

極値理論を用いたトラフィックピークレート推定の 誤差低減手法

下川 尚人^{†1} 内田 真人^{†2} 鶴 正人^{†2}

ネットワーク品質の最悪状態を予測することは、ネットワークの設計・制御・管理において重要である。そのため、トラフィックレートが従う確率分布の裾野領域の振る舞いを把握する必要がある。そこで本研究では、極端な事象の振る舞いを記述するための統計理論である極値理論 (Extreme Value Theory) を用いたトラフィックピークレートの推定法について検討する。さらに、重み係数 α を逐次変更することのできる指数移動平均を適用することで推定誤差を低減する手法を提案する。実トラフィックデータを用いた評価により、提案手法を適用することで、推定に利用できるトラフィックデータが不足した場合においても誤差の低減が可能であることが明らかとなった。

Peak traffic rate estimation method of error reduction using extreme value theory

NAOTO SHIMOKAWA,^{†1} MASATO UCHIDA^{†2}
and MASATO TSURU^{†2}

It is important to estimate the worst quality of network performance for network design, control and management. Therefore, it is necessary to understand the traffic behavior when its volume becomes very large compared to usual status, i.e., the tail behavior of probability distribution of traffic rate. In this paper, we estimate traffic peak rate using an extreme value theory, which is a statistical theory for describing the behavior of extreme events. In addition, we propose a method to decrease the estimation error by using an exponential moving average, where the weight parameter, α , can be sequentially modified. As a result of experimental evaluation with real traffic traces, we show that the estimation error can be decreased by using the proposed method especially even when sufficient amount of data is not available for the estimation of traffic peak rate.

1. 研究背景

ネットワーク品質の最悪状態を予測することは、ネットワークの設計・制御・管理において重要である。そのため、トラフィック流量が平常時に比べて非常に大きくなる場合の振る舞い、すなわち、トラフィックレートが従う確率分布の裾野領域の振る舞いを把握する必要がある。ところが、トラフィックレート推定に関する既存の研究は、その全体的な挙動に注目したものが主流であり、例外的な挙動に注目したものは少なかった。そこで本研究では、極端な事象の振る舞いを記述するための統計理論である極値理論 (Extreme Value Theory) を用いたトラフィックピークレートの推定法について検討する。同様のアプローチによる検討には、ネットワークスイッチの省電力化への応用を目的としたもの¹⁾があったが、推定の際に十分多くのデータを利用できることが前提とされており、推定に利用できるデータの多寡や、推定する期間の長短が推定誤差に及ぼす影響については検討されていなかった。また、推定の誤差を低減する手法についても検討されていなかった。以上の背景から、本研究では、トラフィックピークレート推定のための既存手法¹⁾に対し、重み係数を逐次変更することのできる指数移動平均を適用することによって推定誤差を低減する手法を新たに提案し、実トラフィックデータを用いて提案手法を評価する。

2. 既存手法

既存手法¹⁾におけるトラフィックピークレート推定の概要を以下に示す (図 1 参照)。まず、単位時間あたりのトラフィック流量 (以下、トラフィックレート) を予め定められた計測単位時間毎に計算し、トラフィックレートの時系列データを取得する。本研究では、計測単位時間を 10 ms とし、トラフィックレートの単位は Mbps とする。以下、分析対象となるトラフィックデータが計測された期間を分析期間と呼ぶ。既存手法¹⁾では、分析期間におけるトラフィックレートの時系列データを用いて、その直後に設定された推定期間における最大のトラフィックレート (以下、トラフィックピークレート) を、以下の手順で推定する。ただし、推定期間は計測単位時間で割り切れるものとする。

^{†1} 九州工業大学情報工学部電子情報工学科
Department of Computer Science and Electronics, Computer Sciences And System Engineering,
Kyushu Institute of Technology

^{†2} 九州工業大学ネットワークデザイン研究センター
Network Design Research Center, Kyushu Institute of Technology

まず、分析期間を推定期間と同じ長さのサンプル期間に分割する。ただし、分析期間はサンプル期間で割り切れるものとし、分析期間に含まれるサンプル期間数は n 個とする。また、 i 番目のサンプル期間における j 番目のトラヒックレートを $z_{i,j}$ と定義すると、 i 番目のサンプル期間でのトラヒックレートの集合は

$$(z_{i,1}, z_{i,2}, \dots, z_{i,m}), \quad i = 1, 2, \dots, n \quad (1)$$

と表せる。次に、各サンプル期間内におけるトラヒックピークレートを取得（サンプル）する。サンプル期間 i におけるトラヒックピークレートは

$$z_i = \max(z_{i,1}, z_{i,2}, \dots, z_{i,m}), \quad i = 1, 2, \dots, n \quad (2)$$

となる。この操作を分析期間に含まれるすべてのサンプル期間において繰り返すことでトラヒックピークレートのデータ集合

$$(z_1, z_2, \dots, z_n) \quad (3)$$

を取得することができる。一方、極値理論の結果によると、取得したトラヒックピークレートのデータ集合の要素が従う確率分布は

$$\Pr\{X \leq x\} = \exp\left\{-\left(1 + \xi \frac{x - \mu}{\psi}\right)^{-1/\xi}\right\}, \quad 1 + \xi \frac{x - \mu}{\psi} > 0 \quad (4)$$

と表される一般化極値分布により表すことができる。そのため、既存手法においては、式 (4) のパラメータ ξ, μ, ψ を推定した上で、その期待値を推定期間におけるトラヒックピークレートの推定値としていた。

以下、式 (4) で与えられる一般化極値分布のパラメータ ξ, μ, ψ を推定する手順、及び期待値の計算方法について説明する。このパラメータ推定手法は確率重み付きモーメント法 (Probability Weighted Moment Method) と呼ばれる。はじめに、式 (3) で与えられるトラヒックピークレートのデータ集合の要素を昇順で並び替えたものを

$$(x_1, x_2, \dots, x_n) \quad (5)$$

とする。次に、以下の値を計算する。

$$\hat{w}_0 = \frac{1}{n} \sum_{j=1}^n x_j \quad (6)$$

$$\hat{w}_1 = \frac{1}{n} \sum_{j=2}^n x_j \left(\frac{j-1}{n-1}\right) \quad (7)$$

$$\hat{w}_2 = \frac{1}{n} \sum_{j=3}^n x_j \left(\frac{j-1}{n-1} \frac{j-2}{n-2}\right) \quad (8)$$

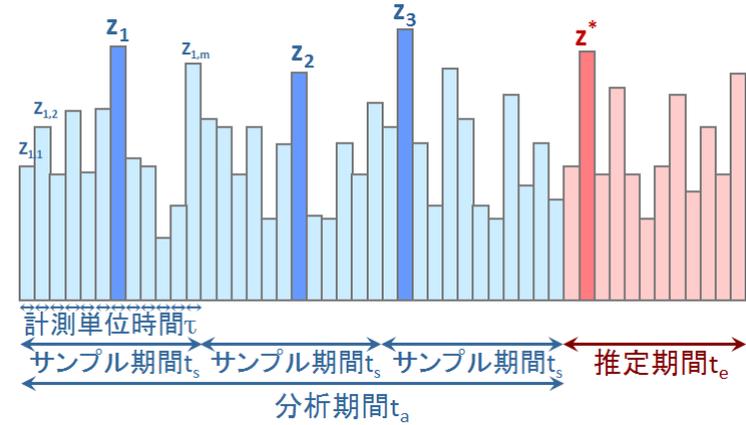


図1 既存手法の概要

このとき、パラメータ ξ の推定値 $\hat{\xi}$ は以下の関係を満たすことが知られている。

$$\frac{3\hat{w}_2 - \hat{w}_0}{2\hat{w}_1 - \hat{w}_0} = \begin{cases} \frac{\log 3}{\log 2}, & \hat{\xi} = 0 \text{ のとき} \\ \frac{3^{\hat{\xi}} - 1}{2^{\hat{\xi}} - 1}, & \hat{\xi} \neq 0 \text{ のとき} \end{cases} \quad (9)$$

この関係を満たす $\hat{\xi}$ を求めるためには、式 (9) の左辺と右辺の差の 2 乗が最小になる $\hat{\xi}$ を求めればよい。すなわち

$$\text{式 (9) の左辺と右辺の差の 2 乗} = \begin{cases} \left(\frac{3\hat{w}_2 - \hat{w}_0}{2\hat{w}_1 - \hat{w}_0} - \frac{\log 3}{\log 2}\right)^2, & \hat{\xi} = 0 \text{ のとき} \\ \left(\frac{3\hat{w}_2 - \hat{w}_0}{2\hat{w}_1 - \hat{w}_0} - \frac{3^{\hat{\xi}} - 1}{2^{\hat{\xi}} - 1}\right)^2, & \hat{\xi} \neq 0 \text{ のとき} \end{cases} \quad (10)$$

が最小になる $\hat{\xi}$ を求めればよい。具体的には、インターネットトラヒックの特性から $-1 \leq \hat{\xi} \leq 1$ であることが予想されることを考慮して、この範囲内で 0.001 刻みで $\hat{\xi}$ の値を変化させて式 (10) を計算することで、これが最小になる $\hat{\xi}$ を近似的に求めることがで

きる．また，パラメータ ψ, μ の推定値 $\hat{\psi}, \hat{\mu}$ は

$$\hat{\psi} = \begin{cases} \frac{2\hat{w}_1 - \hat{w}_0}{\log 2}, & \hat{\xi} = 0 \text{ のとき} \\ \frac{(2\hat{w}_1 - \hat{w}_0)\hat{\xi}}{\Gamma(1-\hat{\xi})(2^{\hat{\xi}} - 1)}, & \hat{\xi} \neq 0 \text{ のとき} \end{cases} \quad (11)$$

$$\hat{\mu} = \begin{cases} \hat{w}_0 - \hat{\psi}\gamma, & \hat{\xi} = 0 \text{ のとき} \\ \hat{w}_0 + \frac{\hat{\psi}}{\hat{\xi}}(1 - \Gamma(1 - \hat{\xi})), & \hat{\xi} \neq 0 \text{ のとき} \end{cases} \quad (12)$$

のように与えられる．ただし， $\gamma \simeq 0.5772156649$ はオイラーの定数である．

上記のようにして求めたパラメータ推定値 $\hat{\xi}, \hat{\mu}, \hat{\psi}$ の結果を用いると，推定期間におけるトラヒックピークレートの期待値は

$$z_{\hat{\xi}, \hat{\mu}, \hat{\psi}} = \begin{cases} \hat{\mu} + \hat{\psi}\gamma, & \hat{\xi} = 0 \text{ のとき} \\ \hat{\mu} - \frac{\hat{\psi}}{\hat{\xi}}(1 - \Gamma(1 - \hat{\xi})), & \hat{\xi} \neq 0 \text{ のとき} \end{cases} \quad (13)$$

と推定できる．既存手法¹⁾では，推定期間におけるトラヒックピークレートの推定値として，式(13)で与えられる期待値を採用している．

3. 提案手法

次節で示すように，既存手法では，推定期間が長くなる場合や分析期間長が短くなる場合等，推定に利用できるデータ数が不足すると誤差が大きくなる問題がある．そこで本研究では，この問題を改善するために，既存手法に対して指数移動平均を適用した改良手法を提案する． k 回目の推定における既存手法における推定値を $z_{\hat{\xi}^{(k)}, \hat{\mu}^{(k)}, \hat{\psi}^{(k)}}$ とすると，指数移動平均を用いた推定値 $\bar{z}_{\hat{\xi}^{(k)}, \hat{\mu}^{(k)}, \hat{\psi}^{(k)}}$ は

$$\bar{z}_{\hat{\xi}^{(k)}, \hat{\mu}^{(k)}, \hat{\psi}^{(k)}} = \alpha z_{\hat{\xi}^{(k)}, \hat{\mu}^{(k)}, \hat{\psi}^{(k)}} + (1 - \alpha) \bar{z}_{\hat{\xi}^{(k-1)}, \hat{\mu}^{(k-1)}, \hat{\psi}^{(k-1)}} \quad (14)$$

と求めることができる．ただし， $\bar{z}_{\hat{\xi}^{(0)}, \hat{\mu}^{(0)}, \hat{\psi}^{(0)}} = z_{\hat{\xi}^{(1)}, \hat{\mu}^{(1)}, \hat{\psi}^{(1)}}$ とする．式(14)を用いることで，重み係数 α によって直前の推定値と過去の推定値に重みを調整することができ，それによって直前の推定結果を補正することができる．なお， $\alpha = 1$ の場合は既存手法と等価である．本研究では，重み係数 α を固定値とする場合を，固定重み指数移動平均法と呼ぶ．また，予測値と実測値との差の絶対値を誤差とした場合における勾配法によって，重み係数 α を逐次変更する手法を，変動重み指数移動平均法と呼ぶ．変動重み指数移動平均法

においては

$$\alpha_{k+1} = \alpha_k - \begin{cases} \varepsilon \frac{d}{d\alpha_k} |z_k^* - \bar{z}_{\hat{\xi}^{(k)}, \hat{\mu}^{(k)}, \hat{\psi}^{(k)}}| & z_k^* \neq \bar{z}_{\hat{\xi}^{(k)}, \hat{\mu}^{(k)}, \hat{\psi}^{(k)}} \text{ のとき} \\ 0 & z_k^* = \bar{z}_{\hat{\xi}^{(k)}, \hat{\mu}^{(k)}, \hat{\psi}^{(k)}} \text{ のとき} \end{cases} \quad (15)$$

に基づいて重み係数 α を逐次変更する．ただし， z_k^* は k 回目の推定期間におけるトラヒックピークレートの実測値である．なお

$$\frac{d}{d\alpha_k} |z_k^* - \bar{z}_{\hat{\xi}^{(k)}, \hat{\mu}^{(k)}, \hat{\psi}^{(k)}}| = \begin{cases} -\varepsilon (z_{\hat{\xi}^{(k)}, \hat{\mu}^{(k)}, \hat{\psi}^{(k)}} - \bar{z}_{\hat{\xi}^{(k-1)}, \hat{\mu}^{(k-1)}, \hat{\psi}^{(k-1)}}), & z_k^* > \bar{z}_{\hat{\xi}^{(k)}, \hat{\mu}^{(k)}, \hat{\psi}^{(k)}} \text{ のとき} \\ \varepsilon (z_{\hat{\xi}^{(k)}, \hat{\mu}^{(k)}, \hat{\psi}^{(k)}} - \bar{z}_{\hat{\xi}^{(k-1)}, \hat{\mu}^{(k-1)}, \hat{\psi}^{(k-1)}}), & z_k^* < \bar{z}_{\hat{\xi}^{(k)}, \hat{\mu}^{(k)}, \hat{\psi}^{(k)}} \text{ のとき} \end{cases} \quad (16)$$

である．

一方，本研究においては，トラヒックピークレートの推定値として，式(13)で与えられる期待値以外に

$$z_{\hat{\xi}, \hat{\mu}, \hat{\psi}} = \begin{cases} \hat{\mu} - \hat{\psi} \log(\log 2), & \hat{\xi} = 0 \text{ のとき} \\ \hat{\mu} - \frac{\hat{\psi}}{\hat{\xi}} (1 - (\log 2)^{-\hat{\xi}}), & \hat{\xi} \neq 0 \text{ のとき} \end{cases} \quad (17)$$

によって与えられる中央値（メディアン，50パーセント点）を用いる場合についても検討する．これは，以下に示すように，予測値と実測値との差の絶対値を誤差としている場合に妥当となるためである．

真のトラヒックピークレートを表す確率変数を Z とし，それが従う確率密度関数を $p_Z(z)$ とおく．以下では， $\mathbb{E}[\cdot]$ は括弧内の確率変数の期待値を表す．このとき，トラヒックピークレートの推定値を θ とし，誤差を $|Z - \theta|$ と定義すると，その期待値は

$$\mathbb{E}[|Z - \theta|] = \int_0^\infty |x - \theta| p_Z(x) dx \quad (18)$$

$$= \int_\theta^\infty (z - \theta) p_Z(z) dz + \int_0^\theta (\theta - z) p_Z(z) dz \quad (19)$$

$$= \int_\theta^\infty z p_Z(z) dz - \int_0^\theta z p_Z(z) dz - \theta \Pr\{Z \geq \theta\} + \theta \Pr\{Z \leq \theta\} \quad (20)$$

$$= \int_0^\infty z p_Z(z) dz - 2 \int_0^\theta z p_Z(z) dz - \theta \Pr\{Z \geq 0\} + 2\theta \Pr\{Z \leq \theta\} \quad (21)$$

$$= \mathbb{E}[Z] - 2 \int_0^\theta z p_Z(z) dz - \theta + 2\theta \Pr\{Z \leq \theta\} \quad (22)$$

のように計算できる．したがって

$$\frac{d}{d\theta} \mathbb{E}[|Z - \theta|] = -2\theta p_Z(\theta) - 1 + 2\Pr\{Z \leq \theta\} + 2\theta p_Z(\theta) \quad (23)$$

$$= -1 + 2\Pr\{Z \leq \theta\} \quad (24)$$

となる．ここで

$$\frac{d}{d\theta} \mathbb{E}[|Z - \theta|] = 0 \quad (25)$$

とおくと

$$\Pr\{Z \leq \theta\} = \frac{1}{2} \quad (26)$$

がなりたつ．すなわち， θ は中央値である．以上より，誤差を $|Z - \theta|$ と定義した場合，その平均を最小化する θ は中央値となることが分かる．

4. 実験結果

4.1 実験概要

本節では，提案手法の評価に利用する実トラフィックデータ，実験の概要と評価方法について説明する．本研究では前章で説明した提案手法を，MAWI Working Group Traffic Archive が提供するトラフィックデータを使用して評価する．このデータは太平洋横断の 1Gbps のリンクで毎日 14 時から 15 分間計測したものであり，本実験では 2011 年 11 月の 1 か月間における平日 20 日分のデータを使用した．また，推定毎の誤差の計算方法は推定したピーク値と実測のピーク値の差の絶対値とする．さらに，各日のデータにおける固有のトラフィック特性に強く影響されることのないよう，20 日分のデータすべての誤差の平均を用いて評価した．また，分析に利用することのできるデータ量とそれに対する推定期間の長さの関係が推定精度に与える影響を評価するために，分析期間に含まれるサンプル期間の個数，もしくは分析期間の長さを固定した上で推定期間の長さを変更することによって評価を行った．具体的には，サンプル期間の個数を 5，10，15，20，25 個と変化させ，それぞれのサンプル期間数において，推定期間を 1 秒から 15 秒まで 1 秒刻みで変更して推定を行った．また，分析期間を 30 秒，1，2，3，4，5 分と変化させ，それぞれの分析期間長毎に推定期間を 1，2，3，5，6，10，15 秒と変化させて推定を行った．

4.2 実験結果・考察

4.2.1 固定重み指数移動平均法を適用した場合

固定重み指数移動平均法を適用した場合の結果を図 2 から 5 に示す．図 2 は分析期間内のサンプル期間数が 5 個の場合，図 3 は 25 個の場合である．また，図 4 は分析期間長が 30

秒の場合，図 5 は 5 分の場合である．これらの図においては，重み係数 α を 0.1 から 1.0 まで 0.1 刻みで変更して推定を行った結果をそれぞれ示している．図 2 と図 4 より，推定に利用できるデータが少ない場合は，推定期間によって誤差が最も小さくなる重み係数 α の値，すなわち，最適な重み係数 α の値が異なることが分かる．さらに，その最適な重み係数 α の値は比較的小さな値となっており，既存手法 ($\alpha = 1$ の場合) における誤差は大きいことが分かる．一方，図 3 と図 5 より推定に利用できるデータが多い場合は，最適な重み係数 α の値が 1.0 付近となっていることが分かる．このことから，推定に利用できるデータが十分ある場合には，過去の推定値を利用せずに，直前の分析期間における推定値のみを利用することが適切であると考えられる．以上のことから，固定重み指数移動平均法を適用する場合は，最適な重み係数 α の値が条件によって変化することが分かる．すなわち，固定重み指数移動平均法を用いて適切な推定を行うためには，最適な重み係数 α を予め設定する必要があると言える．そこで次節では，重み係数 α の値を逐次変更する変動重み指数移動平均法について検討する．

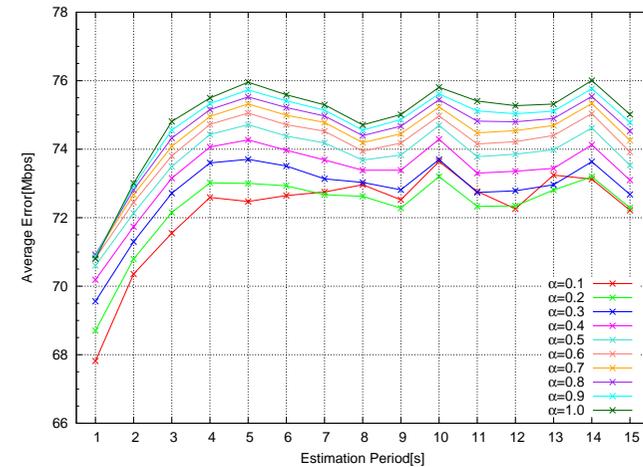


図 2 固定重み指数移動平均法 (サンプル期間数 5 個)

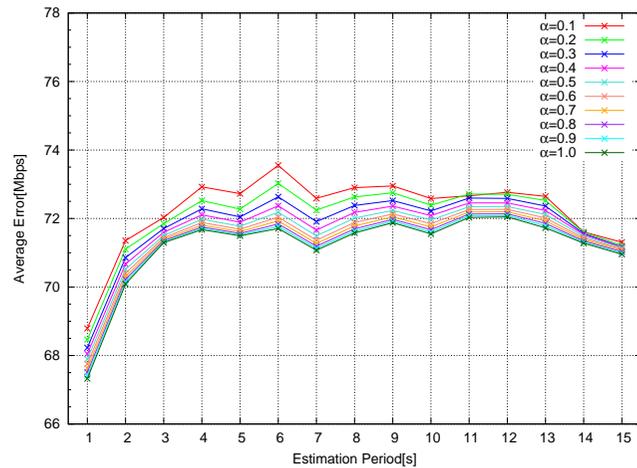


図 3 固定重み指数移動平均法 (サンプル期間数 25 個)

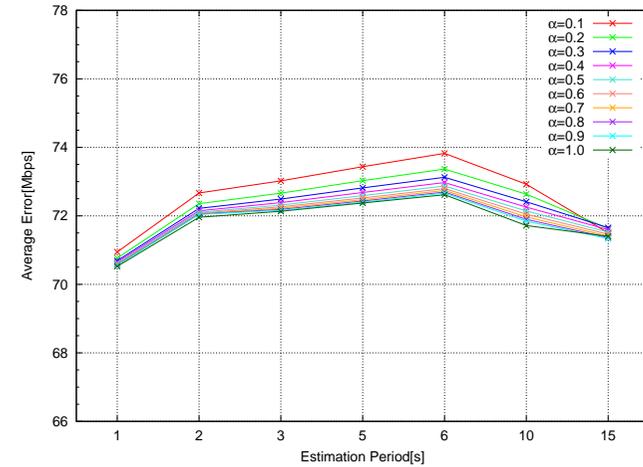


図 5 固定重み指数移動平均法 (分析期間長 5 分)

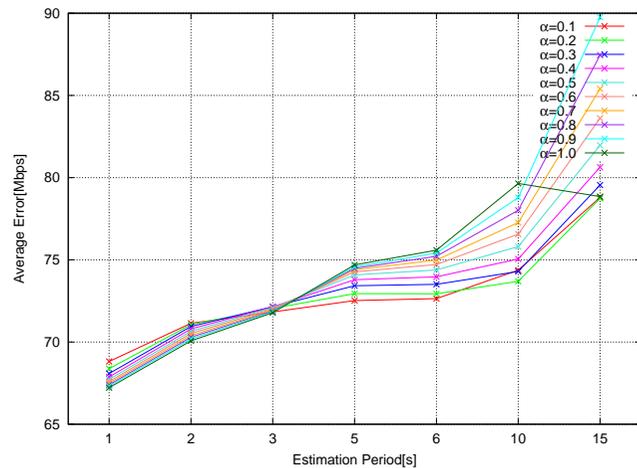


図 4 固定重み指数移動平均法 (分析期間長 30 秒)

4.2.2 変動重み指数移動平均を適用した場合

変動重み指数移動平均を適用した場合の推定結果を図 6 から図 9 に示す。ただし、重み係数 α の初期値は 1.0 とする。図 6 は分析期間内のサンプル期間数が 5 個の場合、図 7 は 25 個の場合である。また、図 8 は分析期間長が 30 秒の場合、図 9 は 5 分の場合である。さらに、図 6 と図 9 のそれぞれについて、推定期間が 3 秒と 5 秒の場合における、ある 1 日分のデータを用いたときの推定における α の推移を図 10 と図 11 に示す。なお、これらの図において、固定重み指数移動平均法との比較のために、対応する条件で、各推定期間において重み係数 α が 0.1 から 1.0 のうちで誤差が最大となった場合の誤差を Maximum Error、最小となった場合の誤差を Minimum Error として示している。

図 6 と図 8 より、変動重み指数移動平均法の方が既存手法に比べ誤差が低減できていることが分かる。これは、推定に利用できるデータ数が少ない場合は、固定重み指数移動平均法の場合と同様に、過去の推定値も利用する変動重み指数移動平均の方が、直前のデータのみ使用して推定を行う既存手法に比べ誤差が低減できるためであると考えられる。なお、図 8 においては、推定期間が 15 秒の場合において既存手法の誤差が小さくなっているが、これは、20 日中の 3 日分のデータを用いた場合の誤差が極めて小さくなった結果、20 日分の誤差の平均が小さくなったためであることを確認している。すなわち、この結果は、この

3日分のデータの固有の特性を反映したものであり、必ずしも一般的な特性を反映しているものであるとは言えないと考えられる。また、図7と図9においては、変動重み指数移動平均法を適用した場合と既存手法を適用した場合の誤差はほぼ同一であることが分かる。前節の結果より、推定に利用できるデータが十分ある場合には、過去の推定値を利用せずに、直前の分析期間における推定値のみを利用すること、すなわち既存手法を用いることが適切であることを明らかにしたが、変動重み指数移動平均法では、重み係数 α を逐次変更することによって、既存手法と同等の効果が得られていると考えられる。以上のことから、提案手法を用いることによって、推定に利用できるデータ数が少ない場合には誤差を低減でき、かつ、推定に利用できるデータ数が多い場合においては既存手法と同程度の推定精度を維持できているということが分かる。つまり、変動重み指数移動平均を適用することによって、どのような条件においても、既存手法に比べ同等かそれ以上の推定精度を達成できているということが分かる。実際、図10に示すように、推定に利用できるデータ数が少ない場合には重み係数 α の値を小さな値に変化させ、図11に示すように、推定に利用できるデータ数が多い場合には重み係数 α の値を最適地の 1.0 付近から大きく変更せずに既存手法に近い推定を行っていることが分かる。

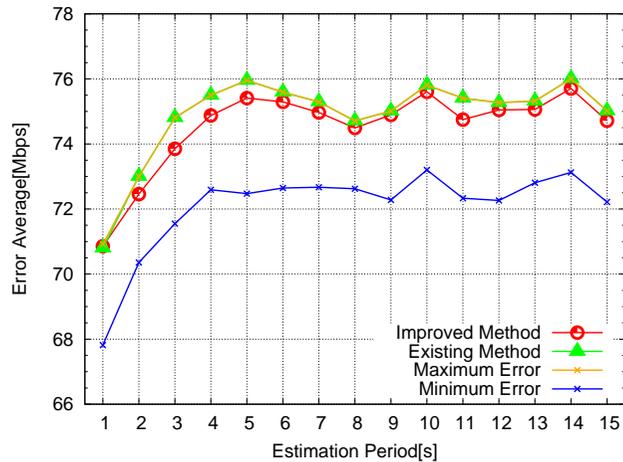


図6 変動重み指数移動平均法 (サンプル期間数 5 個)

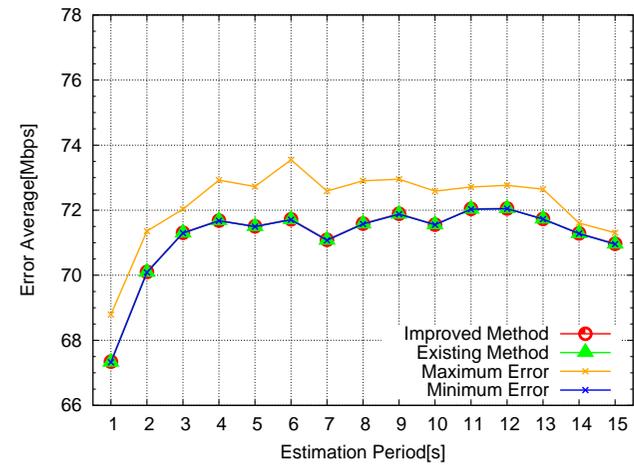


図7 変動重み指数移動平均法 (サンプル期間数 25 個)

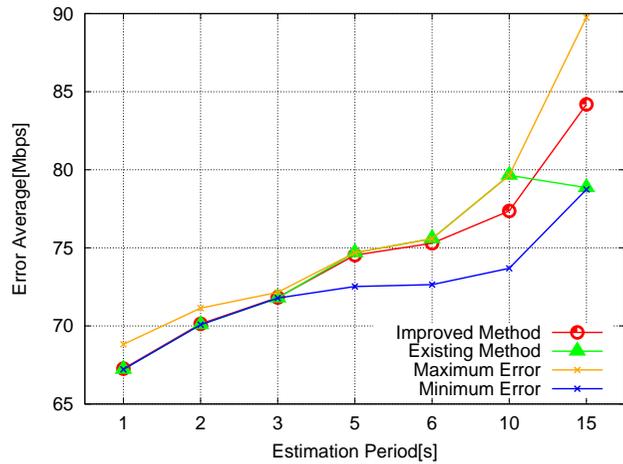


図 8 変動重み指数移動平均法 (分析期間長 30 秒)

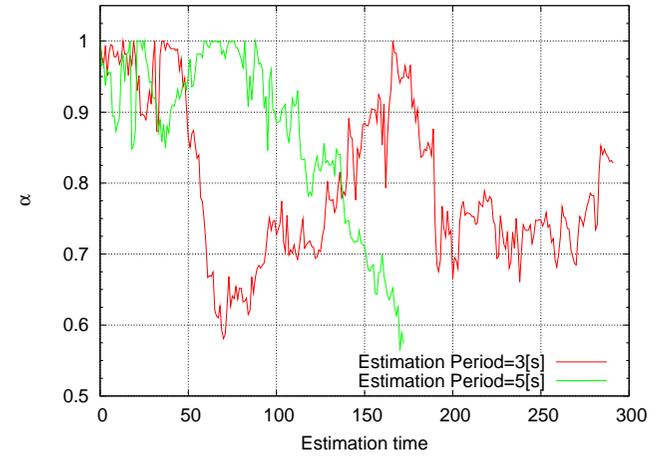


図 10 α 推移例 (サンプル期間数 5 個)

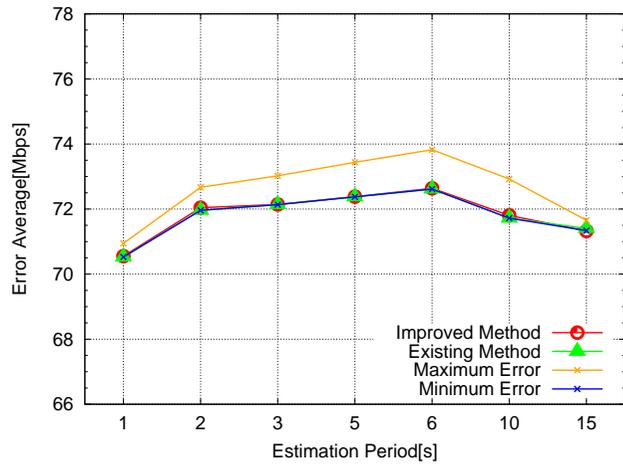


図 9 変動重み指数移動平均法 (分析期間長 5 分)

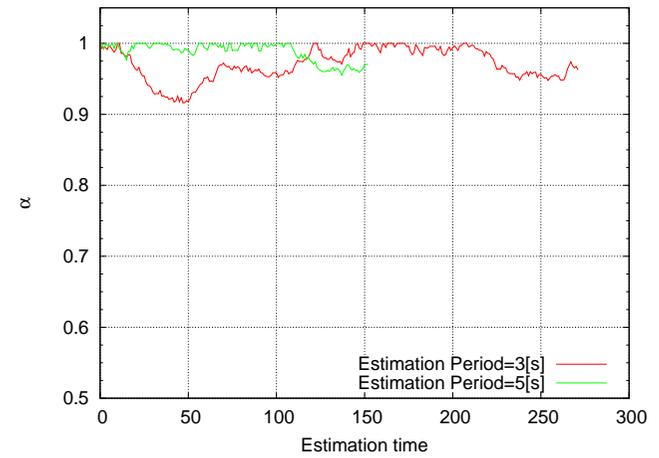


図 11 α 推移例 (サンプル期間数 25 個)

5. ま と め

本研究では極値理論を用いた既存のトラフィックピークレート推定法に対し、重み係数 α を逐次変更することのできる指数移動平均を適用することで推定誤差を低減する手法を提案した。実トラフィックデータを用いた評価の評価、提案手法を用いることによって、推定に利用できるデータ数が少ない場合には既存手法に比べ推定誤差を低減でき、推定に利用できるデータ数が多い場合には既存手法と同等の推定誤差を達成できることを示した。これにより、推定に利用できるデータ数の多寡によらず、トラフィックピークレートを適切に推定することが可能となった。

謝辞 本研究の一部は、新エネルギー・産業技術総合開発機構「グリーンネットワーク・システム技術研究開発プロジェクト(グリーン IT 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。

参 考 文 献

- 1) 金光 宏明, “ピークスルーブット推定に基づく細粒度転送速度制御によるスイッチの省電力化,” 九州工業大学修士論文, 2011 年.

付 録

トラフィックピークレートの推定値として、式 (13) で与えられる期待値を用いた場合の結果を示す。図 12 と図 13 は、分析期間長が 30 秒と 5 分の場合の結果を示したものである。これらの図より、中央値を用いた場合と同様に、推定に利用できるデータ数が少なくなると提案手法の方が誤差低減できるということが分かる。

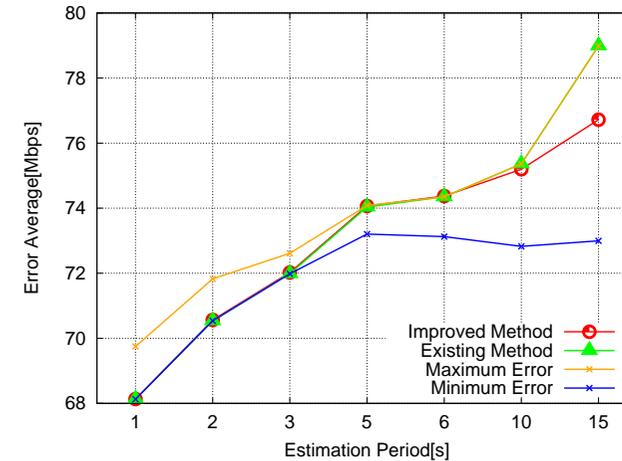


図 12 変動重み指数移動平均法 (分析期間長 30 秒)

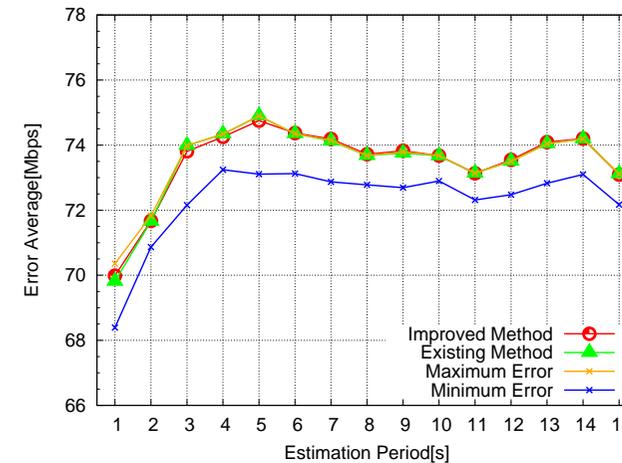


図 13 変動重み指数移動平均法 (サンプル期間数 5 個)