Using Mixed Distribution for Modeling End-to-End Delay Characteristics

Yasuhiro SATO, Shingo ATA, Ikuo OKA
Graduate School of Engineering,
Osaka City University, Osaka, Japan
Email: {sato@n, ata, oka}@info.eng.osaka-cu.ac.jp

Chikato FUJIWARA
Faculty of Modern Management Information,
Osaka Seikei University, Osaka, Japan
Email: fujiwara-c@osaka-seikei.ac.jp

Abstract

An end-to-end round trip time (RTT) is one of the most important communication characteristics for Internet applications. In order to meet the quality of service (QoS) for applications such as streaming services, it is important to understand and predict the delay characteristic accurately. Because characteristics of the packet transmission delay in the Internet vary strongly related to the time-dependent the condition of network. Especially, it is obvious that two or more factors may interfere in observed delay characteristics.

We express these delay characteristics by using the mixture of two or more probability distributions (especially normal distribution), not by single distribution. By investigation the factor which infects the delay distribution, we presume the events that occur in network from the delay characteristic obtained by the observation in the end host.

Keywords: delay characteristics, round trip time (RTT), modeling, mixed distribution, normal distribution

Introduction

- RTT is one of the most important communication characteristics.
  - In TCP, large delay variance leads a failure of appropriate timeout predictions.
  - In Voice Transfer Service (such as Streaming Service), a long delay simply stands for a delay of conversation.
- RTT would be significantly varied when something like congestion and incident occurred.
  - If we could clarify the characteristics of RTT, congestion, incident and the route change may be detected.
- Many researchers have ･analyzed the characteristics of RTT.
  - RTT has a nature of heavy-tailed. (V. Paxson, A. Acharya)
  - RTT is modeled by using single Gamma distribution. (M. Kalman)
- We express the delay characteristics by using the mixture two or more probability distributions.

(2)

1. Introduction

For network applications an end-to-end round trip time (RTT, Round Trip Time) is one of the important communication characteristics. Many research works have been shown that the communication quality of such applications is strongly dependent on the characteristics of the end-to-end delay. There is therefore a possibility that the higher-quality communication can be achieved in the same network environment, by understanding the characteristics of RTTs more accurately.

From the viewpoint of network operators, RTT can also become one of the important metrics. When something like congestion and incident occurred in the network, RTT would be significantly varied. Therefore, if we could clarify the characteristics of RTT, congestion, incident, and any other changes in the route between two end nodes may be detected.

For above backgrounds, several researchers have analyzed the characteristics of RTT so far. In [1],[2], they showed that RTT has time-dependent fluctuations by measuring RTT using probe packets, and also that a distribution of RTT has a nature of heavy-tailed. Additionally, in [3],[4], they modeled RTT using the correlation of RTT value. However, they assumed modeling of RTT in the same segment. When RTT has two or more characteristics by factors which influence network properties (we call as Network Impact Factors (NIFs)). NIFs include e.g., route change, congestion, etc., it is necessary to model individually after distinguishing each characteristic. In [5], they showed that it is possible to model accurately by using gamma distribution in delay characteristics, but an error becomes large when two or more NIFs occurred because they used a single distribution. When we focus on a single NIF to model the delay distribution, RTT values influenced by other NIFs would be considered to error values in this model. As a result, modeling a delay distribution with a single function will have a large error as the increase of NIFs. In this case, modeling with a single distribution function is not suitable. It is desirable to model by mixture of two or more probability distribution, not by single distribution.

We express these delay characteristics by using the mixture of two or more probability distributions (especially normal distribution). Moreover by investigation the factor which infects the delay distribution, we presume the events that occurred in network from the delay characteristics obtained by the observation in the end host.

Measurement Environment

- We measured RTT from Osaka City University to some major WWW servers (called target hosts) distributed in the world.
- The target hosts
  - MIT, FBI, KGB, google, yahoo, etc.･
- We collected 24-hour of RTTs with 1 sec. period.
- We executed traceroute to obtain the route to target hosts.
- The measurements were performed from January to November, 2004.
- The measured RTT is 0.01 msec. unit.

(3)

2. Measurement Environment

We measured RTT from Osaka City University to some major WWW servers (called target hosts) distributed in the world. We selected WWW servers considering geographic location of servers. We also selected two domestic WWW servers in Japan. We also measure one-hop RTTs to our public web server located on the same segment with the measurement host. The measurements were performed from January to November, 2004. For each target host we collected 24-hours of RTTs with 1 sec. period (total 86,400 samples).

During the measurement of RTTs, we also executed traceroute to obtain the route to target hosts. Furthermore, we measured RTTs not only to target hosts but also intermediate routers to target hosts, which are obtained by traceroute, to investigate the variation of distribution characteristics by increasing the number of hops. We should note here that because we use ICMP packets to measure RTTs (i.e., we used modified ping program), we need to consider some restrictions related to ICMP packets. First, some routers limit a rate of processing of ICMP packets. This means that some RTTs may include the time waiting in the routers which limit the rate of sending ICMP packets. Therefore we set the interval of sending probe packets enough to avoid the influence of rate limitation. Second, some routers or hosts do not respond ICMP probe packets due to the security reason. We eliminate these results when we model the distribution.

Methodology of Analysis

- We analyze the delay characteristics by investigating …
  - the time of day variance.
  - probability density function of obtained RTTs.
- We model probability density function of obtained RTT.
  - Modeling function is the mixture of two or more normal distributions.
  - Deciding the parameters of the model function by the least square method.

(4)

3. Methodology of Analysis

3.1 Probability Distribution of RTT

We analyze characteristics of RTT using probability distribution function. Because the value of RTTs are truncated by the granularity of the clock of the measurement host (e.g., 1 msec), measured RTTs are grouped by the unit of clock on the measurement host. After grouping samples by the value of RTT, we count number of samples for each value of RTT. We then sort the value of RTT in ascending order, and label them from 1 to L, where L is the number of different RTT values. We define x_l ($1 \leqq l \leqq L$) as the value of RTT which is labeled by l, and $d(x_l)$ as the frequency (i.e., the number of samples) of the value of RTT x_l.

The probability density distributions (occurrence probability) at label l is given by

$$p_l = \frac{d(x_l)}{N},$$

where N is number of RTT samples.

Mixed distribution

- This is a distribution that combines two or more probability distribution functions.

$$R_{\text{mixed}}(x) = \sum_{i=1}^{M} a_i p_i(x)$$

$a_i (0 \leq a_i \leq 1)$: mixture ratio

$p_i(x)$: probability function of each distribution

M : number of combined distribution (mixture degree)

- The mixture ratio means weight of each distribution.

(5)

3.2 Normal Distribution

We selected normal distribution as candidate to adequately represent delay distributions. The normal distribution is given by

$$e = \sum_{i=1}^{L} \frac{(q_i - R_{\text{model}}(x_i))^2}{q_i^2}.$$

where μ is the mean of measurements and σ is their variance.

It is not comparable with normal distribution because the obtained RTT is discrete value. Therefore, we divide the probability variable of normal distribution into a constant class value, derive the probability value in each class value by quadrature by parts, and treat as a discrete value. Because the RTT value is 0.01 msec. units, the class value is assumed 0.01 msec. units.

3.3 Mixed Distribution

The mixed distribution is distribution that combines two or more probability distribution functions. In this study, we express the delay characteristics by using the mixture of two or more normal distributions. This slide shows that the probability distribution of the mixed distribution function. Theoretically, the sum of a_i should be 1. For instance, when the mean and variance of normal distribution 1 are denoted μ_1 and σ_1, and the mean and variance of normal distribution 2 are denote μ_2 and σ_2, the distribution function of two mixed normal distributions is given by

$$R_{\text{norm2}}(x) = a_1 \frac{1}{\sqrt{2\pi\sigma_1}} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}} + a_2 \frac{1}{\sqrt{2\pi\sigma_2}} e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}}.$$

In this case, it is theoretically satisfied that a_2=1- a_1. However, we treat as a_1 and a_2 independently because our purpose is to fit the measured RTTs to modeling functions as much part of distribution as possible, not entirely.

Decision of Appropriate Model

1. We pick up t pairs of (value of RTT x_i, number of samples for x_i) values from measured data.
2. Set parameters of the model function $f(x)$ as initial values.
3. Obtain t pairs of $(x_i, f(x_i))$ for each x_i.
4. Calculate the sum of squared differences D given by

$$D = \sum_{i=1}^{t} (y_i - f(x_i))^2 .$$

5. Set parameters of $f(x)$ to another value and repeat from Step 3 to Step 5 until the minimum value D is found.
6. Obtain the last parameters of $f(x)$ as appropriate parameters.

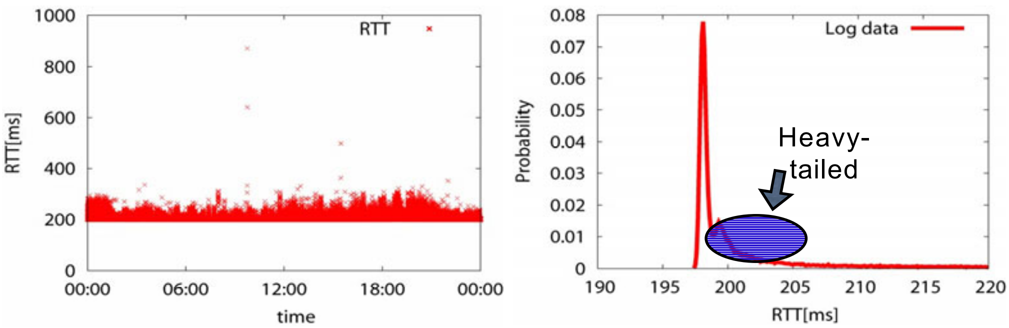
(6)

3.4 Decision of Appropriate Model

The least square method is used for deciding the parameter of the model. In this method, we calculate the sum of squa･red differences between the sampled values and the values obtained by the model distribution function. To obtain the appropriate set of parameters, iterate to calculate the sum of squared differences by changing the parameters of distribution function. We use parameters having the minimum sum of squared differences as an appropriate parameters. This slide shows the procedure.

We use the least square method to determine two types of parameters: one is the mixture ratio a_i, and the other is a set of parameters of normal distribution functions. We iterate above procedure by increasing the mixture degree one by one, and stop the procedure when we finally obtain the minimum value of D. The detailed process is described in Subsection 4.2.

Overview of Measurement Results



- There are some points having large values of RTTs.
  - It is because the network congestion occurred suddenly due to the significant increase of traffic.

(7)

4. Analytic Results

4.1 Overview of Measurement Results

The left figure shows time-dependent variations of RTTs and the right one shows the probability distribution of google. The horizontal axis of both are time measured, and vertical ones show the value of RTTs.

From the left figure, we can observe that RTTs have a time-of-day variance; in busy hours (e.g., 12pm), many people use the Internet simultaneously, and the congestion of the network becomes increased. On the other hand, during silence hours (e.g., 0am), values of RTTs are small because most of hosts do not generate traffic at that time. Moreover, there are some points having large values of RTTs (called *spikes*). It is because the network congestion occurred suddenly due to the significant increase of traffic. Unlike above type of variance, spikes may occur regardless of time-of-day.

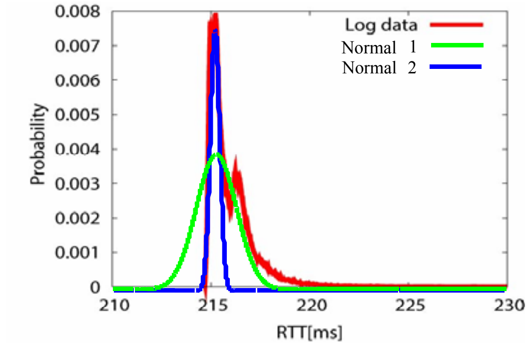
Next, as shown in the right figure, we can observe that there are two or more peaks in the distribution of RTTs. One of the reasons can be considered that a route change is occurred during the measurement. If we use a single distribution function to model RTTs, a large error is appeared at the second peak of the distribution. Moreover, these distributions have a heavy-tailed property (i.e., a small number of RTTs have very large values) it is also reported in [2].

Modeling by Normal Distribution

- The result of normal distribution is quite far from the measured RTTs.

↓

- There are two peaks while the normal distribution has a single peak.
- The distribution of RTTs has more heavy-tailed property.



It cannot model both of two regions.

(8)

4.2 Parameters Estimation

In modeling by normal distribution, parameters which should be estimated are the mean and the variance of RTTs. In case of mixed distribution, on the other hand, mixture ratios are also required in addition to parameters of each sub-distribution.

4.3 Modeling by Normal Distribution

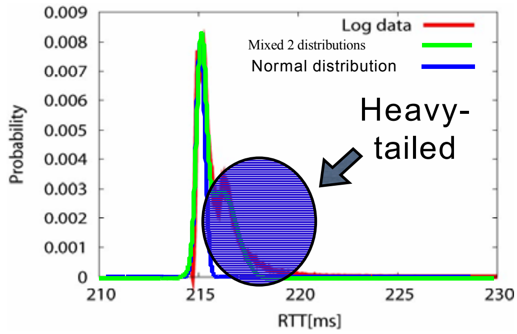
First, we show the result of modeling by single normal distribution in this slide. In this figure we compare the distribution of measured RTTs and the normal distribution function with parameters calculated by the process shown in Subsection 3.4. The measured RTTs are collected from MIT on June 26, 2004. “Normal1” is a result of modeling from data the entire distribution. “Normal2” is a result of modeling of the match to the first peak. We can observe from this figure that the result of normal distribution is quite far from the measured RTTs. We consider following two reasons for this result. First, there are two peaks in MIT data while the normal distribution has a single peak. Second, the distribution of RTTs has more heavy-tailed property, i.e., the density is suddenly increasing at around 215 msec, and gradually decrease until 220 msec. Moreover, probabilities around 215 msec are significantly high (up to 0.008) rather than other regions. Even if the normal distribution can fit either the region at peak (around 215 msec), or the region at tail part (from 216 to 220 msec), it cannot model both of two regions.

Modeling by Mixed Distribution I

- The second peak is considered as heavy-tailed part in modeling by a single distribution.

Modeling by the mixed distribution.

It is possible to model a part of heavy-tailed by using the mixed distribution.



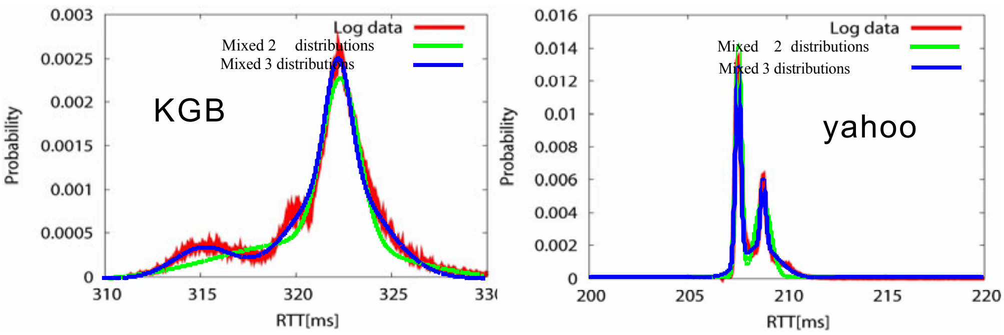
Probability distribution (MIT-1)

(9)

4.4 Mixed Distribution

As described before, expressing the delay distribution from a current result by single distribution is low accuracy, and insufficient to model. Hence, we model the delay distribution by using the mixed distribution. This slide shows that we could express the delay characteristics by using the mixed distribution. In the figure, the second peak is considered as heavy-tailed part in modeling by a single distribution. However, using the mixed distribution, the heavy-tailed part is modeled more accurately. Thus, it is possible to model a part of heavy-tailed by using the mixed distribution.

Modeling by Mixed Distribution II



When the mixture degree is two, modeling is not enough.

By increasing the mixture degree, we could model more accurately.

(10)

These figures compare the probability distributions of the mixed distribution with the measurement data on each site. From these figures, we can model the distribution of RTTs quite well by using mixed 3 distributions. We use a sum of squared relative differences (we refer as *relative error* in this study) between sampled value and model function as the metric of appropriateness of modeling, which is given by

$$e = \sum_{i=1}^{L} \frac{(q_i - R_{\text{model}}(x_i))^2}{q_i^2}.$$

Using mixed distribution can decrease the relative error greatly compared with modeling by normal distribution. Another notable observation is that generally the relative error can be decreased by increasing the mixture degree. Our results imply that there is a kind of optimal mixture degree for each measured data sets. As noted in Section 1, if we focus on a single network impact factor on modeling, values affected by other factors may be considered as error. That is, we can consider that the mixture degree is strongly relative to the number of NIFs, especially it is equal to the number of NIFs. Of course the number of NIFs is finite, the value of mixture degree is also finite. That is why we consider that the mixture degree has an optimal value.

Totally, 38% of all data sets can be modeled by single distribution. However, 62% of data sets cannot be modeled. By using mixed distribution with two normal distributions, 55% of total (i.e., 89% of rest) datasets can be modeled. Furthermore, 6% of total (i.e., 10% of rest) can be modeled by mixed three distributions.

Inferring Network Impact Factors

- It is thought that variance of the delay characteristics only by the change in the route hardly change.
  - The change in the route was not admitted from the result of traceroute.

	a_1		a_2	
	μ_1	σ_1	μ_2	σ_2
UCLA	0.34		0.26	
	287.24	0.010	288.53	0.012
yahoo-1	0.40		0.13	
	207.53	0.018	208.80	0.017
yahoo-2	0.32		0.24	
	198.39	0.016	199.68	0.018

We consider that it depends on the load-balancing in the lower layer.

(11)

5. Inferring Network Impact Factors

There is a possibility to be able to presume the factors that occurred in network, by focusing on mixture ratio, mixture degree and the parameters.We found that in yahoo-1 and yahoo-2 in Table, σ_1 and σ_2 are very close values. These sites can be presumed to generate the delay by the change in the route because it is thought that variance of the delay characteristic only by the change in the route hardly changes. However, the change in the route was not admitted from the result of traceroute investigated in parallel with this measurement data. Therefore, we consider that it depends on the load-balancing in the lower layer that cannot be measured by traceroute, not by the change of route on IP. On the other hand, a big difference is seen in each variance on KGB. In such case, we consider that it is because another delay factor in addition to the route change described on is included.

Next, we focus on the mixture ratio. Because the sum of the area of the probability distribution is 1, the sum of the mixture ratio is 1, too. However, in yahoo-2, the sum of the mixture ratio is 0.74, and modeling is not accurate though the error margin is small. In contrast, the sum of the mixture ratio of MIT is 0.96, which is modeled more accurate. This shows the part of the remainder is not modeled completely by this analysis. Hence, we need modeling that increases the mixture degree and uses excluding normal distribution.

Conclusions and Future works

- We have proposed a new statistical modeling method to analyze RTT distributions by using mixed distributions.
- We have found that the tail part of RTT distribution can be modeled by the mixture of non-heavy-tailed distributions.
- Based on numerical results, we have discussed how to infer network impact factors from RTT distributions.

- We verify whether accurate presumption can be done by increasing the mixture degree, and by using a distribution other than normal distribution.
- We need to consider that a more detailed delay factor can be presumed by clarifying the relation how the factor to change the delay influences the parameters of the mixed distribution.

(12)

6. Conclusions

・ We have proposed a new statistical modeling method to analyze RTT distributions by using mixed distributions. Our numerical results have shown that most of delay distributions are not suitable to model by a single distribution, and they are preferable to be modeled by mixed distribution. Moreover, we have found that the tail part of RTT distribution, which is considered to have a heavy-tailed nature so far, can be modeled by the mixture of non-heavy-tailed distributions (i.e., normal distribution). Based on numerical results, we have discussed how to infer network impact factors from RTT distributions.

For future research topic, it is necessary to verify whether accurate presumption can be done by increasing the mixture degree, and by using a distribution other than normal distribution.

We also need to consider that a more detailed delay factor can be presumed by clarifying the relation how the factor to change the delay influences the parameters of the mixed distribution.

References

[1] V. Paxson, "End-to-End Internet Packet Dynamics," in Proceedings of ACM SIGCOMM '97, pp. 139–154, September 1997.

[2] A. Acharya and J. Saltz, "A Study of Internet Round-Trip Delay," Tech. Rep. CS-TR-3736, January 1996.

[3] J. Bolot, "Characterizing end-to-end packet delay and loss in the Internet," Journal of High Speed Networks, vol. 2, pp. 289–298, September 1993.

[4] Z. Wang, A. Zeitoun, and S. Jamin, "Challenges and lessons learned in measuring path RTT for proximity-based applications," in Proceedings of Passive and Active Measurement Workshop, pp. 151–157, April 2002.

[5] M. Kalman and B. Girod, "Modeling the delays of successively-transmitted Internet packets," in Proceedings of IEEE International Conference on Multimedia and Expo, pp. 2015–2018, June 2004.