗っていただき、研究の方向性の指導や実装の細やかなケア等の技術的な指導から、研究活動の推進に不可欠なコミュニケーション能力までありとあらゆる面で面倒を見ていただきました。氏なしでは卒論執筆だけでなく充実した研究活動を送れませんでした。強く心から感謝致します。

研究に協力をしていただいた、山岸 祐大氏、三條場 直希氏、由井 卓哉氏、碓井 利宣氏、と徳田・村井合同研究室の皆様感謝致します。

特に計測対象であるネットワーク機器の貸出を行なっていただいた、慶應義塾大学政策・メディア研究科後期博士課程 本多 倫夫氏、同研究科修士課程 星 北斗氏に感謝致します。

研究室で苦楽を共にし、道半ばにて敗れつつも卒論執筆の応援をして下さった中村 尚生氏に心から感謝します。彼と一緒に研究をし、交流することでお互いに励まし合い、意識を低める事ができました。

私の4年間の大学生活の健康の源であったSFC合気会と部員全員に感謝致します。

最後に、大学入学からの4年間だけでなく23年間をあらゆる面で支えていただいた父、澁田 一正、母、澁田 京子と私の家族に心から感謝致します。

参考文献

- [1] 内閣府経済社会総合研究所景気統計部. 消費動向調査結果. <http://www.esri.cao.go.jp/jp/stat/shouhi/shouhi.html>, 6 2012.
- [2] 総務省情報通信政策局. 通信利用動向調査報告書世帯編. <http://www.soumu.go.jp/johotsusintokei/statistics/statistics05b1.html>, 6 2012.
- [3] ITpro. 【初級】知っておきたいストレージの基礎 最終回(後半) nas と iscsi の特徴. <http://itpro.nikkeibp.co.jp/article/COLUMN/20060407/234778/>, 4 2006.
- [4] Kazutoshi Fujikawa Takahiro Hirofuchi, Eiji Kawai and Hideki Sunahara. Usb/ip - a peripheral bus extension for device sharing over ip network. <http://static.usenix.org/events/usenix05/tech/freenix/hirofuchi/hirofuchi.pdf>, 4 2005.
- [5] 松谷 健史. Ip 化を実現するバスブリッジの設計と実装. <http://www.sfc.wide.ad.jp/thesis/2008/master/macchan-mthesis.pdf>, 7 2008.
- [6] Raouf Boutab Qi Zhang, Lu Cheng. Cloud computing: state-of-the-art and research challenges. <http://cloud.pubs.dbs.uni-leipzig.de/sites/cloud.pubs.dbs.uni-leipzig.de/files/fulltext.pdf>, 1 2010.
- [7] Inc Microsoft. skype. <http://www.skype.com>, 6 2012.
- [8] OKI. 声と音の技術 voip 技術. <http://www.oki.com/jp/rd/ss/voip.html>, 6 2012.
- [9] IETF. Internet small computer systems interface (iscsi). <http://tools.ietf.org/html/rfc3720>, 6 2012.
- [10] CISCO. Fibre channel over ethernet(fcoe). http://www.cisco.com/web/JP/news/cisco_news_letter/tech/fcoe/index.html, 6 2012.
- [11] 山口実靖, 小口正人, and 喜連川優. 高遅延広帯域ネットワーク環境下における iscsi プロトコルを用いたシーケンシャルストレージアクセスの性能評価ならびにその性能向上手法に関する考察 performance evaluation and improving of sequential storage access using the iscsi protocol in long-delayed high throughput network. <http://www.dbsj.org/journal/vol2/no1/papers/yamaguthi.pdf>, 6 2012.
- [12] Mike Muss. ping.c. <http://minnie.tuhs.org/cgi-bin/utree.pl?file=2.11BSD/src/bin/ping/ping.c>, 12 1983.

-
- [13] 有地博幸, 渡邊義康ほか (東陽テクニカ), and ITmedia. 計る測る量るスペック調査隊: ネットワークスイッチのスループットを調査せよ【前編】 (3/3). http://www.itmedia.co.jp/enterprise/articles/0803/24/news008_3.html, 3 2008.
- [14] Jozsef Imrek Peter Orosz, Tamas Skopko. Performance evaluation of the nanosecond resolution timestamping feature of the enhanced libpcap. http://www.thinkmind.org/download.php?articleid=icsnc_2011_9_50_20258, 10 2011.
- [15] BUFFALO. ルータのスループット値と測定環境について. http://buffalo.jp/products/catalog/network/router_throughput/.
- [16] SPIRENT Communications. Smartbits dual media ethernet modules smartmetrics/terametrics xd. <http://www.toyo.co.jp/file/pdf/spirent/smartbits/pdf/xdmodule.pdf>.
- [17] NetFPGA Group. Netfpga. <http://www.netfpga.org/php/specs.php>.
- [18] lenovo. Thinkcentre s50(8183-64j) - 製品仕様. http://support.lenovo.com/ja_JP/downloads/detail.page?DocID=PD020298.
- [19] ワイエス・ソリューションズ株式会社. 声と音の技術 voip 技術. http://www.yso1.co.jp/lan_cabling/utp/utp_c5e5_st.html.
- [20] IEEE P802.3ab 1000BASE-T Task Force. Ieee p802.3ab 1000base-t task force. <http://www.ieee802.org/3/ab/index.html>.
- [21] CISCO. Lan base ソフトウェア搭載 cisco catalyst 2960-s および 2960 シリーズスイッチ. http://www.cisco.com/web/JP/product/hs/switches/cat2960/prodlit/pdf/product_data_sheet0900aecd80322c0c.pdf.
- [22] CISCO. Cisco catalyst 3750 シリーズ スイッチ. http://www.cisco.com/web/JP/product/hs/switches/cat3750/prodlit/pdf/cat50_ds.pdf.
- [23] Alaxala. Ax3630s シリーズ. http://www.alaxala.com/jp/products/archive/datasheet_pamphlet/pdf/ax3630s_ds_op_ver11_9.pdf.
- [24] BUFFALO. Lsw4-gt-8ep. <http://manual.buffalo.jp/buf-doc/35010554-1.pdf>.
- [25] BUFFALO. Lsw3-gt-5ep. <http://manual.buffalo.jp/buf-doc/35010423-1.pdf>.
- [26] Allied Telesis. ギガビットイーサネット・スイッチ centrecom gs908e ユーザーマニュアル. <http://www.allied-telesis.co.jp/support/list/switch/g908e/m000408d.pdf>.
- [27] BUFFALO. Lsw4-gt-8ep 製品仕様. <http://buffalo.jp/products/catalog/network/lsw4-gt-8ep/index.html?p=spec>.

- [28] 社団法人ネットワークインフォメーションセンター. Ipv4 アドレスの在庫枯渇に関して. <http://www.nic.ad.jp/ja/ip/ipv4pool/>, 2 2011.
- [29] atmarkIT 三木泉. Open networking foundation から見た sdn と openflow. <http://www.atmarkit.co.jp/news/201206/15/onf.html>, 6 2012.
- [30] Urs Hoelzle. Openflow @ googl. <http://opennetsummit.org/talks/ONS2012/hoelzle-tue-openflow.pdf>, 4 2012.
- [31] 日経コミュニケーション 高橋 健太郎. [コア・ネット]mpls で統合網を構築, 耐障害性を徹底強化. <http://itpro.nikkeibp.co.jp/article/COLUMN/20100216/344603/>, 2 2010.
- [32] CISCO. 特集記事 01 [クラウド最前線] 事例 3: Nec ビッグローブ株式会社. http://inspire.cmarket.jp/0005/cover_story1_7.html, 11 2011.