

ネットワークトラフィックの予測とその応用

篠 秀明 阿波 賢二 大久保 智史 八名 和夫
法政大学工学部電子情報学科

あらまし

通信ネットワークへの負荷が増大する現在、トラフィックを分散制御しネットワークを効率的に利用することは非常に有用であり、最適な分散制御を行なうためには未来のトラフィックを予測することが要求される。SNMP (Simple Network Management Protocol) により容易に観測されるネットワークトラフィックは短時間的に無相関に近い性質を有すためそのままでは予測にむかない。そこで本稿ではトラフィックの差分系列が短時間的に負の相関を有することに注目し、差分系列を予測することでネットワークトラフィックを予測した。その結果、現時点を次点の予測値とする予測手法よりも良好な予測結果が得られた。また、トラフィック予測の応用としてネットワークトラフィックの分散制御方法についても述べる。

1 まえがき

近年インターネットへ接続する端末が増加し、ネットワーク上には大量のトラフィックが生じている。データ通信を行なう媒体の通信回線には転送容量が決まっており容量を越える状態またはそれに準じる状態ではスループットの低下、遅延等が起こる。このためトラフィック量を軽減しスループットを向上するようなシステムが要求されるが、現状ではキャッシュサーバの導入や回線容量を増すなどの措置がとられるに留まっている。また、大規模ネットワークではネットワーク構造が複雑になりがちであり無駄になる回線も存在する場合もある。そこで、トラフィックの分散制御等により負荷を分散し効率的にネットワーク運用する技術が必要となる。ここで、未来のトラフィック量を正確に予測できればトラフィックの最適な分散制御を行なう事が可能となる。

SNMP はネットワーク管理システム専用のプロトコルであり、大規模なネットワークでも比較的簡単にネットワークの管理、監視や、トラフィックの監視が可能である。SNMP により得られるトラフィック時系列は短時間的には無相関に近い統計的性質を有している。このことがトラフィック時系列の予測を困難にさせている要因の1つと考えられ、またこのような信号を予測、制御することは工学的に興味深い。そこで本稿ではSNMP により得られるトラフィック時系列を予測し、その応用としてネットワークトラフィックの分散制御方法について述べる。

2 SNMP によるネットワークトラフィック計測

現在のネットワークシステムには管理プロトコルが用意され大規模ネットワークでも簡単にその性能を調査、監視する事が出来る。本章では、ネットワーク管理プロトコルであるSNMP(Simple Network Management Protocol)を利用するネットワーク性能計測ツールを使用して、ネットワークにおけるデータ収集、性能評価を行う方法について述べる。

2.1 SNMP

SNMP は、TCP/IP プロトコル群で標準化されたネットワーク管理プロトコルである。現在のネットワーク機器の多くがこのSNMP に準拠している。

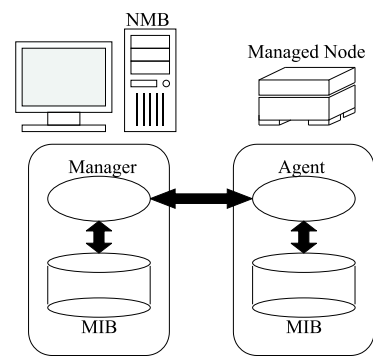


図 1. SNMP 概念図

SNMP が前提とするモデルは、管理対象ノード、ネットワーク管理ステーション、管理情報ベースで構成される (図 1)。

○管理対象ノード (Managed Node)

管理対象となるネットワーク機器である。ここには Agent とよばれるプロセスが存在し、管理対象の状況をモニタリングし MIB と呼ばれるデータベースに記録している。Agent は、ネットワーク管理ステーションからの SNMP によるリクエストを処理する。

○ネットワーク管理ステーション (Network Management Station)

実際の管理実行するプログラムがあり、SNMP を用いて管理対象ノードに対し問い合わせや操作を行う。

○管理情報ベース (Management Information Base: MIB)

管理対象ノードの情報を記録するためのデータベースの構造を規定する。したがって、管理対象ノードとネットワーク管理ステーションでは同じ MIB が使用される。

現在も新しい MIB のノードが加えられ提唱、標準化作業がおこなわれている。

一般的な SNMP の使用方法は、MIB を参照し、ネットワークの管理対象ノードへ定期的な要求を出しネットワーク機器、ネットワークの状態をモニタリングする。ネットワークの状態を調べるために定期的な要求を出す事をポーリング (polling) という。この操作を行う事による問題は、ネットワーク的に遠い位置に存在するネットワーク管理ス

ーションから管理対象ノードに対してポーリングをおこなうとトラフィックが増えてしまい、ネットワークに負荷をかけてしまう事が挙げられる。これを解消する為に管理対象ノードのネットワーク的に近い場所にRMON(Remote Monitoring)を設置する。RMONは、トラフィック計測などの為に設置する事が多く、他のネットワークに負荷をかけないように同じネットワーク内に設置される。

2.2 MIB

MIB(Management Information Base)は、SNMPの重要な構成要素である。SNMPにおける管理対象ノードの情報を記述するデータベースの構造はMIBによって規定されている。管理の対象となる機器においてはSNMP Agentがノードの状態を調べ、MIBに準拠したデータベースに格納する。MIBのデータベースは、木構造を形成する空間を定義している。木の各節点には名前と番号がつけられ、それぞれの葉の部分を実際のオブジェクトになる。これらを表記する場合は、一般に木の根の部分から順番にドットで区切り、ラベルを列挙する。

木の上部構造は、固定である。privateと呼ばれるノード以下において、各企業が自由に使える空間として提供されている。これは、企業別にノードを用意し、各企業の衝突を防ぎ、自由にこの空間を使用できる事を考慮している。図2は、MIBで定義されている主なノードである。

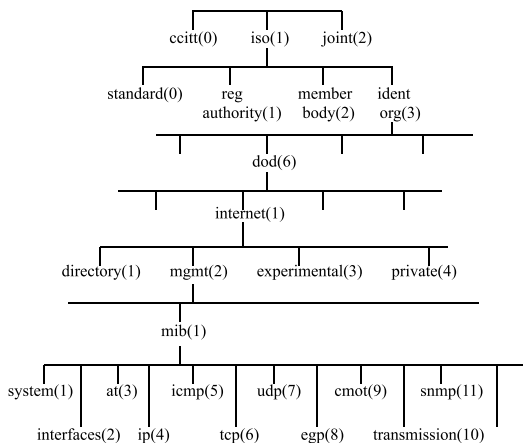


図 2. MIB のツリー構造

2.3 ネットワークトラフィック計測

トラフィック量は、通信回線にどれだけデータ量が送受信されたかを表す指標である。管理対象ノードあるインターフェースを介して送受信されたデータ量は、インターフェースの送信、受信ごとに、最大値 $2^{32}-1$ のカウンタにMIBの累計情報として記憶される。カウンタに記憶された値は、最大値をこえるとゼロに戻る。インターフェースのオクテット数を累積しているMIBのノードは、

iso.org.dod.internet.mgmt.mib.interfaces.ifInOctets,
ifOutOctets(1.3.6.1.2.1.2.10.16)

にインターフェース番号が振られる形となる。例えば、3

番目のインターフェースが受信した総オクテット数を調べると、

iso.org.dod.internet.mgmt.mib.interfaces.ifInOctets.

3(1.3.6.1.2.1.2.10.3)

をポーリングする事で計測できる。ネットワーク管理ステーションからある間隔ごとにこの総オクテット数を計測する事でその間隔のインターフェースにおけるトラフィック量を監視できる。すなわち、ポーリング操作を行った場合、データ系列はその累積値として得られる為、そのサンプリング間隔におけるトラフィック量を求める場合は、1サンプル前の値を差分して得られる。このサンプル間隔を短くする事でどれだけトラフィックが生じているかが計測できるが、MIBの更新時間や、短時間でポーリング操作ではトラフィックが大量に生じてしまう事を考慮して設計する必要性が求められる。

3 トラフィック時系列

図3に1999年12月10日の本学小金井キャンパス内研究セグメントにおける1日分のトラフィック時系列を示す。トラフィック量はSNMPにより5分ごとにネットワーク管理ステーションである本研究室内のマシンから管理情報ノードである本学各ルータを参照し計測した。

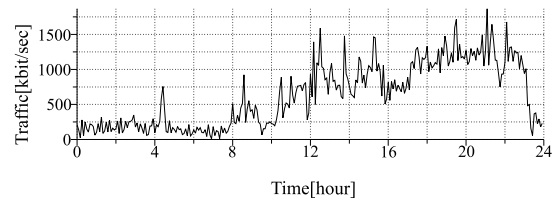


図 3. ネットワークトラフィック時系列

ネットワークトラフィックは各端末からネットワークに対し応答を要求することで生じ、本学のような学校機関などでは日々の授業などの関係でその応答は様々であり、図3を見ても明らかのように日中では授業等の関係でトラフィック量が多く、夜間ではネットワークを利用する人数が少ない為日中に比べてトラフィック量は少ない。また、HTTP(Hyper Text Transfer Protocol)を用いると多くの場合大量なパースト的な転送がみられ、これはネットワークトラフィックの不規則変動として現れる。このように、ネットワークトラフィックは日内変動や不規則変動などの変動成分が複合している事が知られている。

4 トラフィック差分系列の自己相関関数

これまで我々は最適線形フィルタやニューラルネットなどの手法でトラフィック時系列の予測を試みた。しかし、これらは現時点を次点の予測値とする予測(以後シフト予測)よりRMS誤差を比較した限り予測精度が向上されなかった。これはトラフィック時系列の不規則変動成分が短時間的には無相関に近い統計的性質を示しそのままの形では予測に向かない事に起因する。そこで本稿ではトラフィック時系列の差分系列つまり変化量に注目し予測を行った。

本章では、差分系列の予測を行う前にまず差分系列の自己相関関数について述べる。

図4は、図3のトラフィック時系列の差分系列および差分系列の自己相関関数を示している。差分系列は White Gaussian Noise のような周期性のない信号に見受けられるが、差分系列の自己相関関数を調べると1点目に一貫して見られる負の相関がある事ことがわかる。

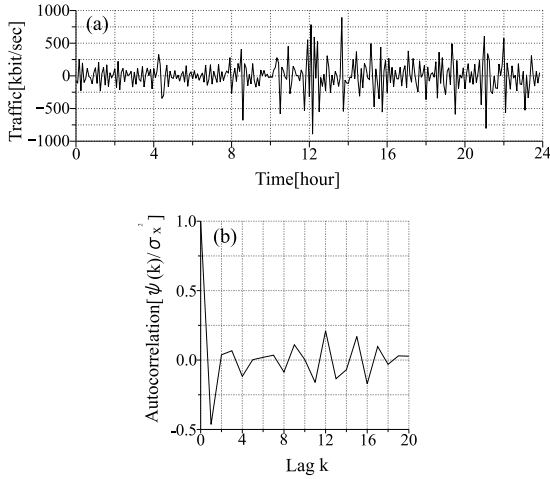


図4. (a) トラフィック時系列の差分系列 (b) 差分系列の自己相関関数

この差分系列の自己相関関数における一点目の負の相関であるが、時系列データに複数見られるパルス状突発的トラフィック(以後スパイク)が大きく影響していることが考えられる。もし、この一点目の負の相関がスパイクによるものであれば、差分系列を予測に用いることはできない。そこで、トラフィック時系列からスパイクの除去を行なった。スパイクの除去は、元データをメディアンフィルタを通したものと、元データを比べ、その差分が設定した閾値より大きいものをスパイクとし、その時の値をメディアンフィルタを通したものと置き換えることで行なった。

$$\text{if } |x[n] - y[n]| > \theta \text{ then } x[n] = y[n] \quad (1)$$

ここで、 $x[n]$ が元データ、 $y[n]$ が7点のメディアンフィルタ出力、 θ は閾値である。本稿では閾値を300kbitとしてスパイクの除去を行なった。

図5(a)がスパイクが除去されたトラフィック時系列、(b)がスパイクを除去したトラフィック時系列からもとめた差分系列の自己相関関数である。スパイクを除去しても、差分系列の自己相関関数の一点目に負の相関が現れている。これは、差分系列が常に増減を繰り返し振動的であることを示しこれは予測の対象となりうる。

また差分系列の自己相関関数は、図6に示すようにサンプリング周期を5分、10分、15分、20分と変化しても負の相関が見られる。これは、差分系列が自己相似性を持つ事を示唆していると考えられる [1]。

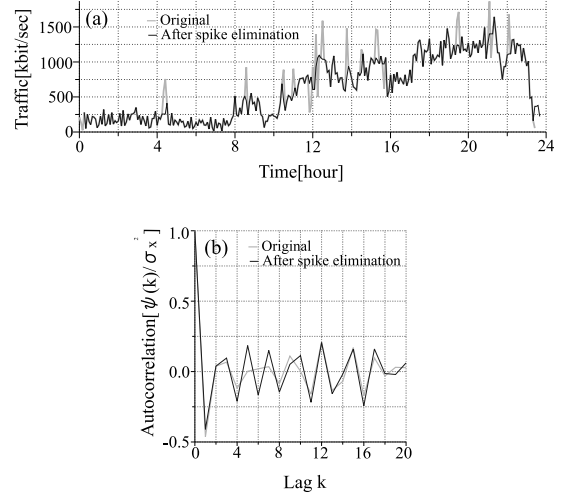


図5. (a) トラフィック時系列におけるスパイク除去 (b) スパイクを除去前後の差分系列の自己相関関数

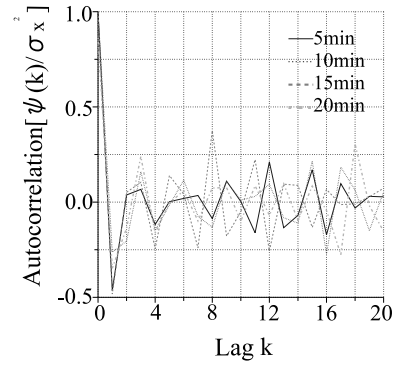


図6. サンプリング間隔をかえた差分系列の自己相関関数

5 トラフィック差分系列を用いた予測

前章で示したように、トラフィック差分系列は1点目に負の相関を有することから、本稿では差分系列を予測することでトラフィック時系列を予測する。今トラフィック時系列を $x[n] (n = 0, 1, \dots, N)$ とすると差分系列 $\Delta x[n]$ は

$$\Delta x[n] = x[n] - x[n-1] \quad (2)$$

となる。次に予測器により予測差分系列 $\Delta \hat{x}[n+1]$ を予測する。本稿では予測器として最適線形フィルタを用いた。

予測差分系列 $\Delta \hat{x}[n+1]$ は過去の入力系列 $\Delta x[n-i]$ を用い次式のように表される。

$$\Delta \hat{x}[n+1] = \sum_{i=0}^N a_i \Delta x[n-i] \quad (3)$$

上式において N は予測次数、 a_i は予測係数である。時刻 $n+1$ における差分系列 $\Delta x[n+1]$ と予測差分系列 $\Delta \hat{x}[n+1]$ との差が予測誤差信号 $d[n+1]$ となり

$$d[n+1] = \Delta x[n+1] - \Delta \hat{x}[n+1]$$

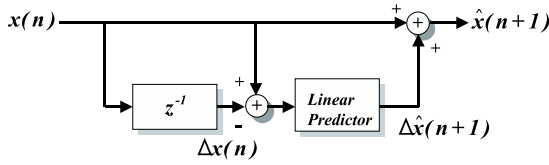


図 7. トラフィック予測システムブロック図

$$= \Delta x[n+1] - \sum_{i=0}^N a_i \Delta x[n-i] \quad (4)$$

で表され、最適予測係数は $d[n+1]$ の分散 σ_d^2 を最小とする a_i の組でありこれは

$$\begin{aligned} \frac{\partial \sigma_d^2}{\partial a_i} &= -2E \left[\Delta x[n+1] - \sum_{i=0}^N a_i \Delta x[n-i] \Delta x[n-i] \right] \\ &= -2E \left[d[n+1] \Delta x[n-i] \right] \\ &= 0 \end{aligned} \quad (5)$$

として求められ、予測誤差系列 $d[n+1]$ の分散 σ_d^2 は、自己相関関数を用いて

$$\sigma_d^2 = \sum_{i=0}^N \sum_{j=0}^N a_i R_{ij} a_j \quad (6)$$

として算出される。最適予測の条件は誤差の 2 乗平均値を最小にすることによるので、上式を a_i で偏微分し 0 と置くと正規方程式

$$\sum_{i=1}^N a_i R_{ij} = R_{0j} \quad (j = 1, 2, \dots, N) \quad (7)$$

を得、これを解き $\Delta \hat{x}[n+1]$ を予測した。

予測差分系列 $\Delta \hat{x}[n+1]$ は現時点のトラフィック時系列に加算され予測トラフィック系列 $\hat{x}[n+1]$ が算出される。つまり

$$\hat{x}[n+1] = x[n] + \Delta \hat{x}[n+1] \quad (8)$$

以上のブロック図を図 7 に示す。

6. トラフィック時系列予測結果

本章では実測トラフィックデータにおける本手法によるトラフィック時系列の予測結果について述べ、シフト予測手法との比較を行なう。

トラフィック時系列は 5 分毎に採取される本学小金井キャンパス内研究セグメントへのトラフィック量であり、図 11 は 1999 年 10 月 22 日 (金) の計測時間 24 時間のトラフィック時系列を本手法により予測を行なった結果である。図 11 は (a) 観測トラフィック時系列 $x[n]$ 、(b) 差分時系列 $\Delta x[n]$ (細線) とその予測時系列 $\Delta \hat{x}[n+1]$ (太線) を示す。ここで線形

表 1. 予測誤差の比較

予測信号	RMS 誤差 [kbit]
$\hat{x}[n+1]$	202.448985
$\hat{x}_s[n+1]$	227.953449

予測フィルタにおいては自己相関関数において 1 点後に高い相関値を示す事を考慮しフィルタ長は 1 点とした。さらに、(c) 差分予測誤差時系列、(d) 観測トラフィック時系列 $x[n]$ (細線) と予測トラフィック時系列 $\hat{x}[n+1]$ (太線)、(e) 予測誤差時系列、(f) 観測トラフィック時系列 $x[n]$ (細線) とシフト予測法 (太線) つまり

$$\hat{x}_s[n+1] = x[n] \quad (9)$$

による予測時系列、(g) にその予測誤差時系列を示している。

予測トラフィック時系列 $\hat{x}[n+1]$ とシフト予測時系列 $\hat{x}_s[n+1]$ を RMS 誤差で比較すると表 1 のように本手法はシフト予測法よりも良好に予測している事がわかる。ところが、差分時系列の予測誤差を見ると差分予測時系列 $\Delta \hat{x}[n+1]$ は差分時系列 $\Delta x[n]$ の良い予測値とはなっていない。これは自己相関関数において図 4(b) から明らかなように差分時系列が広帯域信号である事に起因すると考えられるが、差分予測時系列の分散が小さいため、予測トラフィックに還元した時に大きな変動にならない事、さらに、最適フィルタにより差分時系列の自己相関における 1 点後の負の相関を保っていることで変動の方向を正しい推定している事で予測トラフィック時系列はシフト予測法よりも誤差が小さくなると考えられる。

表 2. 1999 年 10-12 月における予測誤差の比較

Data	RMS 誤差 [kbit]	
	$\hat{x}[n+1]$	$\hat{x}_s[n+1]$
Data1	262.295166	275.702797
Data2	311.034546	311.249300
Data3	206.588548	233.361632
Data4	225.053683	241.944308
Data5	245.089213	258.627773
Data6	230.586041	269.494985
Data7	207.775973	238.777773
Data8	230.028940	265.015063
Data9	186.060363	202.600420
Data10	220.339164	240.220206

表 3. 予測誤差の比較

予測信号	平均 RMS 誤差 ± 標準偏差 [kbit]
$\hat{x}[n+1]$	232.485164 ± 32.964127
$\hat{x}_s[n+1]$	253.699426 ± 27.866163

次に、同計測点における 1999 年 10-12 月の観測トラフィック 10 例についての RMS 誤差を表 2 に、またその平均 RMS 誤差を表 3 に示す。個々の RMS 誤差についても本手法がシフト予測法よりも RMS 誤差が小さい事がわかる。表 2 中の Data5 は比較的 RMS 誤差が大きくシフト予測法と RMS 誤差に近い例であるが、このような例では差分時系列の自己相関関数を見ると 1 点目に負の相関が見られない事から本手法では良好な予測が得られない事がわかる (図 8 参照)。このような例が 10 例中 1 例存在するものの全体として平均 RMS 誤差はシフト予測よりも良好に予測していると言える事がわかる。

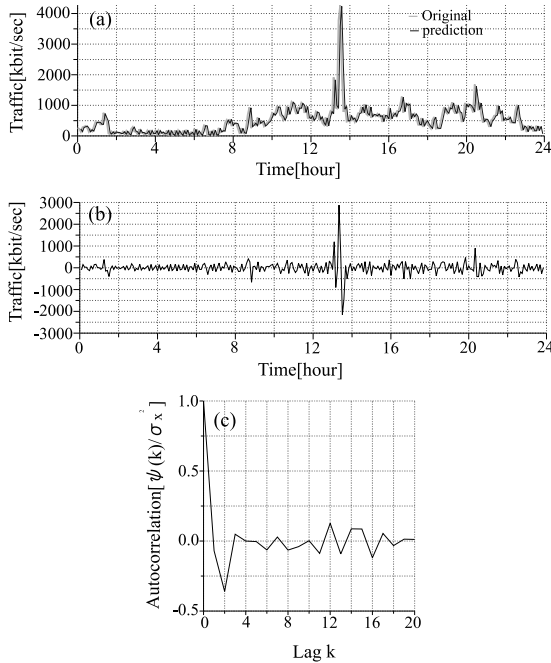


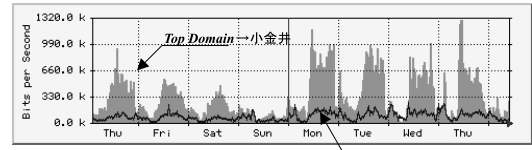
図 8. 予測精度が悪い例 (a) 元データ及び予測トラフィック時系列 (b) 予測誤差時系列 (c) 差分系列の自己相関関数

7 トラフィック予測の経路制御への応用

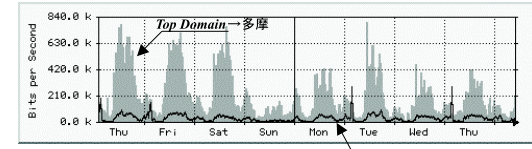
次にトラフィック時系列を正確に予測する事が可能になったとした時、予測をどのように応用するかについて、本章では一例として経路制御への応用を考え、トラフィック予測後の経路制御への応用について考察する。

ネットワーク設計において各経路のトラフィックを均一化し負荷分散をはかることは重要な課題である。従って、ネットワーク経路の帯域は設計時に慎重に評価し決定しなければならない。しかし、必要とされる帯域を全ネットワーク経路についてあらかじめ評価することは容易ではない。また、要求帯域は時間経過と共に変化し設計時の見積もり帯域と実情が合わなくなることも少なくないと考えられる。さらに、日内変動も大きく時々刻々要求帯域は変化している。このような状況において、トラフィック経路を動的に制御することが出来ればネットワーク負荷分散の意味で有用であると思われる。ここでは、現状多くのネットワークで WAN から LAN へ向けた HTTP プロトコルデータが主たるトラフィックとなっていることを考慮し、ネットワーククライアントにおける proxy サーバの振り分けによる動的な制御が可能であること、またトラフィック予測がこの制御にいかに応用されるかを示す。図 9 は法政大学学術教育ネットワークにおけるトラフィックデータである。ネットワーク構成は 3 キャンパスが互いに T1(1.5Mbps) デジタル専用線で結ばれトップドメイン (市ヶ谷) から 6Mbps の帯域で SINET へ接続されている。現在主たるトラフィックは SINET から学内への HTTP プロトコルデータである。図 9 は 8 日間のキャンパス間のトラフィックを計測したものであり週末を挟んで小金井キャンパスの教育用情報端末の利用する proxy サーバを変更した。すなわち、前週において小金井端末の

小金井、トップドメイン間



多摩、トップドメイン間



小金井、多摩間

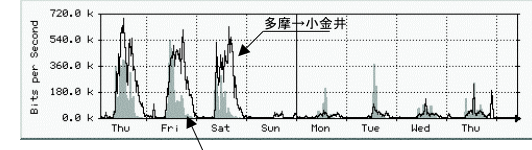


図 9. proxy の振り分けによるトラフィック変化

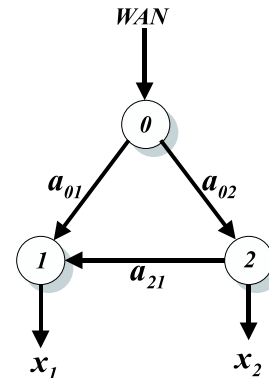


図 10. HTTP proxy の確率的振り分けによるトラフィック制御

HTTP proxy を 6 : 4 でトップドメインと多摩キャンパスに振り分けていたものを後週で 1 : 0 すなわち多摩へのトラフィックバイパスを止めた。このことによって大きくトラフィックが変化している様子がわかる。このことは、主たるトラフィックが HTTP プロトコルによるものであること、及びその結果 HTTP proxy の変更によって大きくトラフィック制御が可能であることを示している。この proxy の動的振り分けとトラフィック予測を組み合わせることで示すようなトラフィック制御が可能となる。

図 10 においてノード 0, 1, 2 はそれぞれ LAN を構成しノード間は互いに接続されている。ノード 0 が WAN との接続点となっている。WAN からノード 1, 2 への HTTP トラフィックに着目する。ノード i から j へのトラフィックを $a_{i,j}$ (bps) としノード i, j が要求する総トラフィックをそれぞれ x_i, x_j とする。 $x_1 > x_2$ の時、ノード 1 の proxy を $1 : k$ でノード 0 と 2 に振り分け、ノード 2 の proxy を $1 : 0$ でノード 0 と 1 に振り分ける。つまり $a_{21} = k a_{01}$, $a_{12} = 0$ とする。ここで $k = (x_1 - x_2) / (x_1 + x_2)$ と設定すれば $a_{01} = a_{02}$ となりノード 1, 2 間の接続をバイパスとして主たる経

路 $0 \rightarrow 1$, $0 \rightarrow 2$ のトラフィックを平均化することができる。 $x_1 < x_2$ の場合は対称的にノード 2 の proxy をノード 0 と 1 に同様に振り分けることによりトラフィックの均一化が図られる。 x_1 及び x_2 は SNMP により一定時間毎の観測が可能であり、前節で述べた手法で予測を行い予測精度に応じて正確な制御が行われることとなる。平均的な週内変動を基本値として直近のデータの平均による修正等による補正を行う程度でも十分実用的な効果が期待される。

8 むすび

本稿ではトラフィック時系列において、差分系列の自己相関関数が 1 点目に負の相関を有する事を利用して、差分系列を予測しトラフィック時系列を予測した。その結果現時点を次点の予測値とするシフト法よりも予測精度が向上する事を示した。しかしながら、比較対象より数 % 予測精度が向上した事を示したに過ぎず予測誤差としては以前高い数値であることから今後更なる改良が必要であると考えている。また、予測精度の向上のためにトラフィック時系列の統計的性質を明らかにすることも必要であろう。本稿で取り上げた予測の応用として取り上げた経路制御は応用の一例に過ぎないが種々のネットワークトポロジーにおいてここで提案する proxy の動的振り分けによるトラフィック制御は有効に応用されるであろう。

参考文献

- [1] W.E. Leland, W. Willinger, M.S. Taqqu, and D.V. Wilson, "On the self-similar nature of Ethernet traffic," SIGCOMM '93, pp.183-193, 1993.
- [2] 小沢 利久, 町原 文明, 石橋 圭介, "マルチメディアトラフィック理論の最新動向," 信学誌, vol.81, no.5, pp.506-515, May. 1998.
- [3] W. Willinger, M.S. Taqqu, R. Sherman, and D.V. Wilson, "Self-similarity through high-variability : statistical analysis of Ether LAN traffic at the source level," SIGCOMM '95, pp.100-113, 1995.
- [4] E.S. Yu and C.Y.R.Chen, "Traffic Prediction Using Neural Networks," in Proc. IEEE GLOBECOM'93, pp.991-995, 1993.
- [5] Stephan Robert, Jean-Yves Le Boudec, "On a Markov modulated chain exhibiting self-similarities over finite timescale," Performance Evaluation 27&28, pp.159-174, 1996.
- [6] 萩原 幸平, 菊地 善行, 治部 将之, 篠 秀明, 八名 和夫, "ネットワークトラフィックの統計的モデル," 統計数理研究共同研究リポート, vol.114, pp104-110, Dec. 1998.
- [7] 篠 秀明, 菊地 善行, 治部 将之, 水田 博久, 八名 和夫, "ネットワークトラフィック累積系列の予測と応用," 1999 年春季電子通信学会総合大会
- [8] 土井 博生, 戸出 英樹, 池田 博昌, "自己相関性を持つトラフィック特性の検討," 1997 年電子情報通信学会総合大会, B-7-200, p.329
- [9] 岩本 厚志, 戸出 英樹, 池田 博昌, "トラフィック予測における線形予測とニューラル予測の比較," 1997 年電子情報通信学会総合大会, B-7-165, p.294
- [10] 戸出 英樹, 岩本 厚志, 池田 博昌, "ニューラルネットワークによるトラフィック予測の検討," 信学技報, pp.69-74, Jun. 1997

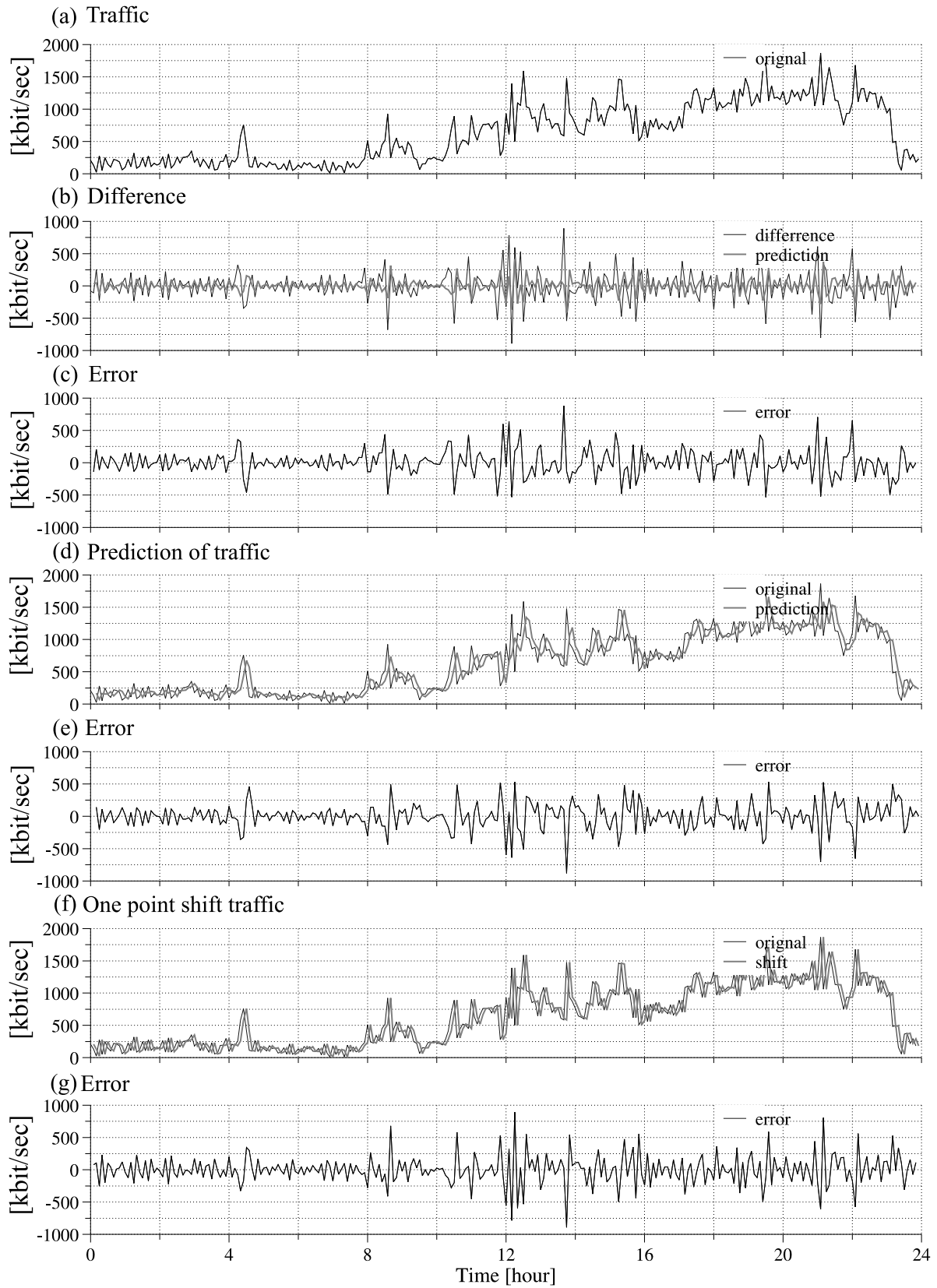


図 11. (a) 観測トラフィック時系列 (b) 差分時系列と予測時系列 (c) 差分予測誤差時系列 (d) 予測トラフィック時系列 (e) 予測誤差時系列 (f) シフト予測時系列 (g) シフト予測誤差時系列

キーワード.

ネットワークトラヒック、線形予測、トラヒック分散制御

Summary.

Network Traffic Prediction and its Application

Hideaki Shino Kenji Awa Tomofumi Okubo Kazuo Yana

Department of Electrical Infomatics , Hosei University

Growing needs for network communication requires efficient traffic control. Network traffic prediction will play a key role for the traffic control. This paper proposes a method for the network traffic prediction using the traffic data obtained by the simple network management protocol (SNMP). The method is based on the presence of short term negative correlation in the difference time series. As an application, This paper proposes a method of traffic control using the traffic prediction and adaptive cache selection.

Keywords.

Network Traffic , Linear Prediction , Traffic Control