

Traffic Data Analysis Based on Extreme Value Theory and Its Applications to Predicting Unknown Serious Deterioration

Masato UCHIDA^{†a)}, Member

SUMMARY It is important to predict serious deterioration of telecommunication quality. This paper investigates predicting such serious events by analyzing only a “short” period (i.e., a “small” amount) of teletraffic data. To achieve this end, this paper presents a method for analyzing the tail distributions of teletraffic state variables, because tail distributions are suitable for representing serious events. This method is based on Extreme Value Theory (EVT), which provides a firm theoretical foundation for the analysis. To be more precise, in this paper, we use throughput data measured on an actual network during daily busy hours for 15 minutes, and use its first 10 seconds (*known* data) to analyze the tail distribution. Then, we evaluate how well the obtained tail distribution can predict the tail distribution of the remaining 890 seconds (*unknown* data). The results indicate that the obtained tail distribution based on EVT by analyzing the small amount of *known* data can predict the tail distribution of *unknown* data much better than methods based on empirical or log-normal distributions. Furthermore, we apply the obtained tail distribution to predict the peak throughput in *unknown* data. The results of this paper enable us to predict serious deterioration events with lower measurement cost.

key words: serious deterioration of the telecommunication quality, prediction, tail distribution, extreme value theory

1. Introduction

The tail distributions of teletraffic state variables, such as throughput, link-usage rate, packet loss rate, queue length, and delay time, are useful for predicting serious deterioration of telecommunication quality. The results of the predictions are also useful because they play a vital role in designing networks and controlling teletraffic effectively. For example, it is clear that the tail behavior (e.g., maximum (worst) value, high quantile value) of throughput and the link-usage rate affect the design of the bandwidth, and the tail behavior of the queue length affects the design of the buffer size. What is more, the volume of teletraffic can be controlled beforehand if we can predict serious deterioration of telecommunication quality. On the other hand, with the growth in Internet traffic and the growing variety of Internet applications, it is becoming more important to predict serious deterioration of telecommunication quality on an IP network, because IP networks that can satisfy the demand for guaranteed quality of service (QoS) will be increasingly required. It follows from what has been said that it is meaningful to analyze the tail distributions of teletraffic state variables. The purpose of this paper is to analyze them.

Up to now, a number of studies have investigated tele-

traffic behavior. For example, numerous papers on aggregated teletraffic have revealed that the burst for a short period and the correlation for a long period of aggregated teletraffic in various networks are remarkable [1], [2]. On the other hand, it is reported that this feature is suppressed with a high load [3]. However, these studies do not give much importance to analyzing the tail distribution of teletraffic data, which is important for evaluating telecommunication quality.

On the other hand, a great deal of effort has been made on tail distribution analysis. For example, in [4] and [5], it is shown that the tail distribution of teletraffic data can be approximated using a “mixture of normal distributions” and using a “Pareto distribution”, respectively. What seems to be lacking, however, is the evaluation of how well the tail distribution obtained by analyzing *known* (*observed, past*) teletraffic data can predict the tail distribution of *unknown* (*unobserved, future*) teletraffic data, where both *known* and *unknown* data are measured on the same network under similar conditions. This insufficiency is a real and substantial problem because if the tail distribution obtained by analyzing *known* data cannot provide an accurate prediction (approximation) for the tail distribution of *unknown* data, then the methods for designing networks and controlling teletraffic using the analysis results will become unreliable.

Against this background, this paper analyzes the tail distribution using *known* data, and focuses on whether the analysis result is appropriate for *unknown* data. This paper also focuses on using a “small” amount (i.e., a “short” period) of teletraffic data as *known* data. In other words, the purpose of this paper is to predict *unknown* serious events that are not included in the small amount of *known* data. Obviously, this is also meaningful from the viewpoint of measurement cost. As the first step in our work, this paper uses throughput data as an example.

To analyze the tail distribution, in this paper, we look at Extreme Value Theory (EVT), which provides a firm theoretical foundation for the analysis. To show that EVT works efficiently in the analysis, we first approximate the tail distribution of a small amount of *known* data based on EVT. Although EVT was simply applied to analyze the tail distribution (of *known* data) in [5], this paper applies EVT more rigorously. Then, we show that the approximated tail distribution of *known* data based on EVT can also approximate the tail distribution of *unknown* data. In addition, we also show that the approximated tail distributions of *known* data based on the empirical distribution or the log-normal dis-

Manuscript received March 31, 2004.

Manuscript revised June 25, 2004.

[†]The author is with NTT Service Integration Laboratories, NTT Corporation, Musashino-shi, 180-8585 Japan.

a) E-mail: uchida.masato@lab.ntt.co.jp

tribution, which is used as the model of the distribution of Web transfer throughput [6], cannot approximate the tail distribution of *unknown* data. Finally, we apply the analysis result based on EVT to estimate the peak value of throughput in *unknown* data.

This paper is organized as follows. In Sect. 2, we survey EVT, which is the theoretical background of this paper. In Sect. 3, we explain the method for analyzing tail distributions. In Sect. 4, we discuss the analysis in [5] and describe its relationship with this paper. In Sect. 5, we present the analysis results and their efficiency. Section 6 is the conclusion.

2. Extreme Value Theory

Consider a random variable X with distribution function F that is defined by

$$F(x) = \Pr\{X \leq x\}.$$

In EVT [7], [8], it is shown that for a large class of underlying distributions F , we can find a constant value ξ and a positive function $\beta(u)$ such that

$$F(x) \approx (1 - F(u))G_{\xi, \beta(u)}(x - u) + F(u) \tag{1}$$

for $x > u$ when $u \rightarrow x_F$, where $x_F \leq \infty$ is called the right endpoint of F and is defined by

$$x_F = \sup\{x \in \mathbb{R}; F_X(x) < 1\},$$

and $G_{\xi, \beta(u)}$ is called the Generalized Pareto Distribution (GPD), which is defined by

$$G_{\xi, \beta(u)}(y) = \begin{cases} 1 - (1 + \xi y/\beta(u))^{-1/\xi} & \xi \neq 0 \\ 1 - \exp(-y/\beta(u)) & \xi = 0 \end{cases}.$$

That is, for a large class of underlying distributions F , as the threshold u is progressively raised, the tail distribution of F converges to a GPD. In addition, this class contains almost all of the common continuous distributions: normal, log-normal, gamma, log-gamma, exponential, χ^2 , t , F , uniform, beta, Pareto, Cauchy, and a mixture of these. This flexibility is useful for analyzing tail distributions, because we can model a tail distribution using a GPD regardless of a underlying distribution F .

3. Analysis Method

Let $D = \{x_1, x_2, \dots, x_n\}$ be originally observed data governed by a distribution function F . Then, for the data D , let $\tilde{F}_D(x)$ be the empirical distribution function. $\tilde{F}_D(x)$ is defined as

$$\tilde{F}_D(x) = \frac{1}{|D|} \sum_{d \in D} I_{\{d|d \leq x, d \in D\}}(d),$$

where $|D|$ is the number of elements in D , and

$$I_A(a) = \begin{cases} 1 & \text{if } a \in A \\ 0 & \text{if } a \notin A \end{cases}$$

is called the indicator function for set A .

In this paper, we use $\tilde{F}(u)$ as the estimator of $F(u)$, and provide the estimators of ξ and $\beta(u)$ using the moment method [9]. The moment estimators of ξ and $\beta(u)$, which are denoted by $\hat{\xi}_D(u)$ and $\hat{\beta}_D(u)$, can be derived as

$$\hat{\xi}_D(u) = \frac{1}{2} \left(1 - \frac{\hat{e}_D(u)^2}{\hat{v}_D(u)} \right),$$

$$\hat{\beta}_D(u) = (1 - \hat{\xi}_D(u))\hat{e}_D(u),$$

where

$$\hat{e}_D(u) = \frac{\sum_{y \in \{d-u|d>u, d \in D\}} y}{|D| - N_D(u)},$$

$$\hat{v}_D(u) = \frac{\sum_{y \in \{d-u|d>u, d \in D\}} \{y - \hat{e}_D(u)\}^2}{|D| - N_D(u)},$$

and

$$N_D(x) = \sum_{d \in D} I_{\{d|d \leq x, d \in D\}}(d).$$

Now, we can provide the estimated tail distribution function as

$$\hat{F}_D(x, u) = (1 - \tilde{F}_D(u))G_{\hat{\xi}_D(u), \hat{\beta}_D(u)}(x - u) + \tilde{F}_D(u), \tag{2}$$

for $x > u$.

Finally, let us denote the arrangement of $D = \{x_1, x_2, \dots, x_n\}$ in ascending order of value by

$$x_{[1]} \leq x_{[2]} \leq \dots \leq x_{[n]}.$$

In this paper, we abbreviate $\hat{F}_D(x, x_{[k]})$ as $\hat{F}_{D,k}(x)$, $\hat{\xi}_D(x_{[k]})$ as $\hat{\xi}_{D,k}$, and $\hat{\beta}_D(x_{[k]})$ as $\hat{\beta}_{D,k}$, when $u = x_{[k]}$.

4. Related Work

Regular variation theory is known as a basis of EVT. In this theory, it is shown that

$$1 - F(x) = cx^{-\alpha}, \quad x \rightarrow \infty, \tag{3}$$

for $x > 1$, where c is a constant value.

In [5], the data of packet interarrival and call holding times are analyzed using the above result, and the accuracy of the analysis is discussed. However, as mentioned in Sect. 1, there is no discussion of the accuracy of the approximation for *unknown* data. In addition, Eq. (1) can include Eq. (3) as a special case because Eq. (3) can be derived by substituting $u = 1$, $\beta(u) = \xi$, $\frac{1}{\xi} = \alpha$, and $F(u) = 1 - c$ in Eq. (1). Therefore, it is sufficient to consider Eq. (1). Note that this substitution assumes that the value of u is small. However, in the following sections, it is shown that if u is small, the analysis results are not valid.

5. Analysis of Real Traffic Data

5.1 Traffic Data

In our analysis, we used one-way traffic traces provided by the Widely Integrated Distributed Environment

(WIDE) project – the largest Internet research community in Japan [10]. The traces were measured on one of the international lines for the WIDE project during daily busy hours (14:00–14:15), and were available from the MAWI (Measurement and Analysis on the WIDE Internet) traffic archive [11]. The measured line was a 100-Mbps Ethernet with an 18-Mbps CAR (Committed Access Rate). More detailed information about the traces is given in [10], [11].

In this section, we use the trace measured on July 7, 2003 (200307071400.dump.gz), and analyze the throughput data of each 10 ms calculated from the trace. We selected this sampling cycle (10 ms) so that we were able to see packet-level telecommunication quality, such as queuing delay. In determining the sampling cycle, we referred to the RED architecture [12], [13]. Further details on this point are given in Appendix A.

Now, we define some notation for this throughput data. First, let us denote the data as $D^{\text{all}} = \{x_1^{\text{all}}, x_2^{\text{all}}, \dots, x_{n^{\text{all}}}^{\text{all}}\}$, where x_t^{all} is the value of throughput at $10 \times t$ ms, and $n^{\text{all}} = 15 [\text{min}] \times 60 [\text{sec}] \times 1000 [\text{msec}]/10 = 90000$. Secondly, let us denote the first 1000 elements in D^{all} as $D = \{x_1, x_2, \dots, x_n\}$. That is, $n = 1000$ and $x_t = x_t^{\text{all}}$ ($t = 1, \dots, n$). Finally, let us denote the remaining 89000 elements in D^{all} as $\bar{D} = \{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_{\bar{n}}\}$. That is, $\bar{n} = 89000$ and $\bar{x}_t = x_{1000+t}^{\text{all}}$ ($t = 1, \dots, \bar{n}$).

In the following sections, we first analyze D based on EVT, and show that the analysis result (estimated tail distribution of D) can approximate the tail distribution of D efficiently. Then, we show that the estimated tail distribution of D can also approximate the tail distribution of \bar{D} efficiently. Here, we call D *known* data, and \bar{D} *unknown* data.

Some figures are listed for reference. The time series of D^{all} is shown in Fig. 1. The time series of the first 1000 elements of D^{all} (i.e., *known* data D) is shown in Fig. 2. The histograms of D and \bar{D} are shown in Fig. 3. The statistical information about D and \bar{D} is shown in Table 1.

We can notice the following information from these figures and table. As shown in Table 1, the skewness of D and \bar{D} take positive values. This indicates that the distributions of D and \bar{D} are positively skewed. In addition, as shown in Figs. 1 and 2, it is entirely fair to say that the teletraffic retains its stationarity within this short time interval. In this paper, we analyze only data with enough stationarity as 200307071400.dump.gz. The histogram of D is rougher than that of \bar{D} as shown in Fig. 3. The range of *known* data D is narrower than that of *unknown* data \bar{D} (compare the values of minimum and maximum of D and \bar{D}) as shown in Table 1. Thus, these characteristics indicate that it may be difficult to estimate \bar{D} from D in a simple manner. However, in this paper, we show that the tail distribution of \bar{D} can be approximated even when such rough data D is used for the analysis if we use EVT.

5.2 Analysis of Known Data

In this section, we analyze the *known* data D and look at how well the analysis results approximate the tail distribution of

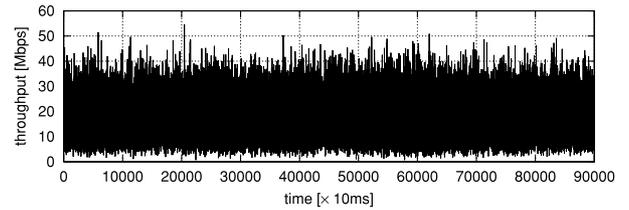


Fig. 1 Time series of throughput.

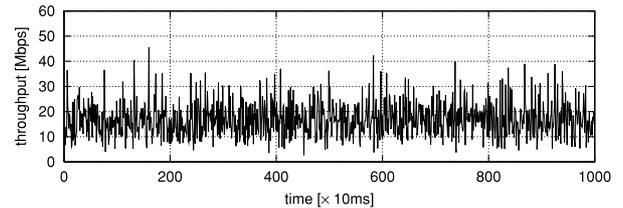


Fig. 2 Stretched image of Fig. 1.

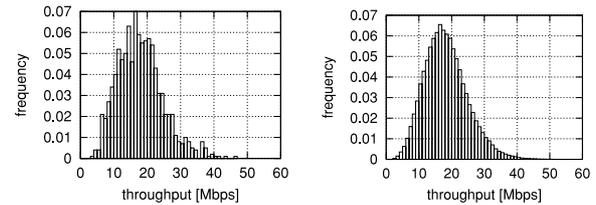


Fig. 3 Histograms of D and \bar{D} (Left: D , Right: \bar{D}).

Table 1 Statistics of D and \bar{D} .

	num	min	max	ave	var	skew
D	1000	2.69	45.52	16.94	45.80	0.67
\bar{D}	89000	0.66	54.66	17.26	42.58	0.57

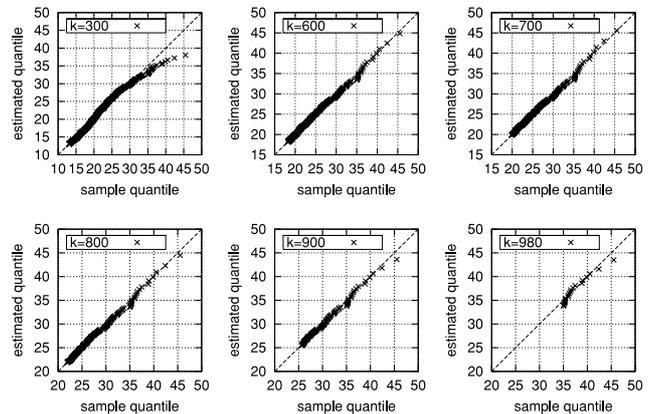


Fig. 4 QQ-plots described by Eq. (4) for $k = 300, 600, 700, 800, 900,$ and 980 .

D . To start, we look at the analysis result using EVT. The sequence of QQ-plots (see Appendix B) is shown in Fig. 4,

$$\left(d, \hat{F}_{D,k}^{-1} \left(\frac{1}{|D|+1} N_D(d) \right) \right), \quad \text{for } d \in \{d \mid d > x_{[k]}, d \in D\}, \quad (4)$$

where $k = 300, 600, 700, 800, 900,$ and 980 . As shown in the sequence of figures, (i) when $k = 300$ the plot clearly deviates from a straight line with gradient of 1 and y-intercept of 0, and (ii) when $k = 600, 700, 800, 900,$ and 980 , each plot approaches a straight line with gradient of 1 and y-intercept of 0. Although the plot for $k = 980$ seems to follow this line less closely than the plots for $k = 600, 700, 800,$ and 900 , the point to observe here is that all plots for $k = 600, 700, 800, 900,$ and 980 are much closer to a straight line with gradient of 1 and y-intercept of 0 than the plot for $k = 300$. This indicates that the tail distribution of *known* data D can be approximated by

$$\hat{F}_{D,k}(x) = (1 - \tilde{F}_D(x_{[k]}))G_{\hat{\xi}_{D,k}, \hat{\beta}_{D,k}}(x - x_{[k]}) + \tilde{F}_D(x_{[k]}),$$

for $x > x_{[k]}$, where $k = 600, 700, 800, 900,$ and 980 .

Considering that the log-normal distribution is used as the model of the distribution of Web transfer throughputs [6], we then look at the QQ-plot using the log-normal distribution. The QQ-plot

$$\left(d, L_{\hat{\mu}_D, \hat{\sigma}_D^2}^{-1} \left(\frac{1}{|D|+1} N_D(d) \right) \right),$$

for $d \in \{d|d > x_{[k]}, d \in D\}$, (5)

is shown in Fig. 5, where L_{μ, σ^2} is the log-normal distribution function with mean μ and variance σ^2 , and $\hat{\mu}_D$ and $\hat{\sigma}_D^2$ are the maximum likelihood estimators of μ and σ^2 for *known* data D . The tail part of the plot in Fig. 5 clearly deviates from a straight line with gradient of 1 and y-intercept of 0, compared with Fig. 4 when $k \geq 600$.

As a result, we can conclude that the tail distribution of *known* data D can be approximated by the GPD better than the log-normal distribution.

5.3 Analysis of Unknown Data

We are now in a position to say that $\hat{F}_{D,k}$ can approximate the tail distribution of *known* data D . However, this does not necessarily mean that $\hat{F}_{D,k}$ can also approximate the tail distribution of F because the *known* data D is only a small sample of underlying distribution F . This leads us to the question of whether $\hat{F}_{D,k}$ can approximate the tail distribution of *unknown* data \bar{D} . This is a real and substantial problem, as mentioned in Sect. 1. Therefore, in this section, we show that $\hat{F}_{D,k}$ can also approximate the tail distribution of

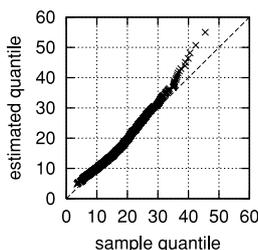


Fig. 5 QQ-plots described by Eq. (5).

unknown data \bar{D} much better than both \tilde{F}_D and $L_{\hat{\mu}_D, \hat{\sigma}_D^2}$. Note that, \bar{D} has a large number of elements, so it is close to the underlying distribution F .

To start, we look at how well $\hat{F}_{D,k}$ approximates the tail distribution of *unknown* data \bar{D} . QQ-plots

$$\left(\bar{d}, \hat{F}_{D,k}^{-1} \left(\frac{1}{|\bar{D}|+1} N_{\bar{D}}(\bar{d}) \right) \right),$$

for $\bar{d} \in \{\bar{d}|\bar{d} > x_{[k]}, \bar{d} \in \bar{D}\}$, (6)

are shown in Fig. 6, where $k = 300, 600, 700, 800, 900,$ and 980 . As shown in the sequence of figures, each QQ-plot for $k = 300, 980$ deviates from a straight line with gradient of 1 and y-intercept of 0. Compared with these, each QQ-plot for $k = 600, 700, 800,$ and 900 is closer to a straight line with gradient of 1 and y-intercept of 0. This means that $\hat{F}_{D,k}$ can approximate the tail distribution of *unknown* data \bar{D} when $k = 600, 700, 800,$ and 900 . These characteristics arise from the following two conflicting findings.

- When the value of $x_{[k]}$ (i.e., the value of k) is small ($k = 300$ in this case), $\hat{F}_{D,k}$ cannot approximate the tail distribution of both *known* data D and *unknown* data \bar{D} because the condition $u \rightarrow x_F$ in Eq. (1) is not satisfied.
- When the value of $x_{[k]}$ (i.e., the value of k) is large ($k = 980$ in this case), $\hat{F}_{D,k}$ cannot approximate the tail distribution of *unknown* data \bar{D} , because $|D| - N_D(x_{[k]})$, which is the number of data used for the analysis, is small so $\hat{F}_{D,k}$ becomes an excessively biased approximation of the tail distribution of D and thus does not have enough flexibility, though the condition $u \rightarrow x_F$ in Eq. (1) is satisfied. In other words, the generality (flexibility) of the $\hat{F}_{D,k}$ to approximate *unknown* data is lost by “overfitting” the small number of analyzing data when k takes a large value, though it is easy to fit the data because the amount of data is small.

These two points of view mean that the values of both $x_{[k]}$ and $|D| - N_D(x_{[k]})$ have to be made large to approximate well the tail distributions of *known* data D and *unknown* data \bar{D} . As a result, we can use

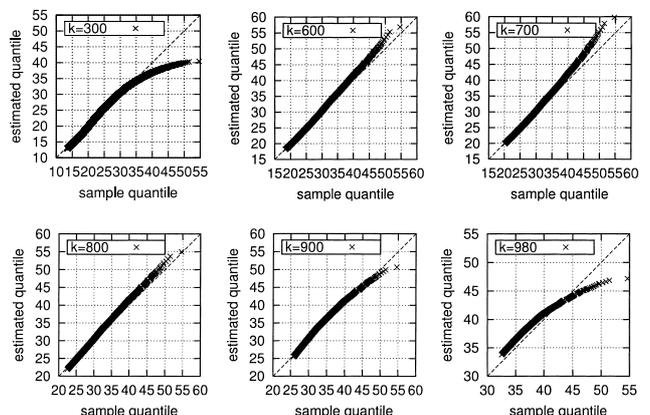


Fig. 6 QQ-plots described by Eq. (6) for $k = 300, 600, 700, 800, 900,$ and 980 .

$$\begin{aligned} \hat{F}_{D,k}(x) &= (1 - \tilde{F}_D(x_{[k]}))G_{\hat{\xi}_{D,k}, \hat{\beta}_{D,k}}(x - x_{[k]}) + \tilde{F}_D(x_{[k]}) \end{aligned}$$

as the estimated tail distribution of *known* data D and *unknown* data \bar{D} for $x > x_{[k]}$, where $k = 600, 700, 800$, and 900 .

So far, we have seen that the tail distribution can be approximated by $\hat{F}_{D,k}$ if k is appropriately selected. Now, we look at how well \tilde{F}_D and $L_{\hat{\mu}_D, \hat{\sigma}_D^2}$ approximate the tail distribution of *unknown* data \bar{D} .

The QQ-plot

$$\left(\bar{d}, \tilde{F}_D^{-1}\left(\frac{1}{|\bar{D}|+1}N_{\bar{D}}(\bar{d})\right) \right), \quad \text{for } \bar{d} \in \{\bar{d} | \bar{d} > x_{[k]}, \bar{d} \in \bar{D}\}, \quad (7)$$

is shown in Fig. 7 (a) and the QQ-plot

$$\left(\bar{d}, L_{\hat{\mu}_D, \hat{\sigma}_D^2}^{-1}\left(\frac{1}{|\bar{D}|+1}N_{\bar{D}}(\bar{d})\right) \right), \quad \text{for } \bar{d} \in \{\bar{d} | \bar{d} > x_{[k]}, \bar{d} \in \bar{D}\} \quad (8)$$

is shown in Fig. 7 (b). As shown in these figures, both QQ-plots clearly deviate from straight lines with gradient of 1 and y-intercept of 0 when the value of \bar{d} is large, as compared with the QQ-plots in Fig. 6 when $k = 600, 700, 800$, and 900 . This means that \tilde{F}_D and $L_{\hat{\mu}_D, \hat{\sigma}_D^2}$ cannot approximate the tail distribution of \bar{D} . More specifically, for Fig. 7 (a), this arises from the *known* data D hardly including outlier elements because the number of elements in D is small. Actually, as shown in Table 1, the maximum value and average value of *known* data D are smaller than those of *unknown* data \bar{D} . The important point to note is that $\hat{F}_{D,k}$ can approximate the tail distribution of \bar{D} even though D hardly includes outlier elements. In addition, as shown in Fig. 7 (b), $L_{\hat{\mu}_D, \hat{\sigma}_D^2}^{-1}(N_{\bar{D}}(\bar{d})/(|\bar{D}|+1))$ becomes much larger than \bar{d} when \bar{d} becomes large. This indicates that the characteristics of the tail part of the log-normal distribution cause an underestimation of the approximation of the tail distribution of \bar{D} .

In the next section, we look at the effectiveness of $\hat{F}_{D,k}$ by estimating the peak throughput applying the analysis results.

5.4 Estimation of Peak Throughput

In this section, we estimate the *known* peak throughput (i.e.,

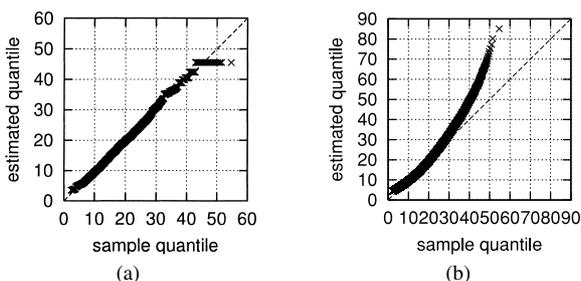


Fig. 7 QQ-plots described by Eq.(7) and Eq.(8) ((a): Eq.(7), (b): Eq.(8)).

the maximum value of elements in D) and the *unknown* peak throughput (i.e., the maximum value of elements in \bar{D}) using $\hat{F}_{D,k}$, and discuss the accuracy of the estimation. The definitions of the *known* peak throughput x_{\max} and the *unknown* peak throughput \bar{x}_{\max} are presented below.

$$\begin{aligned} x_{\max} &= \max_{x \in D} x \\ \bar{x}_{\max} &= \max_{\bar{x} \in \bar{D}} \bar{x} \end{aligned}$$

To estimate *known* and *unknown* peak throughputs, we use the quantile function. That is, for a certain distribution H , we use the $\frac{|D|}{|D|+1}$ -th quantile of H ,

$$H^{-1}\left(\frac{|D|}{|D|+1}\right),$$

as the estimated *known* peak throughput of H , and we use the $\frac{|\bar{D}|}{|\bar{D}|+1}$ -th quantile of H ,

$$H^{-1}\left(\frac{|\bar{D}|}{|\bar{D}|+1}\right),$$

as the estimated *unknown* peak throughput of H . Note the following relationships:

$$\begin{aligned} x_{\max} &= \tilde{F}_D^{-1}\left(\frac{|D|}{|D|+1}\right) = \tilde{F}_D^{-1}\left(\frac{|\bar{D}|}{|\bar{D}|+1}\right), \\ \bar{x}_{\max} &= \tilde{F}_{\bar{D}}^{-1}\left(\frac{|\bar{D}|}{|\bar{D}|+1}\right). \end{aligned}$$

In addition, we denote the estimated *known* peak throughput of $L_{\hat{\mu}_D, \hat{\sigma}_D^2}$ as l_{\max} and the estimated *unknown* peak throughput of $L_{\hat{\mu}_{\bar{D}}, \hat{\sigma}_{\bar{D}}^2}$ as \bar{l}_{\max} , which are defined by

$$l_{\max} = L_{\hat{\mu}_D, \hat{\sigma}_D^2}^{-1}\left(\frac{|D|}{|D|+1}\right)$$

and

$$\bar{l}_{\max} = L_{\hat{\mu}_{\bar{D}}, \hat{\sigma}_{\bar{D}}^2}^{-1}\left(\frac{|\bar{D}|}{|\bar{D}|+1}\right).$$

The relationship between k and $\hat{F}_{D,k}^{-1}\left(\frac{|D|}{|D|+1}\right)$ is shown in Fig. 8, where $x_{\max} = 45.52$ (Mbps) and $l_{\max} = 54.98$ (Mbps). $\hat{F}_{D,k}^{-1}\left(\frac{|D|}{|D|+1}\right)$ approximates the *known* peak throughput x_{\max} better than l_{\max} when k is large (around $k \geq 600$).

The relationship between k and $\hat{F}_{\bar{D},k}^{-1}\left(\frac{|\bar{D}|}{|\bar{D}|+1}\right)$ is shown in Fig. 9, where $x_{\max} = 45.52$ (Mbps), $\bar{x}_{\max} = 54.66$ (Mbps),

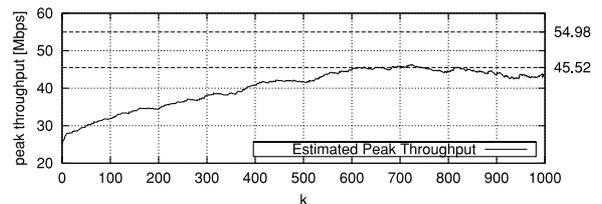


Fig. 8 Estimated *known* peak throughput (relationship between k and $\hat{F}_{D,k}^{-1}\left(\frac{|D|}{|D|+1}\right)$), where $x_{\max} = 45.52$ and $l_{\max} = 54.98$.

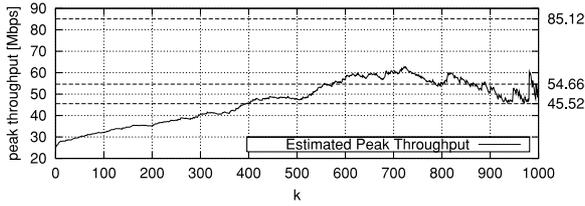


Fig. 9 Estimated unknown peak throughput (relationship between k and $\hat{F}_{D,k}^{-1}\left(\frac{|D|}{|D|+1}\right)$, where $x_{\max} = 45.52$, $\bar{x}_{\max} = 54.66$, and $\bar{l}_{\max} = 85.12$).

and $\bar{l}_{\max} = 85.12$ (Mbps). $\hat{F}_{D,k}^{-1}\left(\frac{|D|}{|D|+1}\right)$ approximates the *unknown* peak throughput \bar{x}_{\max} better than l_{\max} and x_{\max} when k is a medium value (around $600 \leq k \leq 900$).

It should be concluded from what was said above that the estimated *known* and *unknown* peak throughputs based on $\hat{F}_{D,k}$ work more efficiently than \tilde{F}_D and $L_{\hat{\mu}_D, \hat{\sigma}_D^2}$ when k is a medium value. These results agree with the discussion in Sects. 5.2 and 5.3.

5.5 Discussion

5.5.1 Discussion of the Value of k

We decided the appropriate range of k (around $600 \leq k \leq 900$) in the previous sections, but there is room for further investigation because the range was decided using both *known* and *unknown* data. Note that *unknown* data cannot be used for the analysis in practical situations, so we must decide the value of k using only *known* data. To do this, we first recall the results in previous sections.

- The tail distribution of *known* data can be approximated when k is a large value (around $k \geq 600$).
- The tail distribution of *unknown* data can be approximated when k is a large value, except when it is too large (around $600 \leq k \leq 900$).

The above indicates that it is necessary to increase the value of k (i.e., $N_D(x_{[k]})$) so that the value of $|D| - N_D(x_{[k]})$ should not become small too much. Therefore, it seems reasonable to use the smallest value of k that is appropriate for approximating the tail distribution of *known* data (about $k = 600$ in this case). Note that, as shown above, this value of k is also appropriate for the tail distribution of *unknown* data. In this case, $x_{[600]}$ is 18.16 (Mbps), which is almost the same as the average value of D and \bar{D} .

5.5.2 Discussion of the Value of n

In the previous sections, we have seen that the tail distribution of *unknown* data can be well estimated from a small amount of *known* data by using the proposed method. This means that the measurement of teletraffic over a short period (e.g., the first 10 seconds) provides enough data for estimation of the near-future teletraffic (e.g., the following 890 seconds). However, the 10 seconds of teletraffic data might be excessive for estimation of the teletraffic data over the

following 890 seconds. So, in this section, we use various values smaller than 1000 (equivalent to 10 seconds) for the number of *known* data n , and observe how well the tail distribution of the following 89000 *unknown* data (equivalent to 890 seconds) is estimated from the smaller amounts of *known* data. The efficiency of the analysis results is quantitatively evaluated through the estimation of peak throughput. If we can reduce the value of n without decreasing the efficiency of the analyzed result, it is clearly meaningful from the practical point of view, because it means that we can reduce the cost of measuring the data for estimation.

Considering that the estimated *unknown* peak throughput using the proposed method fluctuates with the value of k (as seen in Fig. 9), we use the representative value defined by

$$r_{\max,n} = \frac{\sum_{k=[0.6n]}^{[0.7n]} \hat{F}_{D_n,k}^{-1}\left(\frac{|\bar{D}_n|}{|\bar{D}_n|+1}\right)}{[0.7n] - [0.6n] + 1}, \tag{9}$$

where D_n is the set of the first n elements of D^{all} , \bar{D}_n is the set of the following 89000 elements of D^{all} , and $[\alpha]$ is the closest integer that is equal to or smaller than α . Note that we decided the range of k for the summation in Eq. (9) based on the discussion in the previous sections. That is, as shown in the previous sections, $\hat{F}_{D_n,k}^{-1}\left(\frac{|\bar{D}_n|}{|\bar{D}_n|+1}\right)$ provides a good approximation of the *unknown* peak throughput when $600 \leq k \leq 900$ (see Sect. 5.4), and the value of k should be about 600 (see Sect. 5.5.1) when $n = 1000$. This indicates that the value of k should be equal to or slightly larger than 600. So, in this section, we set $[0.6n] \leq k \leq [0.7n]$ (i.e., $600 \leq k \leq 700$ if $n = 1000$). Although this range might be irrelevant for values of n other than 1000, we can obtain a rough representative value by using this range.

Furthermore, let us define the estimated *unknown* peak throughputs using an empirical distribution ($x_{\max,n}$) and log-normal distribution ($\bar{l}_{\max,n}$) as

$$x_{\max,n} = \tilde{F}_{D_n}^{-1}\left(\frac{|\bar{D}_n|}{|\bar{D}_n|+1}\right),$$

$$\bar{l}_{\max,n} = L_{\hat{\mu}_{D_n}, \hat{\sigma}_{D_n}^2}^{-1}\left(\frac{|\bar{D}_n|}{|\bar{D}_n|+1}\right).$$

Figure 10 shows the error in estimation for each method. The estimation errors of the proposed method (GPD), empirical distribution, and log-normal distribution are defined as $|r_{\max,n} - \bar{x}_{\max,n}|$, $|x_{\max,n} - \bar{x}_{\max,n}|$, and $|\bar{l}_{\max,n} - \bar{x}_{\max,n}|$, respectively, where $10 \leq n \leq 1000$, $|\alpha|$ is the absolute value of α , and $\bar{x}_{\max,n}$ is defined by

$$\bar{x}_{\max,n} = \tilde{F}_{\bar{D}_n}^{-1}\left(\frac{|\bar{D}_n|}{|\bar{D}_n|+1}\right).$$

This figure shows that the proposed method provides a much better approximation of the *unknown* peak throughput than methods using the empirical or log-normal distribution, even when $500 \leq n \leq 1000$. However, we can also see that

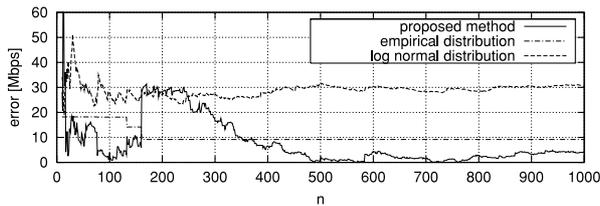


Fig. 10 Relationship between the estimation error and the number of data used in the analysis.

$|r_{\max,n} - \bar{x}_{\max,n}|$ fluctuates strongly when $10 \leq n < 500$ (especially when $10 \leq n < 200$). This is because the estimation result is strongly affected by the fluctuation of data when the number of data is small, so that the estimation result is unstable and unreliable. The above discussion means that we can safely reduce the number of data used in estimating the *unknown* peak throughput (i.e., the value of n) to at least 500 in this example.

6. Conclusion

In this paper, we analyzed the tail distribution of throughput based on EVT (GPD) using a “small” amount of *known* data, and showed that the obtained tail distribution approximates the tail distribution of *unknown* data better than that based on the empirical and log-normal distributions. In addition, we presented an application of the obtained tail distribution to estimate the peak throughput. The results of this paper enable us to

- predict how the telecommunication quality will be worsened on the network (e.g., the deterioration frequency (e.g., high quantile value) and the deterioration degree (e.g., maximum (worst) value) of the telecommunication quality), which can be useful information for operating the network, by analyzing only a small amount of teletraffic data.
- reduce the cost of measurements and the amount of storage space for the teletraffic data to predict serious deterioration of telecommunication quality.

Although the analysis presented in this paper was limited to an example of actual teletraffic data in [11], the results were almost the same even when other data were used. Another example of the analysis using data in [11] is presented in the Appendix C. Furthermore, the analysis result using data which is measured under other network conditions [14] is presented in the Appendix D.

It is important to note that the scope of the EVT-based teletraffic analysis method is not limited to throughput data. This is because the tail distributions of teletraffic state variables, which express rare but important and serious events, are useful in various situations for teletraffic engineering. Therefore, the EVT-based analysis method should be a powerful tool for analyzing various teletraffic data.

Some open problems remain:

- simple methods of deciding the appropriate k .

- simple methods of deciding the the minimum value of n that is sufficient to estimate the tail distribution of *unknown* data efficiently.
- the analysis of teletraffic data concerning teletraffic states other than throughput.
- the analysis of nonstational teletraffic data.

As a final point, a closer look at the fourth item above is worthwhile. In this paper, we have limited the discussion to the analysis of stationary data, and shown that the proposed method works efficiently with such data. However, the proposed method is not directly applicable to the the analysis of nonstationary data, because if the data is nonstationary, *known* and *unknown* data would have different statistical properties, and therefore the estimation results from *known* data would become unreliable for *unknown* data. To overcome this problem, we have to find a way of dealing with stationary period only. In [15], a method for the segmentation of nonstationary data into multiple sets of stationary data has been described. Such an approach will allow us to apply the proposed method to the individual segment of stationary data. That is, such techniques in combination with the proposed method will give us the ability to analyze nonstationary data.

References

- [1] W.E. Leland, M.S. Taqqu, W. Willinger, and D.V. Wilson, “On the self-similar nature of Ethernet traffic,” *IEEE/ACM Trans. Netw.*, vol.2, no.1, pp.1–15, 1994.
- [2] K. Park and W. Willinger, *Self-similar network traffic and performance evaluation*, Wiley Interscience, 2000.
- [3] J. Cao, W.S. Cleveland, D. Lin, and D.X. Sun, “Internet traffic tends toward Poisson and independent as the load increases,” in *Nonlinear Estimation and Classification*, ed. C. Holmes, D. Denison, M. Hansen, B. Yu, and B. Mallick, Springer, New York, 2002.
- [4] V. Bolotin, J. Coombs-Reyes, D. Heyman, Y. Levy, and D. Liu, “IP traffic characterization for planning and control,” *ITC 16*, pp.425–436, 1999.
- [5] S.I. Resnick, “Heavy tail modeling and teletraffic data,” *Annals of Statistics*, vol.25, no.5, pp.1805–1869, 1997.
- [6] H. Balakrishnan, S. Seshan, M. Stemm, and R.H. Katz, “Analyzing stability in wide-area network performance,” *ACM SIGMETRICS Conference on Measurement and Modeling of Computer Systems*, pp.2–12, June 1997.
- [7] R.-D. Reiss and M. Thomas, *Statistical Analysis of Extreme Values: with Applications to Insurance, Finance, Hydrology and Other Fields*, 2nd ed., Birkhäuser, 2001.
- [8] P. Embrechts, C. Klüppelberg, and T. Mikosch, *Modeling Extremal Events for Insurance and Finance*, Springer, 1997.
- [9] J.R.M. Hosking and J.R. Wallis, “Parameter and quantile estimation for the generalized Pareto distribution,” *Technometrics*, vol.29, pp.339–349, 1987.
- [10] Widely Integrated Distributed Environment Project, <http://www.wide.ad.jp>
- [11] Measurement and Analysis on the WIDE Internet, <http://www.wide.ad.jp/wg/mawi>
- [12] S. Floyd and V. Jacobson, “Random early detection gateways for congestion avoidance,” *IEEE/ACM Trans. Netw.*, vol.1, no.4, pp.397–413, Aug. 1993.
- [13] Cisco Systems, “Weighted random early detection on the Cisco 12000 series router,” <http://cco.cisco.com>
- [14] R. Kawahara, K. Ishibashi, T. Hirano, H. Saito, H. Ohara, D. Satoh,

S. Asano, and J. Matsukata, "Traffic measurement and analysis in an ATM-based Internet backbone," *Comput. Commun.*, vol.24, no.15-16, pp.1508-1524, Oct. 2001.

[15] K. Fukuda, H.E. Stanley, and L.A.N. Amaral, "Heuristic segmentation of a nonstationary time series," *Phys. Rev. E*, vol.69, no.2, pp.021108, 2004.

Appendix A: Sampling Cycle

In this paper, we referred to RED (Random Early Detection) [12] in determining the sampling cycle. RED is commonly implemented in routers as a tool for packet-level congestion control, and therefore it is valid to refer to the RED architecture in determining a sampling cycle that will reveal packet-level aspects of telecommunication quality such as queuing delay.

In RED, a small proportion of packets are discarded once the queue in the router has started to fill. More precisely, when a packet arrives, the following events occur.

- The average queue size is calculated.
- If the average queue size is less than the minimum threshold, the arriving packet is queued.
- If the average queue size is between the minimum threshold and the maximum threshold, packets are dropped according to the packet-drop probability.
- If the average queue size is greater than the maximum threshold, the packet is dropped.

The point is that RED discards arriving packets with a certain probability when the average queue size exceeds the minimum threshold. In [13], it is recommended that the minimum threshold should be $0.03B$, where B is the link bandwidth expressed as the amount of data per second. This means that $0.03B$ is equivalent to the traffic volume which can be offered to the link over 30 ms (0.03 seconds). Therefore, observing the teletraffic behavior with the sampling cycle of 10 ms (same order as 30 ms) is meaningful from the viewpoint of packet-level telecommunication quality. Here, it is also important that we consider the volume of offered load in determining the sampling cycle. For example, when the offered load is very high (close to the link bandwidth) or very low (close to 0 Mbps), the sampling cycle has to be made narrower or wider, respectively. However, Tables 1, A·1 and A·2 show that the volume of offered load of the teletraffic data analyzed in this paper is neither particularly high nor low. Therefore, we consider that the 10 ms sampling cycle is valid as long as the offered load is not at either extreme.

Appendix B: QQ-Plot

Let X be a random variable with distribution function H . Then, let us define its quantile function H^{-1} as

$$H^{-1}(p) = \inf\{x : H(x) \geq p\}, \quad 0 < p < 1,$$

where $H^{-1}(p)$ is called the p -th quantile. The definition of $H^{-1}(p)$ means that the value of the variable X becomes

larger than $H^{-1}(p)$ with probability $1 - p$. When p is a large value, $H^{-1}(p)$ is called a high quantile value.

Using sample data $D' = \{x'_1, x'_2, \dots, x'_n\}$ which are governed by a certain distribution, the QQ-plot (quantile-quantile plot) [7] is defined by

$$\left(d', H^{-1}\left(\frac{1}{|D'|+1}N_{D'}(d')\right) \right), \quad \text{for } d' \in D'.$$

If the sample data D' come from the distribution H , the plot will be close to a straight line with gradient of 1 and y-intercept of 0. If the deviation from the line is too strong we can conclude that the sample comes from a different distribution. Therefore, we can visually discriminate different distribution functions using a QQ-plot. It is known that QQ-plots enable a strong discrimination ability, especially in the range of the tail part of the distribution.

Appendix C: Traffic Data on a Different Day

In this section, we use the trace measured on July 9, 2003 (200307091400.dump.gz), and analyze the throughput data of each 10 ms calculated from the trace. The analysis result of this data is shown in some figures and a table. Table A·1 corresponds to Table 1, and Figs. A·1, A·2, ..., A·10 correspond to Figs. 1, 2, ..., 10, respectively.

As shown in these figures and table, the analysis result of 200307091400.dump.gz is almost the same as that of 200307071400.dump.gz, so the argument given in the previous sections is mostly valid here, too. However, there are some points to note.

The fluctuation in 200307091400.dump.gz is weaker than that in 200307071400.dump.gz as shown in Tables 1 and A·1. This is because the throughput range

Table A·1 Statistics of D and \bar{D} .

	num	min	max	ave	var	skew
D	1000	3.16	38.98	17.92	26.62	0.47
\bar{D}	89000	1.05	49.68	17.90	30.51	0.46

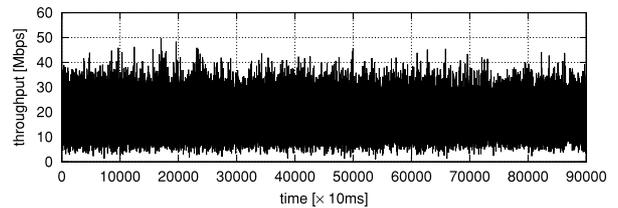


Fig. A·1 Time series of throughput.

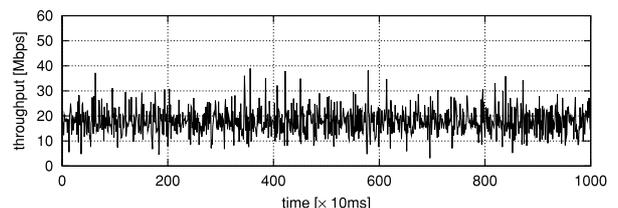


Fig. A·2 Stretched image of Fig. A·1.

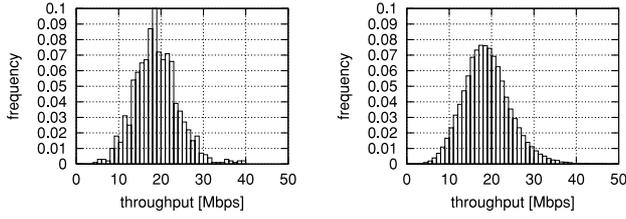


Fig. A·3 Histograms of D and \tilde{D} (Left: D , Right: \tilde{D}).

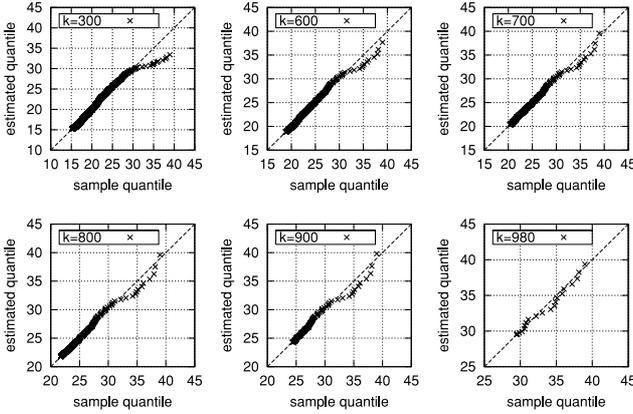


Fig. A·4 QQ-plots described by Eq. (4) for $k = 300, 600, 700, 800, 900,$ and 980 .

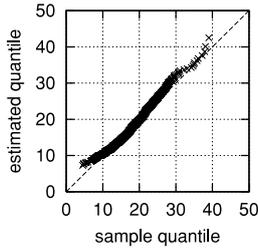


Fig. A·5 QQ-plots described by Eq. (5).

in 200307091400.dump.gz is narrower than that in 200307071400.dump.gz (see maximum and minimum values), and the variance of the throughput in 200307091400.dump.gz is smaller than that in 200307071400.dump.gz. This means that we analyzed two statistically different sets of throughput data.

For $k = 600, 700, 800, 900,$ and 980 , all these QQ-plots are close to a straight line with gradient of 1 and y-intercept of 0 as shown in Figs. A·4 and A·5. This indicates that both $\hat{F}_{D,k}$ and $L_{\hat{\mu}_D, \hat{\sigma}_D^2}$ can well approximate the tail distribution of *known* data in 200307091400.dump.gz. This is the most significant different point between the analysis results of 200307071400.dump.gz and 200307091400.dump.gz. However, this difference is not of major importance to our discussion in this paper because, as shown in Figs. A·6 and A·7(b), $\hat{F}_{D,k}$ approximates the tail distribution of *unknown* data in 200307091400.dump.gz much better than $L_{\hat{\mu}_D, \hat{\sigma}_D^2}$ when $k = 600, 700, 800,$ and 900 . This means that $L_{\hat{\mu}_D, \hat{\sigma}_D^2}$ works well only for *known* data in

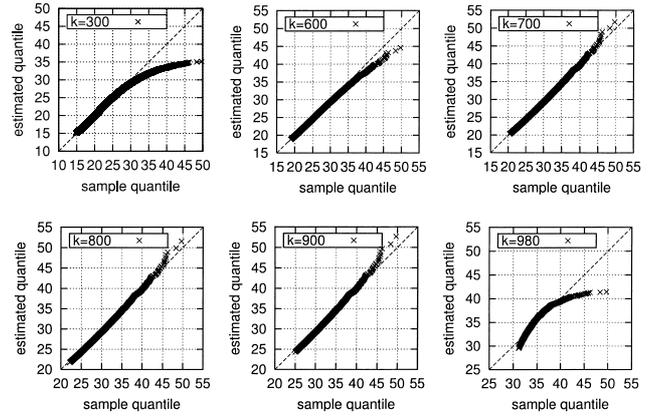


Fig. A·6 QQ-plots described by Eq. (6) for $k = 300, 600, 700, 800, 900,$ and 980 .

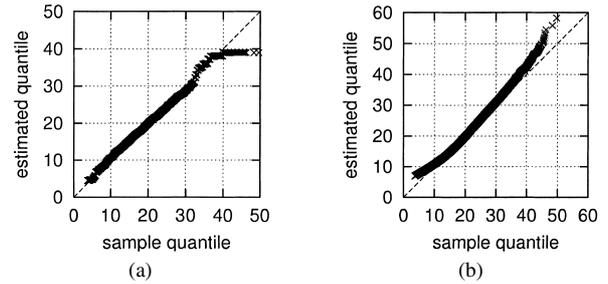


Fig. A·7 QQ-plots described by Eq. (7) and Eq. (8) ((a): Eq. (7), (b): Eq. (8)).

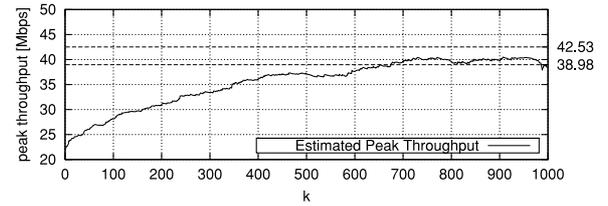


Fig. A·8 Estimated known peak throughput (relationship between k and $\hat{F}_{D,k}^{-1}(\frac{|D|}{|D|+1})$), where $x_{\max} = 38.98$ and $l_{\max} = 42.53$.

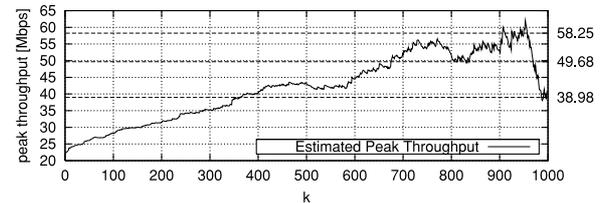


Fig. A·9 Estimated unknown peak throughput (relationship between k and $\hat{F}_{D,k}^{-1}(\frac{|D|}{|D|+1})$), where $x_{\max} = 38.98$, $\bar{x}_{\max} = 49.68$, and $l_{\max} = 58.25$.

200307091400.dump.gz.

As for peak throughput estimation, $\hat{F}_{D,k}$ can approximate *known* and *unknown* peak throughputs better than \tilde{F}_D and $L_{\hat{\mu}_D, \hat{\sigma}_D^2}$ (see Figs. A·8 and A·9) when $600 \leq k \leq 900$. Furthermore, the error in estimation of peak throughput is sufficiently small even when $500 \leq n \leq 1000$.

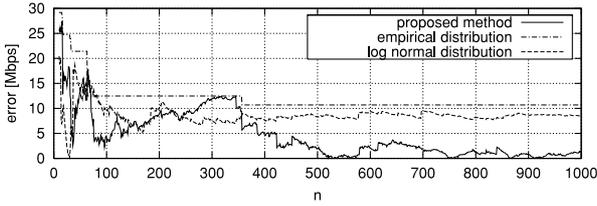


Fig. A-10 Relationship between the estimation error and the number of data used in the analysis.

Appendix D: Traffic Data Measured under Other Network Conditions

In this section, we use a trace measured under other network conditions [14], and analyze the throughput data of each 10 ms as calculated from the trace. We refer to this throughput data as “ocn-sinet” in this paper. The measurements were made on a link connecting NTT’s Open Computer Network (OCN) and the Science Information Network (SINET). OCN is the commercial Internet backbone network operated by NTT, and SINET is the largest Internet backbone network for scientific research institutes in Japan. The link is a 135-Mbps ATM circuit. The measurements were made during daily busy hours over 5 minutes on several weekdays in January 2000. For information on this data, see [14]. Note that this link provides a physically different and spatially separate condition from the WIDE network discussed in the body of this paper.

The analysis results for ocn-sinet are shown in several figures and a table, where $n^{\text{all}} = 5 [\text{min}] \times 60 [\text{sec}] \times 1000 [\text{msec}]/10 = 30000$. Table A-2 corresponds to Table 1, and Figs. A-11, A-12, . . . , A-20 correspond to Figs. 1, 2, . . . , 10, respectively.

As these figures and the table show, the analysis result for this data are almost the same as those for the WIDE data (200307071400.dump.gz and 200307091400.dump.gz), so the arguments given in the previous sections are mostly valid here, too. For example, as shown in Figs. A-14 and A-18, $\hat{F}_{D,k}$ provides a good approximation of the tail distribution and peak throughput of known data when $k = 600, 700, 800, 900$, and 980. Furthermore, as shown in Figs. A-16 and A-19, $\hat{F}_{D,k}$ provides a good approximation of the tail distribution and peak throughput of unknown data when $k = 600, 700, 800$, and 900. These characteristics are the same as the analysis results for the WIDE data. However, there are some noteworthy differences.

The fluctuation in ocn-sinet is weaker than that in the WIDE data (see Tables 1, A-1 and A-2). This is because the throughput range is narrower in the ocn-sinet data than in the WIDE data (see maximum and minimum values), and the variance of the throughput is smaller in the ocn-sinet data than in the WIDE data. This means that the ocn-sinet data was measured under different network conditions from the WIDE data.

The most significant difference between the ocn-sinet and WIDE data is found in Figs. 10, A-10, and A-20. That

Table A-2 Statistics of D and \bar{D} .

	num	min	max	ave	var	skew
D	1000	3.72	28.72	13.24	17.97	0.48
\bar{D}	29000	1.40	32.80	12.73	17.63	0.46

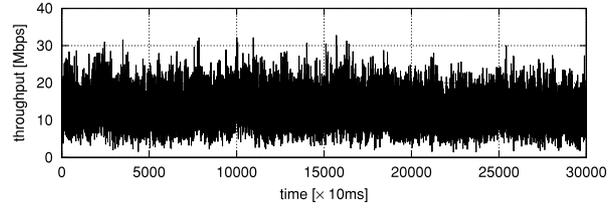


Fig. A-11 Time series of throughput.

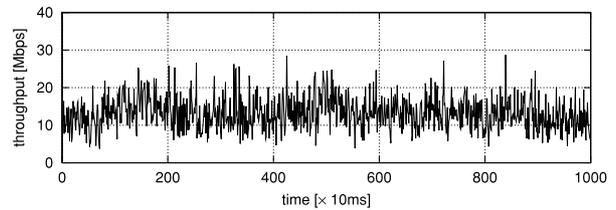


Fig. A-12 Stretched image of Fig. A-11.

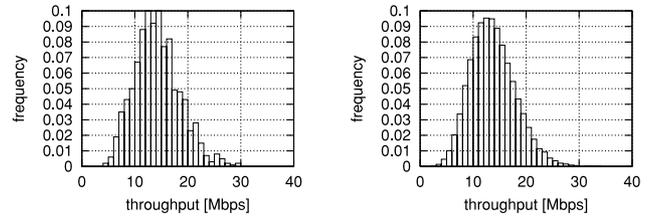


Fig. A-13 Histograms of D and \bar{D} (Left: D , Right: \bar{D}).

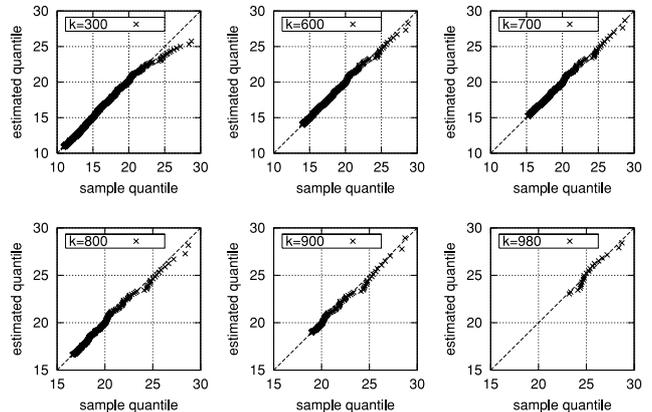


Fig. A-14 QQ-plots described by Eq. (4) for $k = 300, 600, 700, 800, 900$, and 980.

is, $|r_{\text{max},n} - \bar{x}_{\text{max},n}|$ becomes sufficiently small when $300 \leq n \leq 1000$ in Fig. A-20, while $|r_{\text{max},n} - \bar{x}_{\text{max},n}|$ becomes sufficiently small when $500 \leq n \leq 1000$ in Figs. 10 and A-10. This means that we can estimate the unknown peak throughput of the ocn-sinet data from a smaller number of known data than the case for the WIDE data. This reflects the fact

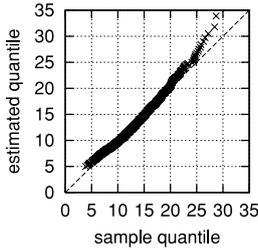


Fig. A-15 QQ-plots described by Eq. (5).

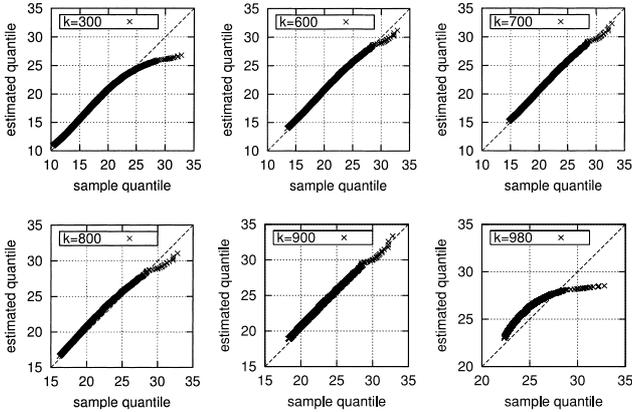


Fig. A-16 QQ-plots described by Eq. (6) for $k = 300, 600, 700, 800, 900,$ and 980 .

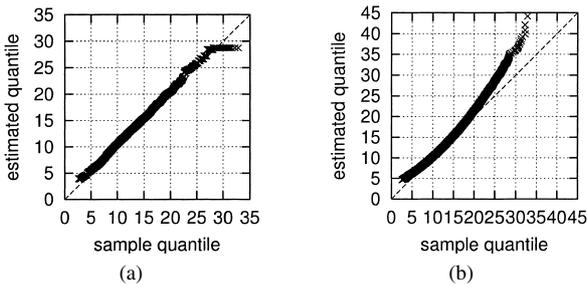


Fig. A-17 QQ-plots described by Eq. (7) and Eq. (8) ((a): Eq. (7), (b): Eq. (8)).

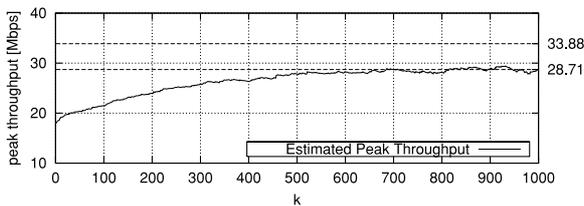


Fig. A-18 Estimated known peak throughput (relationship between k and $\hat{F}_{D,k}^{-1}(\frac{|D|}{|D|+1})$), where $x_{\max} = 28.71$ and $l_{\max} = 33.88$).

that the length of *unknown* data in the *ocn-sinet* data (290 (= 300 – 10) seconds) is shorter than in the *WIDE* data (890 (= 900 – 10) seconds). That is, the smaller the *unknown* data is, the more we can reduce the number of data for estimating the *unknown* peak throughput.

Finally, let us consider the degrees of multiplexing for

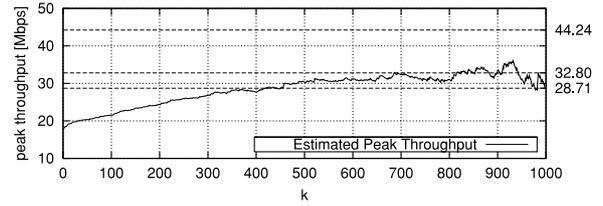


Fig. A-19 Estimated unknown peak throughput (relationship between k and $\hat{F}_{D,k}^{-1}(\frac{|D|}{|D|+1})$), where $x_{\max} = 28.71$, $\bar{x}_{\max} = 32.80$, and $l_{\max} = 44.24$).

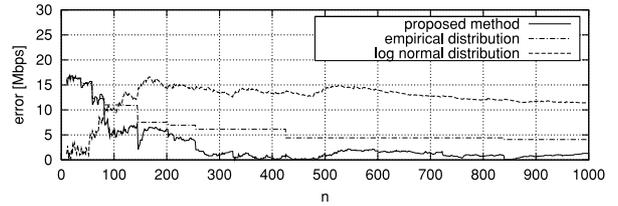


Fig. A-20 Relationship between the estimation error and the number of data used in the analysis.

Table A-3 Multiplexing level of D^{all} (0707 and 0709 correspond to 200407071400.dump.gz and 200407071400.dump.gz, respectively).

network	WIDE		OCN-SINET
	0707	0709	ocn-sinet
flow size [bytes]	4725.62	4458.86	6702.26
flow duration [sec]	20.48	22.06	18.26
flow rate [bytes/sec]	110400.50	100193.06	23015.78
connections [no./sec]	9793.52	11561.99	4786.41

the three sets of teletraffic data analyzed in this paper. Table A-3 gives the various items of statistical information that are relevant to the degrees of multiplexing: average flow size [bytes], average flow duration [sec], average flow rate [bytes/sec], and average number of connections [no./sec], where each flow is identified by a source IP address, destination IP address, source port number, destination port number, and IP protocol. The table clearly indicates different degrees of multiplexing for the *WIDE* and *ocn-sinet* data. In view of the discussion on the analysis results of these three sets for teletraffic data, we are able to say that the proposed method based on EVT is applicable to teletraffic data measured under various network conditions.



Masato Uchida received B.E. and M.E. degrees from Hokkaido University, Sapporo, Hokkaido, in 1999 and 2001, respectively. In 2001, he joined NTT Service Integration Laboratories, Tokyo, Japan. He received the research award (IEICE Communication Quality Technical Group) in 2003, and the young investigators' award (IEICE) in 2004. His research area includes teletraffic engineering and statistical learning theory.