社団法人　情報処理学会　研究報告
IPSJ SIG Technical Report

2004－ＤＳＭ－33　(14)
2004／5／14

高精度時刻PCによるネットワークトポロジ計測手法

北口　善明†　　町澤　朗彦††　　箱崎　勝也†††

† インテック・ウェブ・アンド・ゲノム・インフォマティクス（株）
†† 独立行政法人 情報通信研究機構
††† 電気通信大学
E-mail: †kitaguchi_yoshiaki@webgen.co.jp, ††machi@nict.go.jp, †††hako@is.uec.ac.jp

あらまし　LANにてネットワークを柔軟に利用する方法にVLANがある．このVLANはスイッチ群にて構成されたネットワーク（スイッチ・ネットワーク）上に構築される論理的なLANであり，この利用のためにスイッチ・ネットワークが多く利用されている．このスイッチ・ネットワークではIPレイヤから機器の存在を確認することができず，そのためネットワークトラブルシュート時における問題の切り分けが難しくなっている．本研究では，帯域計測手法を用いることで未知のリンク内におけるスイッチ機器の段数を推定する手法について述べ，提案手法の評価実験の結果を報告する．

キーワード　スイッチ・ネットワーク，ネットワークトポロジ，高精度時刻PC，ネットワーク計測

The network topology measurement technique useing a High-precision PC

Yoshiaki KITAGUCHI†, Akihiko MACHIZAWA††, and Katsuya HAKOZAKI†††

† INTEC Web and Genome Informatics Corp., Tokyo
†† National Institute of Information and Communications Technology, Tokyo
††† University of Electro-Communications, Tokyo
E-mail: †kitaguchi_yoshiaki@webgen.co.jp, ††machi@nict.go.jp, †††hako@is.uec.ac.jp

Abstract　VLAN is in the method of using a network flexibly in LAN. This VLAN is logical LAN built on the network consisted of some switches (Switch Network), and many switch networks are used for this use. In a switch network, existence of apparatus cannot be checked from IP layer. Therefore, it is difficult to specify a problem at the network troubleshooting. This research describes the technique of presuming the number of the switch apparatus in a unknown link and reports the result of an evaluation experiment of the proposal technique.

Key words　Switch Network, Network Topology, High-Precision PC, Network Measurement

1. はじめに

VLANは，物理的な接続形態とは独立に端末の仮想的なグループを設定することにより，端末の物理的な位置を気にすることなくネットワークを構成することができるメカニズムであり，ネットワークの構成を柔軟に変更・管理するために有効である．このVLANを利用するためにスイッチ・ネットワークが多く利用されているが，このスイッチ・ネットワークではスイッチ機器の存在がIPレイヤから確認することができないため，ネットワークトラブルシュート時における問題の切り分けが難しいという点を持っている．今日のスイッチ・ネットワークの運用管理においてはスイッチ機器のネットワークトポロジの把握がトラブル発生時の問題解決に不可欠な要素となっているため，その管理は重要である．この問題を解決する手法としては，同一ベンダの機器を用い独自の管理プロトコルによりスイッチネットワークの管理を行う方法やSNMP/RMON [1] を利用する方法が一般的に用いられている．しかしこれらの方法ではスイッチネットワークの構成を知るためにそのネットワークに対する権限を有する必要があり，また，未対応の機器に対しての利用ができない弱点を持っている．

本研究では管理権限のないリンク内においてもスイッチ機器の存在を把握する手法の提案を行う．計測手法には，サイズの異なるパケットを使いその通過時間からリンクの帯域を推定するOne-Packet計測手法のアルゴリズムを用い，高精度時刻

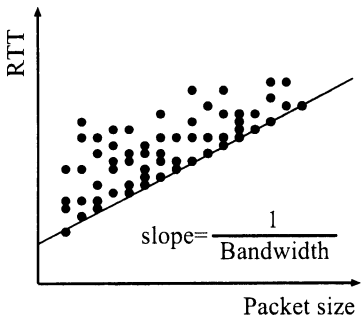

図 1　One Packetk 計測手法での帯域推定方法

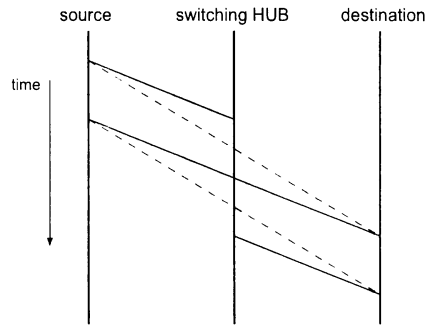

図 2　One Packet 手法での問題点

PC を利用して提案手法の評価を行う．

2. 帯域計測手法

ネットワークの帯域を計測する手段としてこれまでに多くの研究がなされ，それに伴うアプリケーションも数多く開発され広く利用されている．その手法の 1 つとして One-Packet 計測手法があり，以下にその手法についてまとめる．

2.1 One-Packet 計測手法

One-Packet 計測手法は Bellovin [2] により提唱された手法である．サイズを様々に変化させた単一の計測パケットを用い，その転送時間を計測し，各パケットサイズでの最小転送時間を求めその回帰直線の傾きを求める（図 1 参照）．この傾きの逆数が求める帯域の値となる．ここで最小転送時間を用いるのは，計測時に他の要因で遅延が発生した場合の影響を抑えるためである．上記の説明を数式を用いて行うと次のようになる．パケットサイズ p におけるリンク通過時間を $T(p)$ はリンク帯域を B とすると

$$T(p) = \frac{p}{B} + C \qquad (C \text{ は定数}) \tag{1}$$

と表すことができ，この両辺をパケットサイズ p で偏微分することで求まる関数の逆数が推定帯域の値になる．

$$\frac{\partial T(p)}{\partial p} = \frac{1}{B} \tag{2}$$

この計測手法の欠点としては，スイッチ等のストア・アンド・フォワード機器の影響により推定帯域の値が小さくなる点が挙げられる．これは，One-Packet 計測手法がパケットサイズと通過時間の関係から推定することに基因しており，図 2 のように実線のスループットが見かけ上点線のように半分の値として求まることになる [3]．つまり，式 (1) の右辺にスイッチ部分での経過時間が追加されるため，スイッチ機器の段数を x 個とすると次のようになる．

$$T(p) = \frac{p}{B} + \frac{p}{B}x + C \qquad (C \text{ は定数}) \tag{3}$$

この式 (3) を p で偏微分し B で整理すると

$$\frac{\partial p}{\partial T(p)} = \frac{B}{(x+1)} \tag{4}$$

となり，スイッチ機器の段数で帯域が小さく求まることがわかる．

表 1　計測に使用した PC の仕様

PC 1 Specification	
OS	Linux-2.4.18 + PPSkit (RedHat Linux 7.3)
	FreeBSD-4.8 RELEASE
Motherboard	Intel Server Board SCB2 SCSI
CPU	Intel Pentium III 1.13GHz-S
PCI BUS	64bit 66MHz
NIC (drivewr)	Intel Pro/1000 XT (e1000/cm)
	Netgear GA620T (acenic)
	PCI GN-1000TE (ns83820)
PC 2 Specification	
OS	Linux-2.4.18 + PPSkit (RedHat Linux 7.2)
Motherboard	MSI MS-6351 (Ver5)
CPU	Pentium III 996.68MHz
PCI BUS	32bit 33MHz
NIC (drivewr)	Intel Pro/1000 XT (e1000)

3. 提案手法

前章に記述したように，One-Packet 計測手法ではスイッチ機器により推定帯域の結果に直接影響を受けることから，計測リンク帯域が求まっている条件下においてこの特長を利用した場合，同リンク内に存在するスイッチ機器の数を推定することが可能であると考えられる．本手法を確立するためにあらかじめ帯域やスイッチ機器の段数が分かっているリンクにて評価実験を実施した．次にその評価実験のための計測システムについてまとめる．

3.1 評価計測システム

One-Packet 計測手法ではパケットサイズの変化に対する遅延時間の変化が求まれば良いため絶対時刻の同期は必要ではない．この理由から，タイムスタンプとして PCC（Processor Cycle Counter）を用いネットワーク経由で比較する手法を用いた．PCC の値に利用する PC の CPU 周波数をかけることで時間に直すことができる．またこの手法では，プロトコルとして UDP を用い，片道遅延時間を用いた帯域推定手法を用いている [4]．さらに，PC 自身の時刻精度の影響を抑えるために

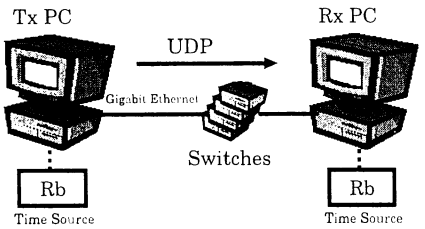

図 3　評価実験時の機器構成

表 2　各 NIC におけるスイッチの段数と計測帯域
上：帯域　下：標準誤差
Table 2

swicth	Pro/1000 XT	GA620T	GN-1000TE
0	443.47 Mbps	389.99 Mbps	459.94 Mbps
	0.316 %	10.28 %	0.337 %
1	307.50 Mbps	365.47 Mbps	315.07 Mbps
	0.241 %	13.03 %	0.257 %
2	234.99 Mbps	227.89 Mbps	239.21 Mbps
	0.194 %	5.96 %	0.189 %
3	190.33 Mbps	197.60 Mbps	192.75 Mbps
	0.168 %	6.85 %	0.164 %
4	160.12 Mbps	159.37 Mbps	161.69 Mbps
	0.155 %	3.89 %	0.169 %
5	138.17 Mbps	137.90 Mbps	139.14 Mbps
	0.151 %	3.87 %	0.144 %
6	121.20 Mbps	129.59 Mbps	122.09 Mbps
	0.132 %	3.68 %	0.140 %
7	108.11 Mbps	109.57 Mbps	108.94 Mbps
	0.137 %	2.67 %	0.132 %

我々の開発した高精度時刻 PC [5] [6] を用い，外部から供給する周波数源としてルビジウム基準信号発振器を使用した．このルビジウム基準信号発振器を用いることで PC の時刻安定度を数百ナノ秒台で維持することが可能である [7]．

以上のシステムおよび計測手法を用いることで，精度の良い評価実験を可能としている．

3.2 計測形態

表 1 に計測に用いた高精度時刻 PC のスペックをまとめる．PC1 では NIC の違いと OS による影響を調査し，PC2 はハードウエア（HW）の違いによる影響を比較測定するためにそれぞれ用いた．PC1 と PC2 の相異点としては PCI バスのクロックスピードであり，この影響を後程の評価実験にて検証する．図 3 に今回の評価実験時の構成を示す．二台の PC 間をギガビットイーサネットで結び，間に入れるスイッチを 0 段から 7 段にまで変化させて，帯域計測を実施した．計測時は他のトラフィックからの影響を排除するため，計測パケットのみが流れる環境とし，16 バイト刻みのパケットにて各々の試行回数を 10 回として $10ms$ 間隔で送信し，16 バイトから 1440 バイトまでの計 900 パケットを用いて実施した（計測期間は約 $10s$）．また，今回利用したスイッチは SMC 社の EZ8508T2 ですべて同種のものを用い，VLAN やポートミラーリング等の設定は行わない状態で使用した．

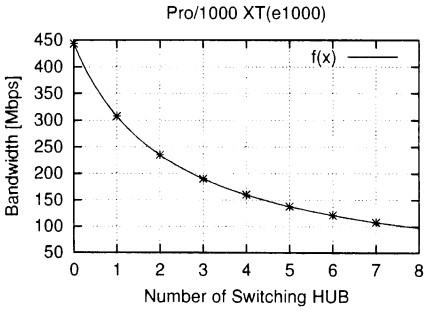

図 4　NIC 毎によるスイッチの段数と計測帯域の関係

4. 評価実験

本手法では PC をベースとした計測方法とするため，利用する PC による影響を調査する必要がある．この章では提案手法を計測 PC の環境を変化させて計測した結果を元に考察を行う．

4.1 NIC による影響

表 2 に，三種類の NIC を用いた計測結果をまとめる．NIC には Intel Pro/1000 XT，Netgear GA620T，PCI GN-1000TE を用い，Linux のデバイスドライバ e1000，acenic，ns83820 をそれぞれ使用している．この計測結果より，スイッチがない場合とスイッチが 1 段の場合を比較した場合，当初予想していたように計測される帯域が単純に半分になっていないことがわかる．この原因は，パケットの経過時間を観測した区間にイーサネット以外の影響があるためだと考えることができる．

帯域計測時に得られるパケットの転送時間 $T(p)$ には，計測リンクにおける伝送遅延だけではなく，送信側 PC でタイムスタンプを打ってから実際に回線上にパケットが送出されるまでの時間 T_s およびパケットが受信 PC の NIC に取込まれてからタイムスタンプが呼出されるまでの時間 T_r も含まれているとできるためで，式 (1) は正確には，

$$T(p) = T_s + \frac{p}{B} + T_r + C \tag{5}$$

となる．これら NIC 内でかかる時間は NIC メモリの帯域に因るものと考え，送受信それぞれの NIC 帯域を B_s, B_r とし，さらに，x 段のスイッチ機器の影響を追加すると

$$T(p) = \frac{p}{B_s} + \frac{p}{B} + \frac{p}{B}x + \frac{p}{B_r} + C \tag{6}$$

となり，両辺を p で偏微分することで，求まる推定帯域とスイッチ機器の段数との関係は次式のようになる．

$$\frac{\partial p}{\partial T(p)} = \frac{BB_sB_r}{B_sB_r(1+x) + B(B_s + B_r)} \tag{7}$$

この式 (7) を用い，測定した計測値とスイッチ機器の段数の関係式を最小二乗法にて求める．計測に用いたリンク帯域はギガビットイーサネットであるので式 (7) の B の値を 1000 として，B_s と B_r の値を求めると表 3 のようになった．求められた関数と測定値をプロットしたものを図 4 に示す．この結果から

表 3 NIC と NIC 帯域の関係

	Pro/1000 XT	GA620T	GN-1000TE
T_s	1594.70 Mbps	1322.64 Mbps	1702.05 Mbps
T_r	1594.70 Mbps	1322.64 Mbps	1702.05 Mbps

表 4 OS, HW の違いとスイッチの段数による計測帯域
上：帯域 下：標準誤差
Table 4

swicth	Linux(PC1)	FreeBSD(PC1)	Linux(PC2)
0	443.47 Mbps	444.58 Mbps	320.43 Mbps
	0.316 %	0.700 %	19.6 %
1	307.50 Mbps	297.13 Mbps	258.34 Mbps
	0.241 %	2.57 %	15.7 %
2	234.99 Mbps	235.55 Mbps	214.11 Mbps
	0.194 %	2.00 %	13.0 %
3	190.33 Mbps	189.46 Mbps	182.74 Mbps
	0.168 %	1.98 %	11.1 %
4	160.12 Mbps	157.49 Mbps	159.83 Mbps
	0.155 %	1.21 %	9.93 %
5	138.17 Mbps	138.19 Mbps	142.93 Mbps
	0.151 %	1.41 %	8.92 %
6	121.20 Mbps	120.95 Mbps	127.06 Mbps
	0.132 %	0.222 %	7.78 %
7	108.11 Mbps	107.96 Mbps	116.72 Mbps
	0.137 %	0.213 %	7.25 %

送受信の NIC での帯域は等しいとみなすことができ，各 NIC 毎に異なる値を持つことがわかった．また図 4 より，求めた関数は計測値とほぼ一致する曲線となっていることがわかる．したがって，利用する各 NIC におけるこの帯域値をあらかじめ取得することでスイッチの段数と計測値の関係式が求まり，スイッチ機器の段数推定が可能となることがわかる．

4.2 PC 内部バスの影響

前述の NIC 毎の影響が他の要因に起因するものであるか検証するために次の検証を実施した．

- OS を変更した場合
- HW を変更した場合

これらの計測を Pro/1000 XT の NIC に限定して実施した．表 1 に示した PC1 の OS を FreeBSD に変更した場合の計測結果と，OS は Linux のままで，ハードウェアを PC2 とした場合の計測結果をそれぞれ表 4 にそれぞれ示す．比較対象データとして PC1 にて Linux を用いた場合の結果もまとめて表記している．また，表 4 の計測結果を用い，前節で示した評価関数から OS と HW の組合せによる NIC 帯域を求める方法を取り，その結果を表 5 にまとめて示す．

これらの結果を見ると，OS の変更に伴う影響はほとんどないことがわかる．OS が Linux の場合と同様に FreeBSD の場合でも NIC 帯域としている値が約 1600Mbps と同じ値となっている．これに対して HW の違いによる影響は顕著に現れていることがわかる．PC2 の場合に NIC 帯域値が PC1 の場合の 1/5 程度と大きく異なっている．これは，PCI バスの転送速度が 32bit/33MHz で 1066Mbps，64bit/66MHz で 4266Mbps であることの影響が大きいと想定できる．今回得られた NIC 帯域としている値と PCI バスの関係の詳細を求めることは本提案手法において重要と考えており今後の課題としている．

表 5 OS, HW と NIC 帯域の関係

	Linux on PC1	FreeBSD on PC1	Linux on PC2
T_s	1594.70 Mbps	1574.45 Mbps	680.947 Mbps
T_r	1594.70 Mbps	1575.31 Mbps	680.946 Mbps

5. おわりに

本稿では，帯域計測手法である One-Packet 手法を用いて，管理者権限のない未知のリンクにおけるスイッチ段数を推定する手法を提示し，その可能性を示した．本手法の評価実験を行った結果，ギガビットネットワークのような広帯域ネットワークにて One-Packet 計測手法を利用すると，NIC や PC 内部の影響が大きく現われることがわかり，本手法の利用においては，利用する HW と NIC による NIC 部分での帯域を事前に求める必要があると言えた．またこの影響は OS の実装には非依存であることも求まった．

今後の課題としては，NIC 帯域とした部分の詳細な解析を行うこととルータを越えるパス内での本手法の有効性を調査することを考えている．

謝辞 本研究の一部は通信・放送機構「ギガビットネットワーク研究開発プロジェクト」の支援を受けている．ここに記して謝意を表す．

文 献

[1] S. Waldbusser, "Remote Network Monitoring Management Information Base," RFC1757, Feb. 1995.
[2] S. M. Bellovin, "A Best-Case Network Performance Model." http://www.research.att.com/~smb/papers/netmeas.ps, Feb. 1992.
[3] R. S. Prasad, C. Dovrolis and B. A. Mah, "The Effect of Layer-2 Switches on Pathchar-like Tools," Proceedings of ACM Internet Measurement Workshop 2002, pp.321–322, Nov. 2002.
[4] 町澤朗彦, 北口善明, 岡沢治夫, 中川晋一, "ネットワークを用いた CPU 動作周波数期間安定度の精密計測," DICOMO2002, シンポジウム論文集, pp.563–566, July 2002.
[5] H. Okazawa, A. Machizawa, S. Nakagawa, Y. Kitaguchi, T. Asami and A. Ito, "Advanced NTP Synchronization Device for Internet Monitoring Tools," Proc. INET2001 <http://www.isoc.org/inet2001/CD_proceedings/T42/inet2001.html>, Stockholm, June 2001.
[6] Y. Kitaguchi, H. Okazawa, S. Shinomiya, Y. Kidawara, K. Hakozaki and S. Nakagawa, "Development of a High-Accurate Time Server for Measurements of the Internet," Lecture Note in Computer Science 2344, pp.351–358, Jan. 2002.
[7] 北口善明, 町澤朗彦, 中川晋一, 西村道明, 日迫彰, 箱崎勝也," PC における時刻精度の精密計測とその評価," NS 研究会, 信学会, Nov. 2004.