PAPER

Characterization of Host-Level Application Traffic with Multi-Scale Gamma Model

Yosuke HIMURA[†a)], Kensuke FUKUDA[††], *Members*, Patrice ABRY[†††], Kenjiro CHO[††††], *Nonmembers*, *and* Hiroshi ESAKI[†], *Member*

SUMMARY In this paper, we discuss the validity of the multi-scale gamma model and characterize the differences in host-level application traffic with this model by using a real traffic trace collected on a 150-Mbps transpacific link. First, we investigate the dependency of the model (parameters α and β, and fitting accuracy ϵ) on time scale Δ, then find suitable time scales for the model. Second, we inspect the relations among α, β, and ϵ, in order to characterize the differences in the types of applications. The main findings of the paper are as follows. (1) Different types of applications show different dependencies of α, β, and ϵ on Δ, and display different suitable Δs for the model. The model is more accurate if the traffic consists of intermittently-sent packets than other. (2) More appropriate models are obtained with specific α and β values (e.g., $0.1 < \alpha < 1$, and $\beta < 2$ for $\Delta = 500$ ms). Also, application-specific traffic presents specific ranges of α, β, and ϵ for each Δ, so that these characteristics can be used in application identification methods such as anomaly detection and other machine learning methods.

key words: *traffic analysis, model evaluation, traffic characterization, multi-scale gamma model*

1. Introduction

Appropriate Internet traffic models are essential for evaluating the performance of network systems (i.e., network and hardware design, or QoS), traffic classification, and anomaly detection. A lot of attention has been focused on the model and characterization of the statistical distributions of traffic (i.e., flow volume [1] and traffic flow duration [2], long-range dependencies [3], application traffic (e.g., regular Web applications [4], P2P applications [5], and video streaming [6]), worm propagation [7] and other anomalous traffic [8])). In this paper, we focus on the multi-scale gamma model first proposed in [9] to model aggregated traffic and applied to detecting anomaly traffic [10]. This statistical model approximates the histogram of the number of packets arriving during a unit time as a gamma distribution on multiple scales. It can express exponential and Gaussian distributions, as well as their hybrids, which have been used to model Internet traffic. In addition, the gamma distribution is uniquely determined by using only two parameters: α determines the shape of the distribution, and β the scale. Moreover, these parameters can be easily estimated by using the moment method. Even though there have been several evaluations and validations on this model, the following questions are still open: "On which time scale is the approximation better?," "What kind of application traffic is more suitable for the model?," and "Can it characterize the difference in the type of application traffic existing in aggregated one?." Indeed, appropriate models of differences in host-level application types are essential for realistic simulation of host behaviors (i.e., reproduce host traffic according to a certain type of application), traffic engineering (e.g., put priority on streaming traffic), detecting anomalous hosts, and discovery of emerging software.

In this paper, we discuss the validity of the multi-scale gamma model and characterize the differences in application traffic with this model by using a real traffic trace collected on a 150-Mbps transpacific link. Before the analysis, we propose a timeout setting technique that is based on a stochastic theory of packet arrival to remove *meaningless* 0 s in data, which cannot be handled by common statistic distributions. We also evaluate several metrics for the fitting accuracy to reveal their applicability to the wide variety of traffic data. Then, we first discuss the dependency of the model (parameters α and β, and fitting accuracy ϵ) on time scale Δ, and find suitable time scales for the model. Second, we inspect the relations among α, β, and ϵ, in order to characterize the differences in the types of applications. The main findings of the paper are as follows. (1) Different types of applications show different dependencies of α, β, and ϵ on Δ, and display different suitable Δs for the model. The model is more accurate if the traffic consists of intermittently-sent packets than other. (2) More appropriate models are obtained with specific α and β values (e.g., $0.1 < \alpha < 1$, and $\beta < 2$ for $\Delta = 500$ ms). Also, application-specific traffic presents specific ranges of α, β, and ϵ for each Δ, so that these characteristics can be used in application identification methods such as anomaly detection and other machine learning methods. Our contributions are (a) provision of a timeout technique to deal with many 0 s in data, (b) practical evaluation of goodness-of-fit metrics for traffic analyses, and especially (c) characterization of the difference in the types of host-level traffic applicable to machine learning-based methods.

Manuscript received December 22, 2009.
Manuscript revised June 6, 2010.

[†] The authors are with The University of Tokyo, Tokyo, 113-8656 Japan.
[††] The author is with National Institute of Informatics/PRESTO, JST, Tokyo, 101-8430 Japan.
[†††] The author is with École Normale Supérieure de Lyon, France.
[††††] The author is with Research Laboratory, Internet Initiative Japan, Tokyo, 101-0051 Japan.
a) E-mail: him@hongo.wide.ad.jp

DOI: 10.1587/transcom.E93.B.3048

2. Preliminaries

In this section, we argue preliminaries needed for the traffic analyses with the multi-scale gamma model. First, we define the type of traffic, and describe the traffic data used in our analysis. Second, the multi-scale gamma model is introduced, and we explain the timeout method to handle many 0 s in data. Third, we evaluate the usability of several goodness-of-fit metrics for traffic analyses. Finally, we summarize these preliminaries and display the flow of our main evaluation.

2.1 Definition of Host-Level Application Traffic

The multi-scale gamma model was originally proposed as a model of aggregated traffic. Here, we apply it to the characterization of *application traffic* in aggregated traffic, which is defined as traffic composed of a typical kind of packets (e.g., web, DNS, e-mail, etc.) sent by a host. For the characterization of application traffic, it is essential to aggregate flows. For example, portscan traffic consists of a set of mice flows targeting numerous ports of a host. Another example is P2P application traffic composed of server and client flows. Thus, such combinations of flows produce characteristics of application traffic.

Since it is difficult to perfectly identify application traffic, we used *host-level traffic* for our analyses, which is defined as a set of packets from one host (i.e., a set of packets all of which have the same source IP address). We classify host-level traffic into *host-level application traffic* which is host-level traffic identified to be mostly composed of packets generated by a typical sort of application. For the identification of host-level application traffic, we extended the classification heuristics based on port numbers and connection patterns [11]; This heuristics classifies host-level traffic with regard to the degree of *abnormality*, and we modified it to classify the type of traffic into seven categories: FLOOD, SCAN, WWWS, WWWC, DNS, MAIL, and P2P. Table 1 lists a brief description of the categories and examples of the classification heuristics. We can capture typical applications by using this heuristics, even though there is no ground-truth of Internet traffic. For practical analyses, we focused on host-level application traffic consisting of more than 1000 packets in a trace, so as to obtain more reliable statistics. Hereafter, we refer to host-level application traffic as application traffic.

2.2 MAWI Dataset — Real Backbone Traffic Traces

Different from simulated or synthesized data, we use a real traffic trace in our evaluation. Selection of a measurement point is one of the concerns for real traffic analysis: (a) end hosts (e.g., construct web servers and client hosts, and observe their in/out traffic), or (b) backbone link. The former is practical for evaluating microscopic behavior of software, because the contents of all packets are previously known. However, it is unsuitable for characterizing the behavior of host-level application traffic existing in aggregated one; Preparation of all possible application traffic is unpractical. That may change in time and space. The behavior of prepared servers does not necessarily represent "typical" behavior of the service. Also, it is unsure whether the statistical property of combination of such flows resembles that of real aggregated traffic. Therefore, we chose to analyze backbone traffic with our heuristics of application identification, because backbone traffic is generated by different kinds of host behavior and application usages.

Our evaluation was performed with the MAWI dataset [12] — real backbone traffic traces measured on a transpacific link between the U.S. and Japan. The data consist of 15 minutes of pcap traces (from 14:00 to 14:15, JST) from 2001. The payloads of all the packets have been removed, and both source and destination IP addresses are anonymized; The prefix structure is preserved. In this study, we analyze the 2007-09-15 traffic data from a 150-Mbps link; the packet-rate and bit-rate are 21.3-Kpps and 109-Mbps, and its breakdown is listed in Table 2 (the result of host classification with our heuristic shown in Table 1).

Table 1 Categories of host-level application traffic and examples of classification heuristics.

category	explanation	example of heuristics
SCAN	The host is a port scanner	If (1) the ratio of SYN flagged packets is more than 20%, (2) the number of peers is more than 30 and (3) the number of packets sent to one host is less than 5, then the host is classified into SCAN
FLOOD	The host is performing a flooding attack	If the ratio of SYN flagged packets is more than 20% and the traffic is not classified into SCAN, then the host is regarded as "an attacker of SYN flooding
WWWC	The host is a WWW client	If the ratio of HTTP request packets is more than 50%, then the host is regarded as "a web client"
WWWS	The host is a WWW server	If the ratio of HTTP answer packets is more than 50%, then the host is regarded as "a web server"
DNS	The host is mainly sending DNS packets	If packets of source port 53 account for more than 50%, the host is regarded as "a DNS server"
MAIL	The host is mainly sending e-mail packets	If the ratio packets with source port 25 is more than 50%, the host is regarded as "a mail server"
P2P	The host is a P2P application user	If higher source and destination ports account for more than 50% of total port usage, respectively, and the packet amount of the most dominant host-to-host flow *accounts for* less than 30%, then the host is regarded as "a P2P application user"

Table 2 Breakdown of the dataset: the number of hosts identified as the corresponding category. "others" is the hosts which cannot be classified into any of the seven categories.

DNS	MAIL	SCAN	FLOOD	P2P
174	45	56	53	15

WWWS	WWWC	others	total
410	428	275	1456

2.3 Multi-Scale Gamma Model

Let Δ [sec] be the time length to aggregate packets and let $X_\Delta(t)$ be the number of packets that arrive during the time of t-th Δ. The procedure of model is mentioned as follows.

(1) For a traffic trace whose length is L [sec], construct the sequence of $X_\Delta(\cdot) = \{X_\Delta(1), \ldots, X_\Delta(T)\}$, where $T = \frac{L}{\Delta}$ (the number of the observed $X_\Delta(\cdot)$).
(2) Then, the sequence is converted into the histogram $H_\Delta = \{m_o, m_1, \ldots\}$, where m_i is the number of $X_\Delta(\cdot)$ whose value is equal to i.
(3) Finally, H_Δ is fit with a gamma distribution $f_{\alpha,\beta}(x) = \frac{1}{\beta\Gamma(\alpha)}\left(\frac{x}{\beta}\right)^{\alpha-1}\exp\left(-\frac{x}{\beta}\right)$, where $\Gamma(\cdot)$ is the gamma function. Since the gamma distribution is a normalized probability distribution, it should be adjusted to the histogram by the multiplication of T; H_Δ is approximated as $T \times f_{\alpha,\beta}(x)$.
(4) Obtain several realizations of the traffic for several time scales Δ_j $(j = 1, \ldots, J)$.

α determines the shape of the histogram, and β the scale. The distribution is close to $\frac{\beta}{x}\exp(-\frac{x}{\beta})$ for $\alpha \approx 0$, exponential for $\alpha \approx 1$, and Gaussian for larger α. On the other hand, a smaller β produces narrower distribution, and a larger β wider, keeping the same shape. The gamma distribution does not assume the equivalence between average and variance of observed data as Poisson one does. Note that a set of continuous distributions (the gamma distributions) is applied to approximating discrete histograms (H_Δ), but this does not matter because α and β are obtained from any kind of traffic, and large values of T make the discrete nature invisible.

α and β are easily estimated by using the moment method. The theoretical average $E(X)$ and variance $V(X)$ of a gamma distribution $f_{\alpha,\beta}(X)$ are expressed as $E(X) = \alpha\beta$, $V(X) = \alpha\beta^2$. Therefore, α and β of an observed histogram H_Δ are estimated as $\hat{\alpha} = \frac{\hat{E}_\Delta(X)^2}{\hat{V}_\Delta(X)}$, $\hat{\beta} = \frac{\hat{V}_\Delta(X)}{\hat{E}_\Delta(X)}$, by computing $\hat{E}_\Delta(X) = \frac{1}{T}\sum_{t=1}^{T} X_\Delta(t)$, $\hat{V}_\Delta(X) = \frac{1}{T}\sum_{t=1}^{T}(X_\Delta(t) - \hat{E}_\Delta(X))^2$. With the moment method, we only need to compute the average and variance obtained from a longitudinal pattern of traffic; Accordingly, we can easilly discuss the interpretation of results (e.g., the shape of longitudinal pattern). Also, we have compared the moment method to Maximum Likelihood Estimator (MLE) and Minimum Squared-error Estimator (MSE) in our preliminary investigation. Our conclusion from the comparison is that the moment method is enough for estimating gamma distribution's parameters (e.g., $\epsilon = 0.4$ for MoM and $\epsilon = 0.2$ for MLE and MSE on average on 100 ms, where ϵ is a goodness-of-fit metric discussed later).

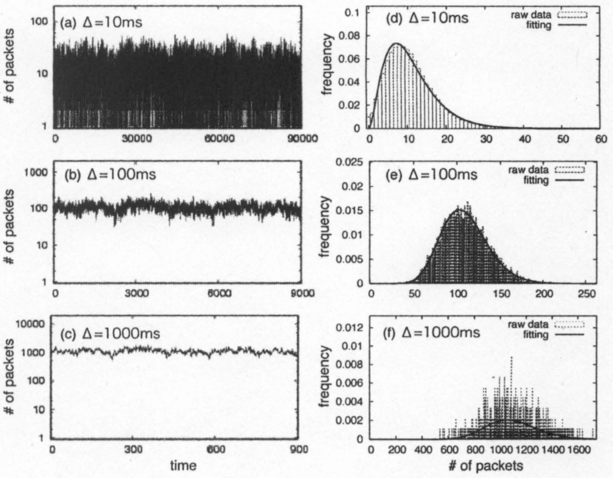


Fig. 1 Example of multi-scale gamma model with traffic generated by a host (left: longitudinal traffic on multiple scales, right: corresponding histograms (box) and estimated ones (lines)). Traffic is converted into a histogram of the number of packet arriving during a unit time, and this histogram is approximated as a gamma distribution, which is uniquely determined by two parameters α and β. This approximation is performed on multiple scales by using several unit times.

The gamma distribution is also characterized by the reproductive property; For any two independent random variables X_i $(i = 1, 2)$ that follow gamma distributions $f_{\alpha_i,\beta}$, their summation $X_1 + X_2$ follows a $f_{\alpha_1+\alpha_2,\beta}(\cdot)$. Therefore, if $X_\Delta(\cdot)$ that follows $f_{\alpha,\beta}(\cdot)$ can be treated as a random variable (i.e., any $X_\Delta(i)$ and $X_\Delta(j)$ are independent), the $X_{2\Delta}(\cdot)$ follows $f_{2\alpha,\beta}(\cdot)$ because $X_{2\Delta}(t) = X_\Delta(2t-1) + X_\Delta(2t)$. In other words, from the viewpoint of random process, α should be proportional to Δ, and β should be constant : $\alpha \propto \Delta, \beta \sim const.$

On the other hand, from the viewpoint of self-similar process [13], the average and variance follow $E_\Delta(X) \propto \Delta, V_\Delta(X) \propto \Delta^{2H}$ (or $V_\Delta(\frac{X}{\Delta}) \propto \Delta^{-2(1-H)}$), and therefore $\hat{\alpha} \propto \Delta^{2(1-H)}, \hat{\beta} \propto \Delta^{2H-1}$. $H \in (0, 1)$ is the Hurst parameter, which characterizes self-similarity. For $H = 0.5$, the process $X_\Delta(\cdot)$ indicates short-range dependency, i.e., it follows $\alpha \propto \Delta, \beta \sim const$ as above-mentioned. For $H > 0.5$, higher H leads to stronger long-range dependency of the process $X_\Delta(\cdot)$.

Figure 1 shows an example of the traffic approximation with the multi-scale gamma model, where (a), (b) and (c) display traffic (a sequence of $X_\Delta(\cdot)$ on different Δs), and (d), (e) and (f) display the corresponding histograms. Smaller Δ makes $X_\Delta(\cdot)$ lower, and the shape of the histogram is exponential, i.e., smaller α. Conversely, larger Δ makes $X_\Delta(\cdot)$ higher, and the shape of the histogram is Gaussian, i.e., higher α.

2.4 Timeout Setting to Remove Many Meaningless 0 s in Data

Traffic may include a long *silent period* where the host does not generate packets (e.g., there is no active flow), then there are a lot of continuous $X_\Delta(\cdot) = 0$. At that case, the histogram will be distorted by the such 0 s which are unrelated to the nature of the application behavior. Thus, the multi-scale gamma model cannot handle huge amount of such 0 s by itself, as other distributions cannot. One basic way to solve this problem is to exclude all the 0 s, but this is irrelevant, because the multi-scale gamma model substantially deals with 0 s in traffic, e.g., an exponential distribution ($\alpha = 1$) must produce a high probability of $X_\Delta(\cdot) = 0$.

To identify such *meaningless* 0 s, we use a timeout technique based on the Poisson arrival theory, which approximates the probability distribution of $X_\Delta(\cdot)$ as $\frac{\lambda^k e^{-\lambda}}{k!}$, where $k = X_\Delta(\cdot)$ and $\lambda = E_\Delta(X)$. If $X_\Delta(\cdot) = 0$ continues τ times, the probability is $(\frac{\lambda^0 e^{-\lambda}}{0!})^\tau = e^{-\lambda\tau}$, and we empirically regard the minimum τ that satisfies $e^{-\lambda\tau} < 0.01$ as the threshold of timeout τ_{th}; If $X_\Delta(\cdot) = 0$ continues more than τ_{th} times, we ignore these 0 s, otherwise we consider them as *meaningful* 0 s. For example, when we let τ be 3 and we observe 00000, then we remove all the five 0 s. Also, zero-inflated models and this technique have the same effect, so these models can similarly solve the "many-0 s" problem as well.

2.5 Empirical Selection of Goodness-of-Fit Metrics Suitable for Traffic Analyses

Metric requirements and candidates: To evaluate "what kind of traffic is reliably approximated with the model on which time scales," we need a goodness-of-fit metric, which is a similarity between a observed histogram and its estimated one. Such metric should be

- independent of the number of observed points T.
- independent of the number of histogram's bins B.
- compatible with human's sense, i.e., appropriate for explaining the similarity.

In other words, the first two conditions require metrics to be independent of time-scale Δ for a same goodness of fitting. Known metrics for goodness-of-fit are the P values (significances) of (a) Pearson's χ^2, (b) Kolmogorov-Smirnov, or (c) Anderson-Darling statistics, because they have been theoretically well-studied. However, Claffy et al. found that these metrics are inappropriate for wide-area network traffic data in the context of packet sampling [14], and they proposed Fleiss' ϕ coefficient by evaluating the influence of the number of observed points (sampling rate) on statistical information such as the packet size distribution. They evaluated the following metrics for the histogram similarity.

- Pearson's χ^2 statistics: $\chi^2 = \sum_{i=1}^{B} \frac{(O_i - E_i)^2}{E_i}$
- The P value of Pearson's χ^2 statistics: $P_{\chi^2} = 1 - \int_0^{\chi^2} f(x; B-1-2)dx$, where $f(x;k)$ is the χ^2 distribution of degree-of-freedom of k[†]
- Fleiss' ϕ coefficient: $\phi^2 = \frac{\chi^2}{\sum_{i=1}^{B}(E_i + O_i)}$
- Paxson's metric: $\epsilon^2 = \frac{1}{B} \sum_{i=1}^{B} \frac{(O_i - E_i)^2}{E_i^2}$
- L1 norm: $L_1 = \sum_{i=0}^{N} |m_i - \hat{m}_i|$
- relative L1 norm: $L_1' = \frac{L_1}{T}$

where O_i and E_i are the observed and expected counts of the histogram's i-th aggregated bin, B is the number of aggregated bins. $\hat{m}_i$ is the estimated value of m_i, that is, $\hat{m}_i = T \times \int_{i-0.5}^{i+0.5} f_{\hat{\alpha},\hat{\beta}}(x)dx$; Gamma distributions are discretized to be more comparable with discrete traffic data. N is a constant value to be used as an upper bound of the $X_\Delta(\cdot)$s for all H_Δs. Consequently, every histogram is uniformly represented with the form of $H_\Delta = \{m_0, m_1, \ldots, m_N\}$ by using a fixed N. O_i should avoid containing zero or only a few counts (e.g., less than 5) in order to compute more reliable statistics [15]. To achieve this condition, we adopt the way to use flexible (unfixed) widths for each O_i, by letting the bin include at least 10%[††] of the total count T as follows: $O_j = \sum_{i=N_j}^{N_{j+1}-1} m_i$, satisfying $O_j > 0.1 \times T$, and the range of O_j is $[N_{j-1}, N_j - 1]$. This is iteratively computed from $j = 1, N_0 = 0$ to $j = B$.

Since their evaluation was focused on the effect of packet samplings on the distribution of the packet size and packet inter-arrival time, we re-evaluated these metrics with respect to the time scale. Moreover, we additionally studied the following metrics.

- Kolmogorov-Smirnov statistics: $KS = \sup|F_i - \hat{F}_i|$
- The P value of Kolmogorov-Smirnov statistics: $P_{KS} = 1 - 2\sum_{i=1}^{\infty}(-1)^{i-1}\exp(-2i * T * KS^2)$
- Sum-of-square error: $SSE = \sum_{i=0}^{N}(P_i - \hat{P}_i)^2$

where P_i is the probability of "$X_\Delta(\cdot) = i$" (i.e., $P_i = \frac{m_i}{T}$), and F_i is the normalized count of the cumulative histogram (i.e., $F_i = \sum_{h=0}^{i} P_h$). We also examined the Anderson-Darling statistics, but it cannot be computed, when H_Δ includes at least one $X_\Delta(\cdot)$ satisfying "$F_{\hat{\alpha},\hat{\beta}}(X_\Delta(\cdot)) = 0$ *or* 1," where $F_{\hat{\alpha},\hat{\beta}}(\cdot)$ is the cumulative distribution of the estimated gamma function $f_{\hat{\alpha},\hat{\beta}}(\cdot)$; This was common case for our experiment. Note that the some of the uses of above metrics are not standard from the viewpoint of statistical theory (e.g., applying KS to measuring the similarity between continuous distribution and discrete one). Nevertheless, we need an empirical metric to quantify the fitting accuracy, even though some of the metrics do not necessarily follow their theory.

Practical usability of metrics with real traffic: At first, we evaluated the dependency of metrics on the number of observed points T, and we found that ϵ, ϕ, SSE, L_1', and KS are suitable for traffic analyses but χ^2 and L_1 are unsuitable; The details are described in Appendix A. Secondly, we also evaluated the practical usability of metrics with real traffic data, which have various characteristics of longitudinal patterns and the shapes of their histograms; This evaluation can also discuss the dependency of the metrics on the

[†] $f(x;k) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{(k/2)-1} e^{-x/2}$. Since we estimate two parameters ($\hat{\alpha}$ and $\hat{\beta}$) from H_Δ, the degree-of-freedom k (= $B - 1$) must be reduced by 2.

[††] The term *10%* can be replaced with other values if we can keep the condition on O_i for each fitting.

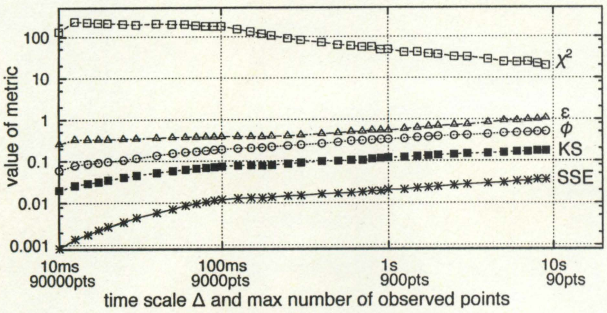


Fig. 2 Evaluation of goodness-of-fit metrics: Median values of metrics among about 1200 identified hosts. ϵ, ϕ, SSE, KS present the same tendency, and practically ϵ exhibits the smallest dependency on Δ. The distribution of ϵ will be discussed in the following section (Fig. 5).

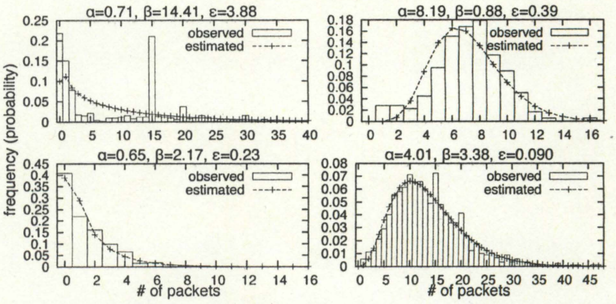


Fig. 3 Examples of fitting (box: observed histogram, line: estimated histogram) and the values of parameters ($\hat{\alpha}$, $\hat{\beta}$, and ϵ).

number of bins B. The desirable situation is that the values of the parameters (ϵ, ϕ, SSE, L_1', and KS) do not fluctuate according to the changes in Δ, considering the independency of n shown in the previous discussion. We evaluated this applicability with approximately 1200 pieces of identified host-level application traffic shown in Table 2, and Figure 2 depicts the result. The x-axis is Δ (and the upper bound of the number of obtained points T), the y-axis is the value of parameters, and each symbol displays the type of metrics, plotting median values among the about 1200 trials (we excluded L_1 and L_1' from the figure for the visibility). The values of ϵ, ϕ, L_1', KS, and SSE gradually increase when Δ becomes larger, which means that the model is less accurate for larger Δ; The difficulty in approximating histograms on larger Δ should result from the high variability in obtained data $X_\Delta(\cdot)$. Even though we have shown that ϵ, ϕ, L_1', KS, and SSE can be used for the goodness-of-fit, the difference in their dependencies on Δ let us to choose the most appropriate one. For example, SSE highly changes for small Δ, that is, SSE can easily be smaller for exponential histograms; Actually, when we compare two different-shaped histograms of similar errors for human intuition, we found a big difference in the values of SSE. Thus, SSE is strongly affected by the shape of histograms, and the metric is less suitable for quantifying fitting accuracies. Finally, we adopted ϵ as one of the most relevant parameters because of the smallest dependency on Δ[†], and this should also lead to the independency of the number of histogram's bins B[††]. Appendix B shows that similar evaluation results will be obtained with KS and ϕ as well as ϵ.

To summarize, χ^2 and L_1 are completely unsuitable because of the dependency on T. In addition, even though the other metrics are free from T, the slight difference in the dependency on Δ recommends us to use ϵ. Note again that high value of ϵ indicates worse fitting, and vice versa. Figure 3 displays several examples of the fitting and corresponding ϵs. By looking at numerous figures for the fitting, we determined $\epsilon = 1$ to be used as a criteria to separate good fittings and bad ones. This threshold was empirically set because there is no scientific way to judge such a "good or bad," which is visually decided by human.

2.6 Summary of Preliminaries and Evaluation Flow

Here we display the flow of our evaluation by summarizing the preliminaries.

1. Aggregated traffic is divided into a set of host-level traffic, and each of the traffic is categorized with our heuristics.
2. Each application traffic is approximated as gamma distributions on multiple scales with our timeout method (preprocess to remove *meaningless* 0 s).
3. Obtain $\hat{\alpha}, \hat{\beta}$, and ϵ for each Δ from a piece of traffic, and investigate the correlation among them considering the category.

3. Evaluation

3.1 Time Scale Dependency

Figure 4 displays the time scale dependency of (a) $\hat{\alpha}$, (b) $\hat{\beta}$, and (c) ϵ for each category (DNS, MAIL, SCAN, FLOOD, WWWS, WWWC, and P2P). The x-axis is the value of Δ, the y-axes are the values of the parameters, each symbol represents each category, plotting the values of median among the hosts identified as the corresponding category with log-log scales.

3.1.1 Dependency of $\hat{\alpha}$ and $\hat{\beta}$ (Figs. 4(a) and 4(b))

For smaller Δ, the estimated values of $\hat{\alpha}$s monotonically increase, while those of the $\hat{\beta}$s are stable; This follows the theory discussed in Sect. 2.3. Histograms evolve their shapes from $\frac{\beta}{x}\exp(-\frac{x}{\beta})$ to exponential (or to Gaussian), keeping their scales. The order of the $\hat{\alpha}$ values per application is

[†] From the viewpoint of statistics, the best metric of goodness-of-fit is the P value of the statistics, which is a normalized metric to be comparable among all fittings. However, 31.2% of P_{KS} values were computed to be 0 to the 20th decimal place as well as 76.2% of P_{χ^2} values, for $\Delta = 100$ ms as an example.

[††] We have confirmed that KS and ϕ can be used as well as ϵ; Pearson's correlation coefficient among the metrics are 0.78 for ϵ and KS, and 0.96 for ϵ and ϕ, which were computed with the same set of hosts as Fig. 2 on $\Delta = 500$ ms in log-log plot.

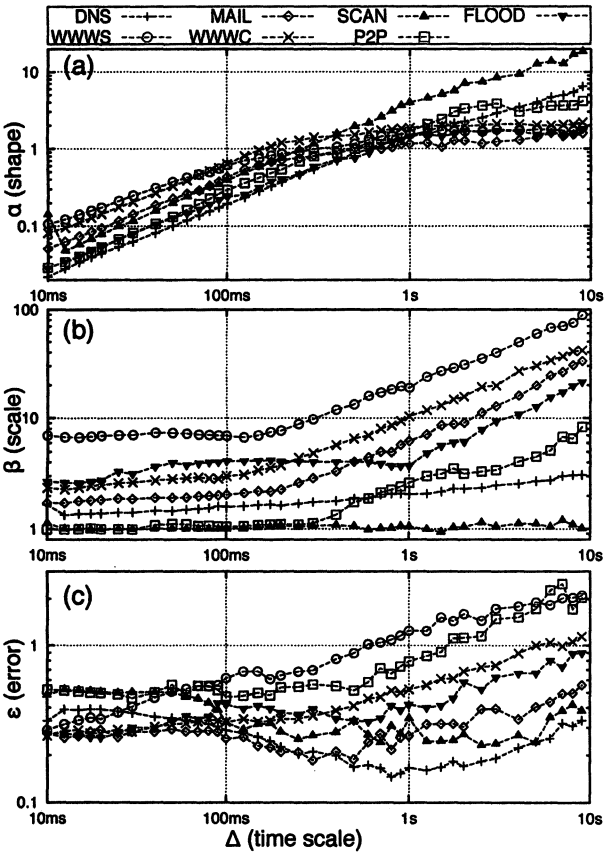


Fig. 4 Time scale analyses on the parameters ($\hat{\alpha}$, $\hat{\beta}$, and ϵ). One plotted point is the median value among all host classified as the corresponding category.

as follows: DNS < P2P < FLOOD < SCAN < MAIL < WWWC < WWWS < 1. Higher $\hat{\alpha}$ is observed in higher $X_\Delta(\cdot)$ values. For example, DNS hosts continuously send constant amounts of packets, so the value of $X_\Delta(\cdot)$ is generally low. Then, the histogram is mainly represented by m_0 and m_1 (not by higher m_i), making its shape close to $\frac{\beta}{x}\exp(-\frac{x}{\beta})$ (i.e., lower $\hat{\alpha}$). Contrarily, for another example, WWWS hosts abruptly send large amount of packets (e.g., transferring large files), so $X_\Delta(\cdot)$ can be higher (i.e., higher $\hat{\alpha}$). Also, the order of $\hat{\beta}$ is as follows: P2P < SCAN < DNS < MAIL < WWWC < FLOOD < WWWS; Higher $\hat{\beta}$ is confirmed in the high variance of $X_\Delta(\cdot)$. For example, WWWS hosts transfer large files and stop sending packet to wait for the next requests, resulting in a high $X_\Delta(\cdot)$ variance which makes a large-scale histogram, i.e., higher $\hat{\beta}$.

On the other hand, for larger Δ, the shape of the plots changes (except for DNS and SCAN) with the Δ specific to the type of application: 200 ms for WWWC, 300 ms for WWWS, 500 ms for MAIL, 700 ms for P2P, 1 s for FLOOD, and no such Δ for DNS and SCAN — we refer these Δs as *inflection points*. From the inflection points, the values of $\hat{\alpha}$ are stable, whereas those of $\hat{\beta}$ increase, i.e., histograms enlarge in scale, keeping their shape. This disagrees with the theory, which implies that $X_\Delta(\cdot)$ is no longer regarded as a random variable. The inflection point can be a measure for the degree of longitudinal dependency. Lower inflection point leads to stronger dependency such as WWWC and WWWS traffic; Indeed, their traffic patterns should be characterized by their contexts (e.g., file transferring between web servers and clients, or human's mouse clicking on the Web). Conversely, higher or no inflection point is derived from the weaker longitudinal dependency of traffic. For example, SCAN hosts arbitrarily send packets without interaction among hosts, and DNS hosts mainly send only one packet to a requester host, so that there are weaker longitudinal relations among packets generated by such hosts. Traffic pattern of flooding attack is generally intermittent and spiky, and this is a plausible reason of FLOOD's inflection point at larger time scale. Packets arbitrarily sent by flooding-attacker hosts are less related due to the lack of application protocol mechanism, but it forms traffic dependency in larger time scales, which might have caused the inflection point for large Δ (e.g., 1 s). We have confirmed this by manually inspecting multi-scale traffic pattern for each individual host.

From the viewpoint of self-similarity, the Hurst parameter in host-level traffic is almost 0.5 (i.e., $\hat{\alpha} \propto \Delta$ and $\hat{\beta} \sim$ *const*) for Δ smaller than inflection points, and almost 1 (i.e., $\hat{\alpha} \sim$ *const* and $\hat{\beta} \propto \Delta$) for Δ larger than those points. Similar results of the Hurst parameter in Internet traffic have also been discussed at the level of aggregated traffic [3], [16], or sub-aggregated (hashed) traffic [17]. This figure confirms that the characteristics of "$H \sim 0.5$ on microscopic scales" is independent of the type of application, mainly due to protocol mechanisms.

In summary, the difference in the type of application traffic is characterized by the values of $\hat{\alpha}$ and $\hat{\beta}$ as well as their dependencies on Δ. We emphasize that multi-scale view is more appropriate for characterizing application traffic.

3.1.2 Dependency of ϵ (Fig. 4(c))

We confirm two differences among all the types of application traffic in the figure; One is the difference in Δ producing smaller ϵs (better fitting), and the other is in their ranges; ϵs for DNS, MAIL, and SCAN vary according to Δ. DNS is well fit for Δs ∈ 500 ms-2 s, MAIL for Δs ∈ 200–500 ms, and SCAN for Δs ∈ $2s$-$5s$[†]. The other categories present that larger Δ yields to larger ϵ. Also, DNS, MAIL, and SCAN are well fit than the others (lower ϵs). One plausible explanation of this is as follows: These kind of hosts intermittently send packets, which makes $X_\Delta(\cdot)$ stable. Then, this leads to less outlier of $X_\Delta(\cdot)$.

3.2 Distribution of Parameters

Here, we discuss the shape of distributions of the param-

[†] The threshold to determine the range of appropriate Δs was set as the Δ value producing ϵ which value is 120% of the minimum ϵ in a line in Fig. 4 for each category.

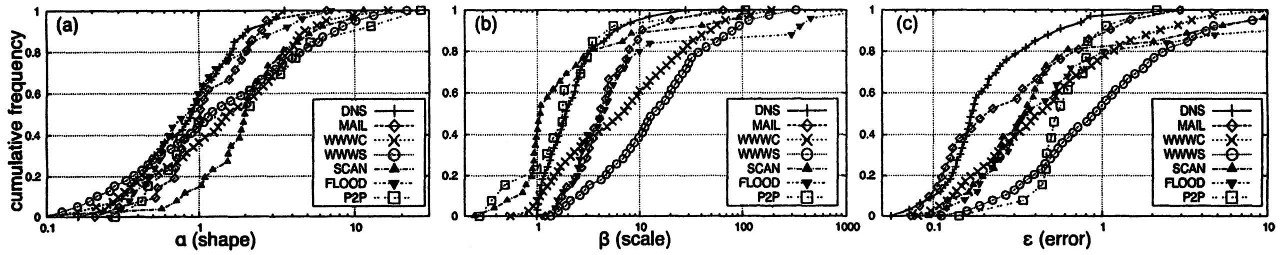


Fig. 5 Cumulative distribution of parameters for $\Delta = 500$ ms: (a) $\hat{\alpha}$, (b) $\hat{\beta}$, and (c) ϵ. The points are sampled with the ratio of (1) 90% for WWWS and WWWC, (2) 80% for DNS, (3) 50% for MAIL, FLOOD, and SCAN, and (4) 0% for P2P.

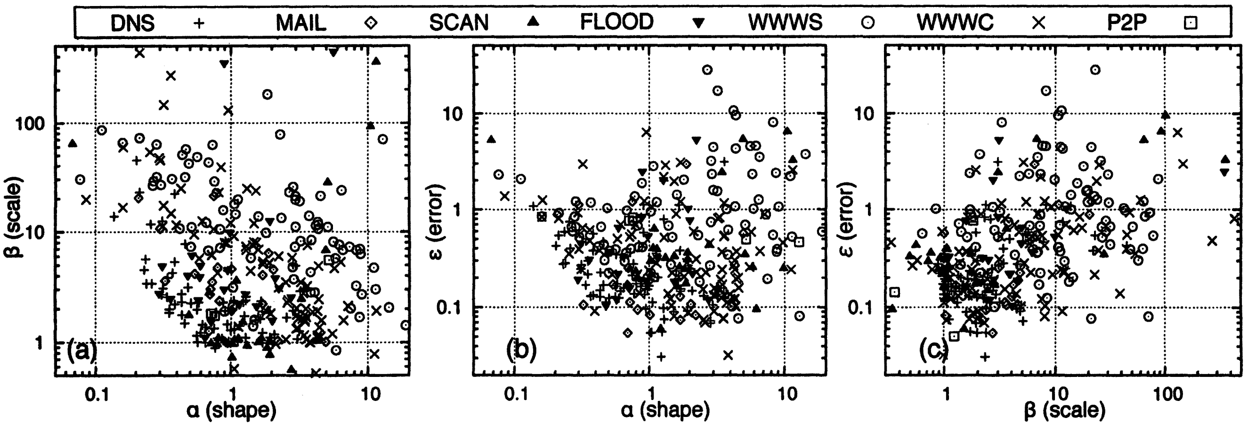


Fig. 6 Parameter correlations among $\hat{\alpha}$, $\hat{\beta}$, and ϵ for $\Delta = 500$ ms: (a) $\hat{\alpha}$ vs. $\hat{\beta}$, (b) $\hat{\alpha}$ vs. ϵ, and (c) $\hat{\beta}$ vs. ϵ. The points are sampled with the ratio of (1) 80% for WWWS and WWWC, (2) 50% for the other categories. Typical ranges of the parameters for each category are displayed in Table 3 in a quantitative way.

eters. Figure 5 shows the cumulative distributions of (a) $\hat{\alpha}$, (b) $\hat{\beta}$, and (c) ϵ for $\Delta = 500$ ms. $\hat{\alpha}$s and $\hat{\beta}$s of MAIL, DNS, SCAN, and FLOOD are densely concentrated on specific ranges rather than the other categories. In particular, FLOOD has two dense areas of $\hat{\beta}$ that derive from two kinds of flooding attackers. Contrarily, WWWS, WWWC, and P2P lead to wide ranges of values, meaning that various usages of HTTP and P2P software (e.g., file transferring and video streaming) produce diverse forms of traffic (heterogeneous $\hat{\alpha}$ and $\hat{\beta}$). On the other hand, for the parameter ϵ, DNS and P2P concentrate the values on $\epsilon \approx 0.2$ and $\epsilon \approx 0.5$, while those of the others are almost uniformly distributed. The amount of inappropriate fittings (e.g., $\epsilon > 1$) is larger for WWWS, WWWC, SCAN, and FLOOD. We obtained similar results for other Δs.

3.3 Parameter Relation

Figure 6 shows the relations among $\hat{\alpha}$, $\hat{\beta}$, and ϵ for $\Delta = 500$ ms. The x-axes and y-axes of all the figures indicate the values of the parameters with log-log scales. The symbols represent the types of application traffic, plotting the parameters of traffic.

$\hat{\alpha}$ vs. $\hat{\beta}$ (Fig. 6(a)): There are specific ranges according to the application type, that is, different types of applications have different traffic behavior. If there were no difference in host behaviors, then a line $\alpha\beta = const$ (number of packets) would theoretically appear in the figure.

- FLOOD is characterized by $\hat{\alpha} \approx 1$ and larger $\hat{\beta}$ (exponential and large-scale histogram).
- WWWS and WWWC exhibit wide ranges of $\hat{\alpha}$s and $\hat{\beta}$s (i.e., hybrid-shape and variable-scale histogram).
- SCAN and P2P have wider $\hat{\alpha}$ ranges and narrower $\hat{\beta}$ ranges than WWWS and WWWC. However, most of SCAN are concentrated on $\hat{\alpha} \approx 1$ and $\hat{\beta} \approx 1$.
- DNS and MAIL more concentrate $\hat{\alpha}$s on $[0.5, 1]$ and $\hat{\beta}$s on $[1, 5]$ (i.e., exponential and small-scale histogram), indicating typical DNS and MAIL traffic patterns.

WWWS, WWWC, and P2P spread wide ranges of $\hat{\alpha}$s (from 0.1 to 5) and $\hat{\beta}$s (from 1 to 50) compare to DNS and MAIL; This is derived from diverse usages of protocols and software as discussed in Sect. 3.2.

$\hat{\alpha}$ vs. ϵ (Fig. 6(b)): Considering $\epsilon = 1$ as an empirical threshold of fitness, there are some specific $\hat{\alpha}$ values with $\epsilon > 1$: $\hat{\alpha} < 0.1$ and $\hat{\alpha} > 1$. In particular, SCAN, FLOOD, WWWS, and WWWC tend to be plotted their parameters on these areas. Fitting of these four categories with these ranges of $\hat{\alpha}$ are less reliable than those with others. On the other hand, in the other range of $\hat{\alpha}$ $(0.1 < \hat{\alpha} < 1)$, the data

Table 3 Ranges of parameters $\hat{\alpha}$, $\hat{\beta}$, and ϵ for $\Delta = 500$ ms (estimated from Fig. 6).

$\Delta = 500$ ms	$\hat{\alpha}$	$\hat{\beta}$	ϵ
DNS	[0.330, 1.988]	[0.970, 5.446]	[0.001, 0.018]
MAIL	[0.442, 2.218]	[1.881, 10.227]	[0.001, 0.027]
SCAN	[0.507, 7.275]	[0.331, 10.640]	[0.002, 0.059]
FLOOD	[0.372, 2.080]	[1.108, 48.962]	[0.003, 0.036]
WWWC	[0.387, 4.916]	[1.366, 31.108]	[0.003, 0.055]
WWWS	[0.292, 5.546]	[3.422, 60.367]	[0.008, 0.109]
P2P	[0.338, 6.480]	[0.392, 7.615]	[0.007, 0.176]

Table 4 The number of hosts which produce relevant fitting ($\epsilon < 1$) for $\Delta = 500$ ms.

DNS	MAIL	SCAN	FLOOD
167 of 174	40 of 45	46 of 56	41 of 53

WWWS	WWWC	P2P	total
221 of 410	325 of 428	13 of 15	853 of 1181

are appropriately fit especially for DNS and MAIL; These two categories concentrate their parameter values on $\alpha \approx 1$ and $\epsilon \approx 0.2$.

$\hat{\beta}$ vs. ϵ (Fig. 6(c)): Fittings with $\hat{\beta} > 2$ tend to result in $\epsilon > 1$, so higher $\hat{\beta}$ leads to less reliable fittings (e.g., about a half of fittings with $\hat{\beta} \approx 10$ are inappropriate). In particular, SCAN, FLOOD, WWWS, and WWWC spread their parameters on the area of $\hat{\beta} > 2$. DNS and MAIL also concentrate their parameters on $\hat{\beta} \approx 2$ and $\epsilon \approx 0.2$.

To summarize, $\hat{\alpha}$, $\hat{\beta}$, and ϵ are related each other, and each category has specific ranges of the parameters. Table 3 lists the ranges of the parameters estimated by each application for $\Delta = 500$ ms. A range represents its average and standard deviation in log scale $10^{\mu \pm \sigma}$: for $\hat{\alpha}$ as an example, $\mu = E[\log_{10}(\hat{\alpha})]$, and $\sigma^2 = V[\log_{10}(\hat{\alpha})]$ (same for $\hat{\beta}$ and ϵ, mutatis mutandis). We also obtained application-specific ranges of $\hat{\alpha}$, $\hat{\beta}$, and ϵ for other Δs.

4. Discussion

Validity of the model: 853 of 1181 identified hosts produced $\epsilon < 0.1$ for $\Delta = 500$ ms; The application breakdown of the validity is shown in Table 4. About 90% of DNS, MAIL, and P2P traffic, and about 80% of SCAN, FLOOD, and WWWC traffic can be fit well, but only 50% of WWWS traffic leads to better fittings. Therefore, the multi-scale gamma model is more applicable to traffic outside of WWWS traffic for this time scale.

Generality of our dataset: To generalize our results, we should compare the results with those obtained for different network conditions, which consist of different (1) link capacity, (2) traffic bandwidth, and (3) breakdown of traffic. Our preliminary work [18] analyzed 8-year-long MAWI traces (15 min. per day), and evaluated the changes in $\hat{\alpha}$ and $\hat{\beta}$ from 2001 to 2008 — during this period, the link capacity was upgraded twice, and the bandwidth and breakdown were evolving. The following is a summary of the results. (a) The $\hat{\alpha}$s and $\hat{\beta}$s values of DNS, MAIL, WWWS, and WWWC were stable for a link, while those of P2P, SCAN, and FLOOD variable over time. (b) The link upgrades increased the $\hat{\beta}$ values of WWWS, WWWC, MAIL, and FLOOD, not affecting the $\hat{\beta}$s of the others and the $\hat{\alpha}$s of all the categories. (c) Furthermore, there are typical $\hat{\alpha}$ and $\hat{\beta}$ values that were specific to the type of application for each link. Hence, from our empirical results we confirmed the stability of the characteristics obtained from the multi-scale gamma model. We have to investigate the changes in ϵ and search for appropriate Δs in our future work.

Application to anomaly detection and other machine learning algorithms: The previous discussion pointed out the specific ranges of the parameters for each application; These parameters can be a part of features to identify the type of application traffic generated by unknown host, or to detect anomaly traffic. Since the use of encrypted packet and dynamic port hides "visible" traffic characteristics (e.g., packet payload and port numbers), our parameters should be promising in traffic classification. However, in Fig. 6 and Table 3, there are overlaps in the parameter ranges among different types of applications, so that other features should be added to our parameters for ensuring a more efficient identification. As a future work, we will investigate the applicability of the characteristics to detailed classifications (e.g., the classification of web traffic into conventional web, SNS, Twitter, WebDAV, VPN, P2P, video streaming, HDTV, and so on) by creating more fine-grained application-level traffic.

Application to reproduction of application traffic: Figs. 4, 5, and 6, and Table 3 explained the typical behaviors of application traffic (specific values of $\hat{\alpha}$ and $\hat{\beta}$). Thus, application traffic (i.e., distribution of the number of packets arriving in a unit time) can be regenerated by setting α and β values. However, note that there should be less reliable reproduction with the α and β values specific to the type of application. An example of the values is $\hat{\alpha} < 0.1$, $\hat{\alpha} > 1, \hat{\beta} > 2$ of SCAN, FLOOD, WWWS, and WWWC for $\Delta = 500$ ms as discussed in Sect. 3.3. To extend the reproduction of the distribution to that of time-series traffic data, this model should be combined with a longitudinal model such as the AFRIMA model [9].

Limitations of the multi-scale gamma model: Here we note the limitations of the model. One is common in statistical manners, that is, only a few data points (e.g., $T = 10$) will lead to unreliable statistics (α, β, and ϵ). The amount of data required for reliable statistics is variously up to the behavior of traffic, and this is currently under investigation. Another limitation is that the model cannot purely reproduce power-low distribution, even though it can represent temporal correlation in a certain time scale by superposing exponential distributions.

5. Conclusion and Future Work

In this paper, we discussed the statistics concerning the multi-scale gamma model and characterized the difference in host-level application traffic, by using a real backbone traffic trace measured on a 150-Mbps transpacific link in

2007. We first discussed the appropriate metric to evaluate the relevance of the model, then we evaluated the relations among the variables of the models (α and β), fitting accuracy ϵ, and time scale Δ. Our main findings were as follows: (1) Different types of applications show different dependencies of α, β, and ϵ on Δ, and they display different suitable Δs for the model. The model is more accurate if the traffic consists of intermittently-sent packets than other. (2) More appropriate models are obtained with specific α and β values (e.g., $0.1 < \alpha < 1$, and $\beta < 2$ for $\Delta = 500$ ms). Also, application-specific traffic presents specific ranges of α, β, and ϵ for each Δ, so that these characteristics can be used in application identification methods such as anomaly detection and other machine learning methods. In the future, we will improve our classification method by introducing the idea of BLINC [19] and use traffic traces from other links.

References

[1] K. Papagiannaki, N. Taft, S. Bhattacharyya, P. Thiran, K. Salamatian, and C. Diot, "A pragmatic definition of elephants in Internet backbone traffic," ACM SIGCOMM IMW'02, pp.175–176, 2002.

[2] J. Charzinski, "Internet client traffic measurement and characterisation results," ISSLS 2000, 2000.

[3] W.E. Leland, M.S. Taqqu, W. Willinger, and D.V. Wilson, "On the self-similar nature of ethernet traffic," ACM SIGCOMM CCR, vol.25, no.1, pp.202–213, 1995.

[4] M.E. Crovella and A. Bestavros, "Self-similarity in world wide web traffic: Evidence and possible causes," IEEE/ACM Trans. Netw., vol.5, no.6, pp.835–846, 1997.

[5] J.A. Pouwelse, P. Garbacki, D.H.J. Epema, and H.J. Sips, "The bittorrent P2P file-sharing system measurements and analysis," USENIX IPTPS'05, 2005.

[6] P. Gill, M. Arlitt, Z. Li, and A. Mahanti, "YouTube traffic characterization: A view from the edge," ACM IMC'07, pp.15–28, 2007.

[7] D. Moore, C. Shannon, and J. Brown, "Code-red: A case study on the spread and victims of an Internet worm," ACM SIGCOMM IMW'02, pp.237–284, 2002.

[8] W. John and S. Tavelin, "Analysis of Internet backbone traffic and header anomalies observed," ACM IMC'07, pp.111–116, 2007.

[9] A. Scherre, N. Larrieu, P. Owezarski, P. Borgnat, and P. Abry, "Non-Gaussian and long memory statistical characterisations for Internet traffic with anomalies," IEEE Trans. Dependable and Secure Computing, vol.4, no.1, pp.56–70, 2006.

[10] G. Dewaele, K. Fukuda, P. Borgnat, P. Abry, and K. Cho, "Extracting hidden anomalies using sketch and non Gaussian multiresolution statistical detection procedure," ACM SIGCOMM LSAD'07, pp.145–152, 2007.

[11] Y. Himura, K. Fukuda, K. Cho, and H. Esaki, "An automatic and dynamic parameter tuning of a statistics-based anomaly detection algorithm," IEEE ICC2009, p.6, 2009.

[12] K. Cho, K. Mitsuya, and A. Kato, "Traffic data reposity at the WIDE project," USENIX 2000 FREENIX Track, 2000.

[13] P. Abry, R. Baraniuk, P. Flandrin, R. Riedi, and D. Veitch, "The multiscale nature of network traffic: Discovery, analysis, and modelling," IEEE Signal Process. Mag., vol.19, no.32, pp.28–46, 2002.

[14] K.C. Claffy, G.C. Polyzos, and H.W. Braun, "Application of sampling methodologies to network traffic characterization," ACM SIGCOMM'93, pp.194–203, 1993.

[15] P.E. Greenwood and M.S. Nikulin, A Guide to Chi-Squared Testing, John Wiley & Sons, 1996.

[16] Y. Ji, "Multi-scale Internet traffic analysis using piecewise self-similar processes," IEICE Trans. Commun., vol.E89-B no.8, pp.2125–2133, Aug. 2006.

[17] P. Borgnat, G. Dewaele, K. Fukuda, P. Abry, and K. Cho, "Seven years and one day: Sketching the evolution of internet traffic," IEEE INFOCOM 2009, pp.711–719, 2009.

[18] Y. Himura, K. Fukuda, K. Cho, and H. Esaki, "Quantifying host-based application traffic with multi-scale gamma model," PAM2009 Student Workshop, p.2, 2009.

[19] T. Karagiannis, K. Papagiannaki, and M. Faloutsos, "BLINC: Multilevel traffic classification in the dark," ACM SIGCOMM'05, pp.229–240, 2005.

Appendix A: Dependency of Metrics on the Number of Obtained Points with Simulated Traffic

We conducted a numerical evaluation of the dependency of metrics on the number of points T to exclude other possible effects on the metrics. The basic idea of this evaluation is to measure the stability of metrics by repeatedly approximating traffic as gamma distributions with probabilistic estimation errors, for each T. The detailed evaluation flow is described as follows.

(1) Set the values of the parameters α and β.
(2) Generate n points following $f_{\alpha,\beta}(\cdot)$. The set of these n points is converted into the corresponding histogram H.
(3) Set the values of the estimated parameter $\hat{\alpha}$ and $\hat{\beta}$ as $\hat{\alpha} = \alpha + e \times \alpha$, $\hat{\beta} = \beta + e' \times \beta$, where e and e' are estimation errors following the standard normal distribution $N(0, 1)$[†].
(4) Compute the fitting error between H and $f_{\hat{\alpha},\hat{\beta}}(\cdot)$.
(5) Repeat (2), (3), and (4) for 10000 times.

Figure A·1 shows the result of the evaluation with $\alpha = 4.0, \beta = 4.0$. The horizontal line represents n, and the vertical line plots median values among the 10000 trials for the several candidate metrics (we removed L_1 and L_1' from the

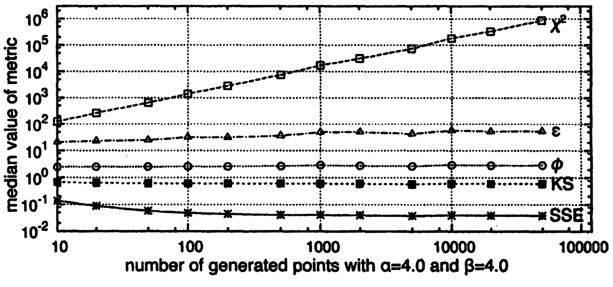


Fig. A·1 Evaluation of goodness-of-fit metrics: Median values of metrics (5000-th best values among 10000 trials) computed by using simulated traffic with $\alpha = 4.0, \beta = 4.0$, and error-added estimation of the parameters. Obviously, χ^2 and L_1 are inappropriate for the fitting accuracy, because they depend on the number of generated points n. On the contrary, ϵ, ϕ, SSE, L_1', and KS are independent of n so that they can be used as the fitting accuracy.

[†] The codomain of the errors is limited to $[-1.0, 2.58]$. The lower bound keeps the parameters more than 0, and the upper one (99%-tile of the distribution) avoids producing outliers.

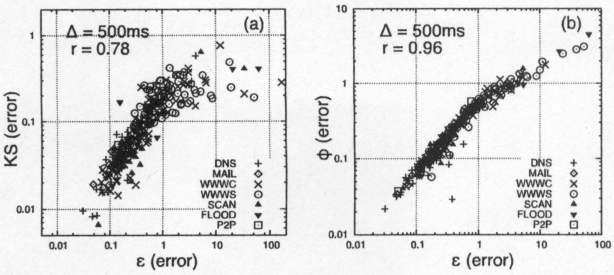


Fig. A·2 Comparing ϵ, ϕ, and KS.

figure for the visibility). We note that the values of the metrics do not represent ideal fittings, because each value means 5000-th best fitting with error-added simulation, and thus the fitting is worse than ideal case; Our motivation to show this figure is to find metric(s) robust to the number of data points. Obviously, χ^2 and L_1 depend on n, which may result from the difference in the dimension between denominator and numerator. Hence, these two metrics are unsuitable for the fitting accuracy. On the contrary, ϵ, ϕ, SSE, L'_1, and KS are independent of n, so that they can be used as the fitting accuracy. We obtained similar results for other αs and βs.

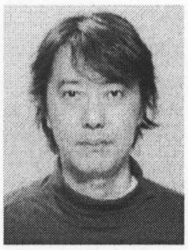
Appendix B: The Use of Other Appropriate Metrics

This appendix explains that the use of ϕ or KS instead of ϵ will lead to similar results, by investigating the correlation among the three goodness-of-fitting metrics. Figure A·2 shows the correlation among the three parameters, plotting hosts (points) for each category (symboles) on $\Delta = 500$ ms with the traffic trace used throughout this paper. This figure displays that ϵ, ϕ, and KS are positively correlated, and the value of Pearson's correlation coefficients are (a) 0.78 and (b) 0.96 in log-log scale. The weaker correlation of 0.78 is mainly derived from the plotted points in the large-error area ($\epsilon > 1$) because of the difference in codomain ($KS \in [0, 1]$, and $\epsilon \in [0, \infty)$); Since worse fittings w.r.t. ϵ (e.g., $\epsilon > 1$) are plotted also on that w.r.t. KS, the use of both the metric is practically appropriate in order to represent bad fitting as well as good fitting. Hence, the examination results with ϕ and KS are probably similar to those with ϵ.

Yosuke Himura is a master course student in Department of Information and Communication Engineering, Graduate School of Information Science and Technology, the University of Tokyo. His research interests are Internet traffic analysis and Internet security.

Kensuke Fukuda is an associate professor at the National Institute of Informatics (NII) and is a researcher, PRESTO, JST. He received his Ph.D. degree in computer science from Keio University at 1999. He worked in NTT laboratories from 1999 to 2005, and joined NII in 2006. His current research interests are Internet traffic measurement and analysis, intelligent network control architectures, and the scientific aspects of networks. In 2002, he was a visiting scholar at Boston University.

Patrice Abry received the degree of Professeur-Agrégé de Sciences Physiques, in 1989 at Ecole Normale Supérieure de Cachan and completed a Ph.D. in Physics and Signal Processing, at Ecole Normale Supérieure de Lyon and Université Claude-Bernard Lyon I, in 1994. Since October 95, he is a permanent CNRS researcher, at the laboratoire de Physique of Ecole Normale Supérieure de Lyon. He received the AFCET-MESR-CNRS prize for best Ph.D. in Signal Processing for the years 1993–1994. His current research interests include wavelet-based analysis and modeling of scaling phenomena and related topics (self-similarity, stable processes, multi-fractal, 1/f processes, long-range dependence, local regularity of processes, inifinitely divisible cascades, departures from exact scale invariance).



Kenjiro Cho is Deputy Research Director at Internet Initiative Japan, Inc. He received the B.S. degree in electronic engineering from Kobe University, the M.Eng. degree in computer science from Cornell University, and the Ph.D. degree in media and governance from Keio University. He was with Sony Computer Science Laboratories, Inc. during 1996–2004, and is with IIJ since 2004. He is also an adjunct professor at Japan Advanced Institute of Science and Technology, and a board member of the WIDE project. His current research interests include traffic measurement and management, and operating system support for networking.

Hiroshi Esaki received Ph.D. from University of Tokyo, Japan, in 1998. In 1987, he joined Research and Development Center, Toshiba Corporation. From 1990 to 1991, he has been at Applied Research Laboratory of Bellcore Inc., New Jersey, as a residential researcher. From 1994 to 1996, he has been at Center for Telecommunication Research of Columbia University in New York. From 1998, he has served as a professor at the University of Tokyo, and as a board member of WIDE Project. Currently, he is executive director of IPv6 promotion council, vice president of JPNIC, IPv6 Forum Fellow, board member of WIDE Project and Board of Trustee of Internet Society.