PAPER

Detection of TCP Performance Degradation Using Link Utilization Statistics*

Keisuke ISHIBASHI[†a)], **Ryoichi KAWAHARA**[††], **Takuya ASAKA**[†††], **Masaki AIDA**[††††], **Satoshi ONO**[†††††], *and* **Shoichiro ASANO**[††††††], *Members*

SUMMARY In this paper, we propose a method of detecting TCP performance degradation using only bottleneck-link utilization statistics: mean and variance. The variance of link utilization normally increases as the mean link-utilization increases. However, because link-utilization has a maximum of 100%, as the mean approaches 100%, the possible range of fluctuation becomes narrow and the variance decreases to zero. In this paper, using the M/G/R processor sharing model, we relate this phenomenon to the behavior of flows. We also show that by using this relationship, we can detect TCP performance degradation using the mean and variance of link utilization. In particular, this method enables a network operator to determine whether or not the degradation originates from the congestion of his/her own network. Because our method requires us to measure only link utilization, the cost of performance management can be greatly decreased compared with the conventional method, which requires dedicated functions for directly measuring the TCP performance.

key words: *TCP, performance, link utilization, measurement*

1. Introduction

With the rapid increases in traffic of new applications such as P2P file-sharing in IP networks, congestion is a significant issue in those networks. Therefore, IP network operators are eager to determine the degradation of the QoS of user traffic (e.g., TCP throughput). They especially want to know whether or not the degradation occurs due to congestion in their own networks.

Currently, most operators simply correlate the QoS with mean link utilization and check that mean link utilization does not exceed a pre-determined threshold (e.g., 50%) [1]. In such methods, however, the threshold is usually determined empirically and is not based on the actual QoS of user traffic. Consequently, the QoS is not guaranteed. Although only measuring or estimating QoS is not sufficient to guarantee QoS, it is one of the requirements for guaranteeing the QoS without some QoS control mechanisms, which are not widely deployed in the current Internet. Therefore, QoS measurement or estimation should be used as one of the functions required to guarantee the QoS.

On the other hand, service level management based on actual (active or passive) measurement can directly monitor user QoS. For example, with netflow [2], we can passively obtain information about the duration and transfer size of each flow, and we can determine the performance of the flow. Moreover, TCP throughput can be actively measured by transferring a certain amount of test traffic [3] or calculated using the packet delay and loss-rate, which can also be measured actively [4]. However, such performance measurement will be relatively costly because it requires dedicated functions for the measurement. Furthermore, with those end-to-end measurements, the network operators cannot determine whether or not the performance degradation originated from congestion in their own networks.

Recently, the processor sharing model (M/G/R/PS model) has been applied to calculate the performance of TCP flows in a link that aggregates many access links (Fig. 1) [5]–[8]. In the M/G/R/PS model, TCP flows are assumed to fully and exclusively utilize their own access link bandwidth when the aggregate link is not congested. And when the aggregate link is congested, they are assumed to fairly share the bandwidth of the aggregate link. (A detailed explanation of the model is given in Sect. 2.1.) The model requires only the mean link utilization and bandwidth ratio of the access link and aggregate link to find degradation of TCP performance. In addition, it directly relates the aggregate-link utilization and TCP performance, so we can determine whether or not the degradation originated from congestion in the aggregate link. As described later, how-

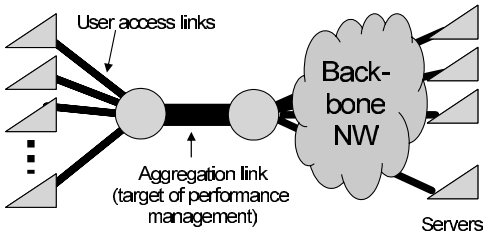

Fig. 1 Network model treated in this paper.

Manuscript received January 14, 2005.
Manuscript revised June 26, 2005.

[†]The author is with the NTT Information Sharing Platform Laboratories, NTT Corporation, Musashino-shi, 180-8585 Japan.
[††]The author is with the NTT Service Integration Laboratories, NTT Corporation, Musashino-shi, 180-8585 Japan.
[†††]The author is with the Graduate School of Informatics, Kyoto University, Kyoto-shi, 615-8510 Japan.
[††††]The author is with the Faculty of System Design, Tokyo Metropolitan University, Hino-shi, 191-0065 Japan.
[†††††]The author is with the Center for Continuing Professional Development, Kogakuin University, Tokyo, 163-8677 Japan.
[††††††]The author is with National Institute of Informatics, Tokyo, 101-8430 Japan.

*Part of this paper was presented at IEEE GLOBECOM2004, in November 2004.

a) E-mail: ishibashi@m.ieice.org

DOI: 10.1093/ietcom/e89–b.1.47

ever, this method has a drawback in its accuracy because it assumes that a flow fully utilizes its access link bandwidth; i.e., it ignores the effect of TCP slow-start or large bandwidth-delay product [9], which would prevent a TCP flow fully utilizing its access link. In that case, actual TCP throughput, which we refer to as "actual maximum TCP throughput" and explain in Sect. 2.2, may be smaller than the access link bandwidth.

We previously extended the model by using the mean number of flows in the link to increase the accuracy [10]. By using the mean number of flows, we can estimate the actual maximum TCP throughput and take into account the effect of slow-start or large bandwidth-delay product in estimating the TCP performance. However, while some routers have a function for presenting information about flows passing through them [2], the measurement of the number of flows is relatively difficult. First, it requires identifying and reconstructing a flow from packets passing through the router, which consumes the processing resources of the routers. Second, it requires holding a table of flow information, so the required memory for a highly aggregated router may become huge and this memory must be sufficiently fast, which will make it expensive. On the other hand, measuring the link utilization only requires counting the size of packets passing through the link, which is an easy task.

In this paper, based on the results of the M/G/R/PS model and its extension [10], we propose using link utilization statistics instead of the number of flows to detect TCP performance degradation. When the link is lightly loaded, the variance of link utilization increases with increasing mean link utilization. This is natural because if flows composing the link traffic are independent of each other, then the variance increases linearly with the number of flows, as Morris showed for Web traffic [11]†.

However, because link utilization cannot exceed 100%, when mean link utilization approaches 100%, the possible range of fluctuation becomes narrow and the variance of link utilization decreases. In that case, the TCP performance may also be degraded because when instantaneous link utilization is 100%, the bandwidth available for a TCP flow at the time should smaller than the flow's original rate. Thus, the decrease in the variance of link utilization is expected to be related to the decrease in TCP performance. In this paper, by assuming that the behavior of TCP flows can be modeled as the M/G/R processor sharing (PS) model, we give a theoretical basis for this relationship. We also show that by using this relationship, we can detect TCP performance degradation using only the mean and variance of link utilization. These values can be obtained from the values in management information bases (MIBs), which are implemented in almost all routers and can be easily obtained through the SNMP protocol. Therefore, the cost of performance management can be significantly reduced.

The rest of the paper is organized as follows. First, in Sect. 2, we show the target network environment of our study and briefly summarize related studies on detecting the performance degradation of networks of this kind, including our previous work. Then, in Sect. 3, we propose a method of detecting TCP performance degradation using the variance and mean of link utilization. For this, we show the relationship between the variances of link utilization and the number of flows. Then we show the relationship between the variance of the flows and the degree of TCP performance degradation. Section 4 presents an evaluation of our method through simulation.

2. Related Work

In this paper, as in [10], we focus on an aggregate link (C [bps]) on which traffic from many user access links (r [bps]) is aggregated (Fig. 1). We also focus on the performance of TCP traffic because current Internet traffic mostly consists of TCP flows [12]. We define the degree of TCP performance degradation as the mean transfer time of TCP flows divided by the mean transfer time when the aggregate link is lightly loaded.

2.1 Application of M/G/R/PS Model to Estimate TCP Performance Degradation

Recently, there have been many studies on applying the M/G/R/PS model to TCP performance evaluation [5]–[8]. The M/G/R/PS model is a generalization of the single-server processor sharing (PS) queuing model. In the single-server PS queuing model, all customers in a system receive services fairly, while in regular queuing models such as FIFO, only customers at the server can receive services, and customers in the queue cannot receive any services [14]. By considering a TCP flow as a customer and considering the bandwidth of the link as processing capacity of the server, and assuming that TCP flows fairly and fully utilize the link bandwidth, we can apply the processor sharing model to model the flow behavior in a congested link. In the M/G/R/PS model, when the number of servers is increased from one to R, a customer can utilize at most only one server even if other servers are not being used. Thus, when the number of customer is smaller than R, the capacity of all R servers is not fully utilized, whereas when the number of customer is greater than R, the capacity of servers is fairly shared. This model fits to the case in Fig. 1, because in that case, when the aggregate link is not congested, TCP flows utilize their own access link bandwidth, however bandwidth of aggregate link is not fully utilized. And when the aggregate link is congested, they can be assumed to fairly share the aggregate link bandwidth. Using the model, we can then obtain the degree of TCP performance degradation as a function of mean link utilization (ρ) and bandwidth ratio ($R = C/r$) expressed as follows [5], [8]:

$$D(R,\rho) = 1 + \frac{\mathbf{C}(\lfloor R \rfloor, \rho)}{(1-\rho)R}\Big(1 - (1-\rho)(R - \lfloor R \rfloor)\Big), \qquad (1)$$

†When multiple TCP flows come from a client which are not independent and synchronized, we can regard them as one aggregate TCP flow (See Appendix A in [10]).

where $\mathbf{C}(R,\rho)$ is Erlang's C equation, which can be written as [13]:

$$\mathbf{C}(R,\rho) := \frac{\left(\frac{(\rho R)^R}{R!}\right)\left(\frac{1}{1-\rho}\right)}{\sum_{k=0}^{R-1}\frac{(\rho R)^k}{k!} + \left(\frac{(\rho R)^R}{R!}\right)\left(\frac{1}{1-\rho}\right)}. \quad (2)$$

In this model, while flows are assumed to fully utilize the access link when the aggregate link is not congested, this is rarely true because the actual TCP throughput may be limited due to the slow-start effect or a large bandwidth-delay product other than the bandwidth of access link or the congestion in the aggregate link. While Eq. (1) assumes that R is the number of flows to consume the bandwidth of the aggregate link, if TCP throughput is smaller than their access bandwidth even if the aggregate link is not congested, then the number of the flows required to consume the bandwdith should be larger than R. Consequently, this model may not detect performance degradation accurately.

The method proposed in [8] supports the effect of slow-start. However, to calculate the throughput considering the slow-start effect, the method uses RTT and maximum window size as well as the link utilization and link capacity. Those parameters are not easy to measure for each flow. On the other hand, as described later, our method only uses the link utilization statistics and link capacities, which are easy to measure.

2.2 Detection of Performance Degradation Based on Flow Information

In [10], to increase the accuracy of estimation of TCP performance degradation, we extended the model by replacing the access link bandwidth r in (1) with the actual maximum throughput of a flow $r'(\leq r)$, which reflects the effect of throughput degradation factors other than aggregate-link congestion. More precisely, the actual maximum throughput r' is defined as the average throughput of flows when the aggregate link is not congested. Thus, r' can also be calculated as the bandwidth usage of the aggregate link divided by the number of flows in the link when it is not congested.

Therefore, bandwidth ratio R should be also replaced by *virtual* bandwidth ratio $R_V := C/r'$. Let the link utilization at this time be ρ. Then, the bandwidth usage of the link is ρC, and the mean number of flows is $\rho C/r'$, or ρR_V. Thus, R_V can be obtained by dividing the average number of flows in the the link by the mean utilization of the link (ρ) when the link is not congested. In [10], we also verified through various simulations that we can accurately calculate the degree of performance degradation when the link is congested by using R_V in (1).

However, as was noted in Sect. 1, the number of flows is harder to measure than link utilization. In the next section, we show that we can detect the TCP performance degradation without directly estimating R_V by using the link utilization statistics.

3. Proposed Method

In this section, first we give the relationship between the statistics of flow and link utilization under the assumption that the behavior of flows is modeled by the M/G/R/PS model, and then we describe how we can detect the performance degradation using the variance and mean of link utilization.

Here, we define several terms (including some that were defined earlier).

- R_V: The capacity of the aggregate link divided by the actual maximum throughput of a flow. R_V can also be regarded as the average number of flows required to consume all the bandwidth on the aggregate link.
- $F(t)$: the number of simultaneous flows in the aggregate link at time t
- $S(t)$: $\min(F(t), R_V)$. $S(t)$ is introduced to relate the number of flows to the link utilization. In the following, we assume that the link utilization is proportional to the number of flows. However, the number of flows can grow infinity while the link utilization has a maximum value of one. If we limit R_V, then $S(t)$ actually becomes proportional to link utilization.
- $U(t)$: a binary function such that $U(t) = 1$ if the aggregating router transmits a packet at time t, and $U(t) = 0$ otherwise (Fig. 2) Note that $\mathrm{E}[U(t)] = \rho$.
- $U_\tau(t)$, $S_\tau(t)$: averaged time-series of $U(t)$ and $S(t)$ over the time scale τ. Here, $U_\tau(t) = \int_0^\tau U(t+\tau)\,d\tau/\tau$ and this is the link utilization measured with time-interval τ.
- $V_U(\tau)$ and $V_S(\tau)$: the variances of $U_\tau(t)$ and $S_\tau(t)$, respectively. $V_U(\tau)$ and $V_S(\tau)$ are decreasing functions of τ because averaging over τ makes the fluctuation in the interval disappear. These are also called Variance Time Plots (VTPs), which are used to calculate the degree of traffic self-similarity.
- τ_0 :=Packet size divided by the bandwidth of the aggregate link.

In the rest of this section, we show the relationship between the variance of utilization and the TCP performance degradation. First, we write the variance of link utilization with measurement interval τ, $V_U(\tau)$, as

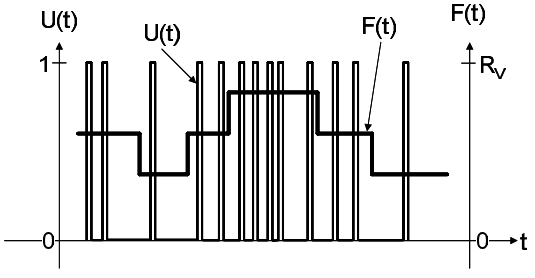

Fig. 2 Sample plot of $F(t)$ and $U(t)$.

$$V_U(\tau) = \frac{V_U(\tau)}{V_S(\tau)/{R_V}^2} \frac{V_S(\tau)}{\mathrm{Var}[S(t)]} \frac{\mathrm{Var}[S(t)]}{{R_V}^2}. \quad (3)$$

Then, we evaluate each factor on the right hand side of (3) to find the relationship between $V_U(\tau)$ and $D(R_V, \rho)$, the degree of TCP performance degradation.

3.1 Relationship between Link Utilization Variance and the Variance of the Number of Flows

Here, we show the relationship between the variance of the number of flows and the variance of link utilization.

First, we assume that the probability of a packet being transmitted at time t (i.e., $\Pr[U(t) = 1]$) is proportional to $S(t)$, which is the number of active flows at time t limited by R_V. This probability, i.e., packet intensity at time t, may depend on many factors, but the most significant factor is naturally assumed to be the number of active flows. We also assume that packets in a flow are arriving at random[†]. Then, we can obtain the relationship between the variance of link utilization and the variance of the number of flows measured over the k-packet time-scale $k\tau_0$ ($k \geq 1$) as follows (the proof is given in Appendix A):

$$V_U(k\tau_0) \simeq \frac{V_S(k\tau_0)}{R_V^2} + \frac{1}{k}\left(\frac{2\rho}{3}(1-\rho) - \frac{\mathrm{Var}[S(0)]}{R_V^2}\right). \quad (4)$$

The second term on the right hand side of (4) decreases as k increases in inverse proportionality. The first term also decreases as k increases, but the rate of decrease is smaller and is determined by the time scale of flow activity, as shown in Sect. 4. Therefore, for τ in the range such that the second term of (4) is negligible compared with the first term, we have

$$\frac{V_U(\tau)}{V_S(\tau)/R_V^2} \simeq 1. \quad (5)$$

3.2 Relationship between the Variance of the Number of Flows and TCP Performance Degradation

As noted in the definition, $V_S(\tau)$ is a decreasing function of τ because averaging over τ makes the fluctuation in the interval disappear. This decrease is determined by the time scale of the fluctuation in the number of flows. When τ is not so large compared with the flow time-scale, we can consider $V_S(\tau)/\mathrm{Var}[S(t)]$ to be almost constant as ρ changes, as shown in Sect. 4.

3.3 Variance of Number of Flows

Now, we evaluate $\mathrm{Var}[S(t)]$. The behavior of the number of flows in the M/G/R/PS model is the same as in the M/M/R model [14]. Thus, when congestion at the aggregate link rarely occurs (i.e., $S(t) \simeq F(t)$), the number of flows follows a Poisson distribution [14], where the variance and mean are the same. Then, we have the following equations: $\mathrm{Var}[S(t)] \simeq \mathrm{Var}[F(t)] \simeq \mathrm{E}[F(t)] \simeq \rho R_V$, where the variance of $S(t)$ increases linearly as the link utilization increases. However, when congestion frequently occurs and $\Pr[F(t) > R_V] > 0$, the variance of $S(t)$ is not the same as ρR_V. In particular, when the aggregate link is heavily congested and the number of flows is almost always larger than R_V (i.e., $\Pr[F(t) > R_V] \simeq 1$), then $S(t)$ remains constant with value R_V and its variance is close to zero. Here, we can derive the above behavior of the variance for $S(t)$ as follows (the proof is given in Appendix B):

$$\mathrm{Var}[S(t)] = \rho R_V(1 - \mathbf{C}(R_V, \rho)). \quad (6)$$

Thus, the decrease in $\mathrm{Var}[S(t)]$ as the link utilization approaches one can be expressed with Erlang's C equation.

3.4 Variance of Link Utilization Averaged over τ

By assuming that $V_S(\tau)/\mathrm{Var}[S(t)]$ is almost constant as ρ changes, as stated in subSect. 3.2, and by substituting (5) and (6) into (3), for the range of τ such that the second term of (4) is negligible compared with the first term, we have

$$V_U(\tau) \simeq \mathrm{const} \times \frac{\rho}{R_V}(1 - \mathbf{C}(R_V, \rho)). \quad (7)$$

Thus, as described in Sect. 1, the variance of link utilization (measured over time-scale τ) increases linearly with increasing mean link utilization ρ when the aggregate link is lightly loaded, but as the link utilization approaches one, its variance decreases to zero, and this decrease is expressed with Erlang's C equation.

Note that the degree of TCP performance degradation in the M/G/R/PS model is also expressed with Erlang's C equation as in (1). In the next subsection, we show that we can detect the performance degradation using the above correspondence between (1) and (7).

3.5 Detection of TCP Performance Degradation

Let V_0 be

$$V_0 := \lim_{\rho \to 0} \frac{V_U(\tau)}{\rho}. \quad (8)$$

When $\rho \to 0$, congestion hardly occurs and $\mathbf{C}(R_V, \rho) \to 0$[††]. Thus, from (7), we have $V_0 = \mathrm{const.}/R_V$ and

$$\mathbf{C}(R_V, \rho) = 1 - \frac{V_U(\tau)/\rho}{V_0}. \quad (9)$$

Comparing (9) and (1), we find that

$$I(\rho) := \frac{1 - \frac{V_U(\tau)/\rho}{V_0}}{1 - \rho} \quad (10)$$

[†]In fact, packets do not arrive at random. In particular, when the window is smaller than the delay-bandwidth product, packets arrive in bursts with the RTT interval as a window. Thus, we can expect there to be some effects of burst arrival. We consider this case in Sect. 4.2.

[††]For actual detections, we can also estimate V_0 by taking the sample mean of $V_U(\tau)/\rho$ when the link is lightly congested and $\mathbf{C}(R_V, \rho)$ can be considered as 0.

is proportional to $D(R_V, \rho) - 1$. Therefore, when performance at the flow level starts to fall, $I(\rho)$, which can be obtained through measurement of link utilization alone, also starts to increase.

Unlike our flow-measurement based method [10], this method does not enable us to directly calculate the degree of TCP performance degradation ($D(R_V, \rho)$) itself. However, to simply detect whether or not performance is being degraded, it is sufficient to check for increases in $I(\rho)$ as long as the following necessary condition holds: the ratio $V_S(\tau)/\mathrm{Var}[S(t)]$ must be constant as link utilization increases. In that case, we can manage the bandwidth by using a pre-determined threshold for $I(\rho)$, where as long as $I(\rho)$ does not exceed the threshold, the performance at the flow-level is not degraded.

To enable this the method to be applied in actual operations, the threshold value for $I(\rho)$ should be determined. $I(\rho)$ is proportional to the degradation factor of TCP performance, but the proportionality cannot be calculated. Thus, it is difficult to directly derive the threshold from a tolerable degradation level. A possible method of determining the threshold is monitoring the behavior (deviation) of $I(\rho)$ when link utilization is low and performance cannot be degraded (*e.g.,* $\rho < 0.5$), and setting a threshold as the value $I(\rho)$ rarely exceeds, such as three times the standard deviation.

In summary, the method of estimating the degradation of TCP performance proposed in this paper requires only the mean and variance of the utilization of the aggregate link, whereas the method proposed in [10] uses the mean number of active TCP flows and the mean utilization of the aggregate link. The statistics of link utilization is easier to measure than mean number of flows.

4. Simulation Evaluation

We evaluated our proposed method of detecting TCP performance degradation using the *ns-2* simulator [15]. First, we evaluated two cases where the bandwidth-delay product (BDP) is larger or smaller than the window-size to see when the assumptions are reasonable.

Unless specifically mentioned, the simulation conditions were as follows: Bandwidth of aggregate link C = 22.5 [Mbps]; bandwidth of access links r = 1.5 [Mbps], flow arrival was a Poisson process; size of files carried by the flows followed an exponential distribution with a mean of 1 [MB]; advertised window size: 64 [KB], packet size: 1 [KB]; (fixed); round-trip propagation delay: 20 or 1000 [ms]; and simulation length: 3600 [s] ×5 times (we used the mean of the results in the evaluation).

4.1 Small RTT

First, to evaluate (4), which gives the relationship between the variance of link utilization and that of the number of flows, we plotted $V_U(\tau)$, $V_F(\tau)/R_V^2$, and the right side of (4) (Fig. 3) when the link utilization was 0.6[†]. When τ was larger than 10 [ms], the rate of decrease of the variance of link utilization was small and when τ was larger than 100 [ms], the second term of (4) was negligible compared with the first term, and $V_U(\tau) \simeq V_S(\tau)/R_V^2$. Abe [16] also found a small decrease in $V_U(\tau)$ for $\tau < 100$ [ms] in actual measurements in a back-bone IP network.

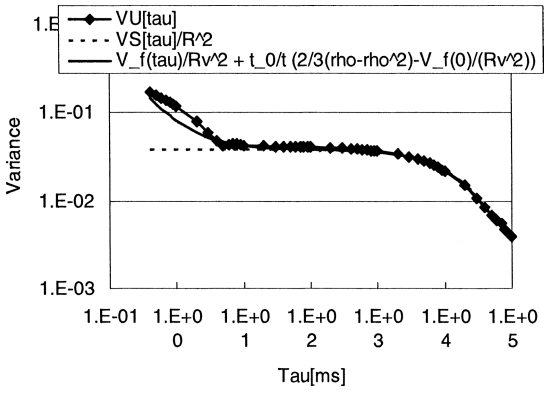

Fig. 3 Comparison of $V_U(t)$ and $V_F(t)$ (RTT= 20 [ms]).

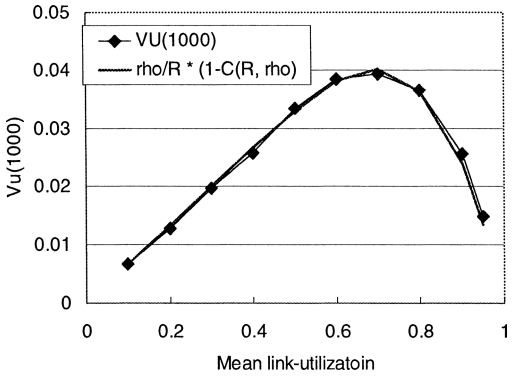

Fig. 4 Variance of link utilization (RTT= 20 [ms]).

Then, to evaluate (7), we compared the variance of link utilization with the measurement interval τ = 1000 [ms] and $\frac{\rho}{R_V}(1 - \mathbf{C}(R_V, \rho))$ (Fig. 4). In this case, since $V_S(\tau)/\mathrm{Var}[S(t)] \simeq 1$ and $V_U(\tau) \simeq V_S(\tau)/R_V^2$, the calculated value agreed with the actual variance. The variance increased linearly with ρ until $\rho = 0.6$. When ρ exceeded 0.6, the variance started to decrease.

We also verified the assumption that $V_S(\tau)/\mathrm{Var}[S(t)]$ is almost constant as mentioned in Sect. 3.2. To evaluate which values of τ lead to $V_S(\tau)/\mathrm{Var}[S(t)]$, which can be considered almost constant, we plotted the ratio $V_S(\tau)/\mathrm{Var}[S(t)]$ for various values of τ (Fig. 5). The ratio was almost constant as ρ varied. Thus, for any value for measurement interval τ in Fig. 5, our method should detect performance degradation at the flow level.

Next, we evaluated how TCP performance degradation is detected using $I(\rho)$. For comparison, as a conventional method to estimate TCP performance using only the link utilization, we calculated the degree of TCP performance

[†]At link utilization of 0.6, $\mathbf{C}(R_V, \rho) \simeq 0$ and $S(t) \simeq F(t)$. Thus, we used $V_F(\tau)$ instead of $V_S(\tau)$ in the evaluation.

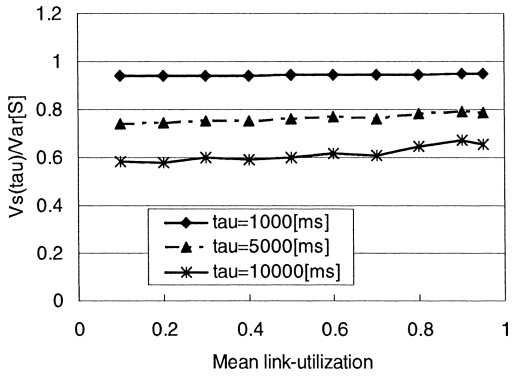

Fig. 5 Ratio of $V_S(\tau)$ to Var$[S(t)]$.

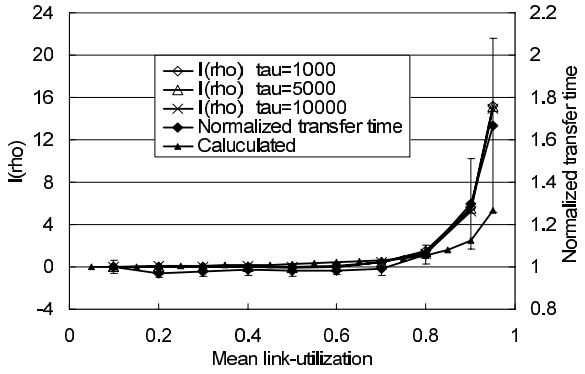

Fig. 6 Degree of performance degradation (RTT= 20 [ms]).

degradation using TCP throughput modeling [4]. Specifically, we first calculated the mean queueing delay for packets using the M/M/1 queueing model whose parameter is mean link utilization. Then we calculated the degree of TCP performance degradation using TCP throughput modeling using the packet delay, which is the sum of the propagation delay and the queueing delay [4].

Figure 6 shows the results. Here "normalized transfer time" indicates the degree of TCP performance degradation obtained from the results of simulations, and "calculated" indicates the estimate obtained using the TCP throughput modeling described above. For $I(\rho)$, we plotted the results measured with τ = 1000, 5000, and 10000,

The $I(\rho)$ values calculated with different measurement intervals agreed. Furthermore, the actual degree of TCP performance degradation and $I(\rho)$ started to increase when the mean link utilization exceeded 0.7, while performance degradation calculated using TCP throughput modeling underestimated the actual degree of degradation. Note that the two quantities are plotted on different scales and that their ratio is nearly R_V. Thus, $I(\rho)$ is not a direct estimator of the degree of TCP performance degradation. However, because both quantities started to increase at almost the same value of ρ, it may be possible to maintain TCP performance by simply observing $I(\rho)$.

We also evaluated a heterogeneous condition on access links. We changed the bandwidth of access links from

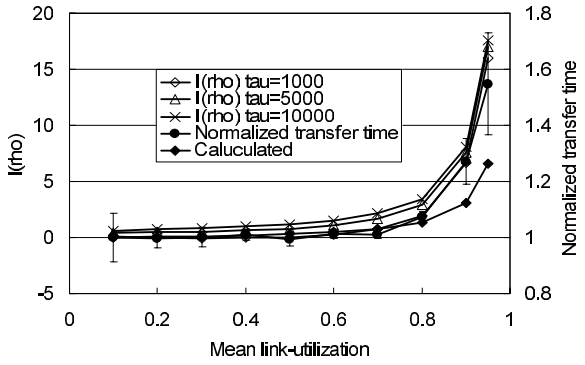

Fig. 7 Degree of performance degradation (heterogeneous bandwidth of access links).

1.5 Mbps to a mixture of 0.5 Mbps and 4.5 Mbps. We set equal transfer volumes through these two types of links. Figure 7 shows the results. Even in this heterogeneous environment, our method could detect the TCP performance degradation accurately. In addition, the results show that confidence intervals were smaller than those in the homogenous conditions. One of the possible reasons is that in our heterogeneous environment, the mean transfer times of flows in small capacity links were large, while the performance degradation and its deviation in those links were small. This is because in a heterogeneous environment, the bandwidth of the aggregate link is shared in a max-min fairness manner [17].

As for the heterogeneous conditions other than the access link bandwidth, RTT values may vary among TCP connections. Specifically, TCP throughput is roughly determined by the minimum value of the maximum window size divided by RTT or by the least bandwidth in the end-to-end path. In the network supposed in this paper (Fig. 1), the bottleneck is access link bandwidth. Therefore, we think that the heterogeneous RTT condition can be essentially reduced to the heterogeneous access bandwidth case.

Because flow-size distribution in the Internet has been reported as heavy tailed [18], we also evaluated the case for Pareto distributed flow size. Shape parameter for the distribution was set to 1.5, which is the default value of the *ns2* simulator. Other parameters were the same as those used in this subsection. Figure 8 shows the results. The evaluation results showed that our method can be applicable for the Pareto distributed file size.

4.2 Large RTT

Next, we evaluated the case where the window size was smaller than the bandwidth-delay product by changing RTT to 1000 [ms]. Other conditions were the same as in Sect. 4.1. Under these conditions, TCP flows could not fully utilize the access-link bandwidth and the packets that made up flows arrived as bursts with the RTT interval.

Figure 9 shows the variance for various measurement time scales. Unlike the variance in Fig. 3, $V_U(\tau)$ did not quickly converge to $V_S(\tau)/R_V^2$, and a significant discrepancy

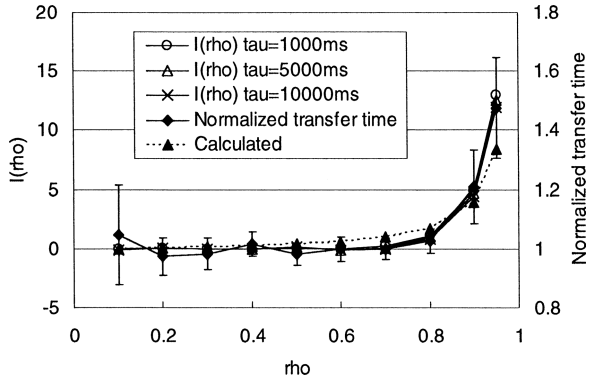

Fig. 8 Degree of performance degradation (Pareto distributed flow size).

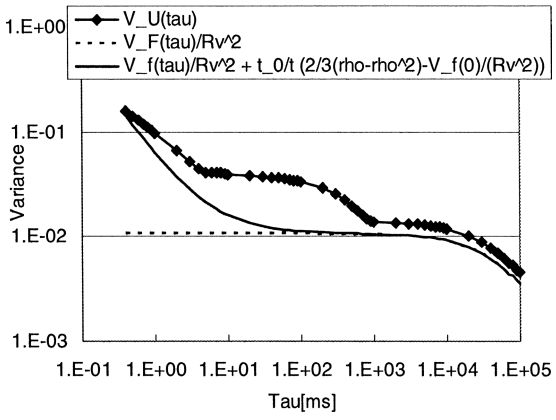

Fig. 9 Comparison of $V_U(t)$ and $V_F(t)$ (RTT= 1000 [ms]).

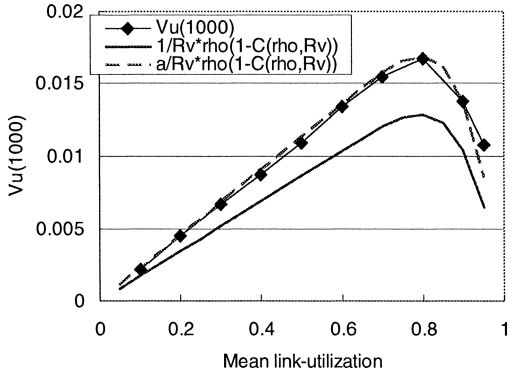

Fig. 10 Variance of link utilization (RTT= 1000 [ms]).

appeared when τ was shorter than RTT (= 1000 ms). As mentioned in the footnote in Sect. 3.1, this is because the packets arrived as bursts of windows rather than at random, which increased the burstiness of the traffic. However, when τ was longer than RTT, we observed that $V_U(\tau) \simeq V_S(\tau)/R_V^2$. This is because the bursty nature of packet arrival at intervals of the RTT was not apparent when we measured the link utilization with an interval longer than RTT.

Figure 10 shows the variance of utilization. Here, $V_U(\tau)$ differs from $\frac{\rho}{R_V}(1 - \mathbf{C}(R_V, \rho))$. However, to calculate $I(\rho)$, it is not necessary for $V_U(\tau)$ to agree with

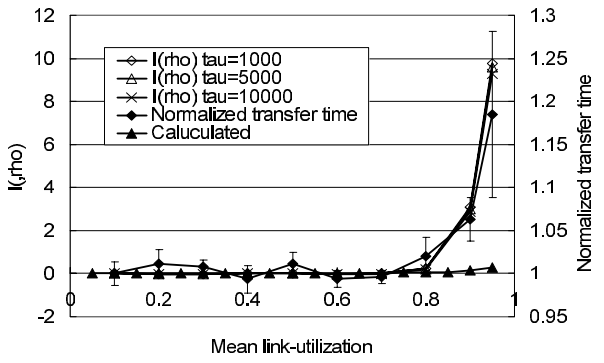

Fig. 11 Degree of performance degradation (RTT= 1000 [ms]).

$\frac{\rho}{R_V}(1 - \mathbf{C}(R_V, \rho))$ but the ratio of the two values should be constant despite the link utilization varying. In Fig. 10, we can see that the latter condition holds from the fact that $\frac{\rho}{R_V}(1 - \mathbf{C}(R_V, \rho))$ multiplied by a constant ("a" in the figure) agrees with $V_U(\tau)$. We can also see that the peak in the variance of link utilization is about $\rho = 0.8$, which is higher than that in the RTT= 20 [ms] case. This is because a large RTT led to a small actual maximum throughput for a flow (r') and large R_V. Thus, $\mathbf{C}(R_V, \rho)$, the probability of the number of flows exceeding R_V, started to increase at a higher link utilization. The variance of link utilization, which is proportional to $\rho(1 - \mathbf{C}(R_V, \rho))$, also started to decrease at a higher link utilization.

Figure 11 shows the $I(\rho)$ measured with τ = 1000, 5000, and 10000, the degree of TCP performance degradation, and the TCP performance degradation calculated using [4]. In this case, TCP throughput modeling significantly under-estimated the TCP performance degradation. This is because, while the packets arrived as window bursts in this case, the M/M/1 modeling cannot take this effect take into account. On the other hand, our method can detect the performance degradation even in this case. This is because our method is based not on the packet-level modeling but on the flow-level modeling (M/G/R/PS model), where the packet-level burst does not have to be taken into account. And by measuring the link utilization with a time-scale larger than RTT, we could can observe the flow behavior and detect the performance degradation of TCP flows.

5. Conclusion and Future Work

We proposed a method of detecting TCP performance degradation using only the mean and variance of bottleneck-link utilization. To do this, we gave the mathematical relationship between the variance of link utilization and the TCP performance using the M/G/R/PS model. Our method has the advantage that a network operator can determine whether or not the degradation originates from congestion of his/her own network. In addition, compared with the conventional method, the management cost can be greatly decreased while providing accurate detection of performance degradation because it requires only measuring link utilization. Simulation showed that our detection method is ap-

plicable to a large RTT and a heterogeneous environment. In this paper, as is described in Sect. 2, we treated the case where only TCP flows exist in the network, whereas current Internet traffic includes some UDP traffic. Because UDP does not have a congestion avoidance mechanism, while TCP does, basically UDP traffic always consume link bandwidth first and TCP traffic uses the remaining bandwidth. Thus, in the evaluation of TCP performance, we can consider that the physical bandwidth is decreased by the amount of UDP traffic and estimate the degree of performance degradation if we can once obtain the ratio of UDP traffic to the total traffic. The extension of our method explained above remains as future work.

References

[1] N. Ansari, G. Cheng, S. Israel, Y. Luo, J. Ma, and L. Zhu, "QoS provision with path protection for next generation SONET," Proc. IEEE ICC2002, 2002.
[2] http://www.cisco.com/warp/public/732/Tech/nmp/netflow/index.shtml
[3] CAIDA cooperative association for Internet data analysis. http://www.caida.org/tools/
[4] J. Padhye, V. Firoiu, D. Towsley, and J. Kurose, "Modeling TCP throughput: A simple model and its empirical validation," Proc. ACM SIGCOMM'98, 1998.
[5] S.B. Fredj, T. Bonald, A. Proutiere, G. Regnie, and J.W. Roberts, "Statistical bandwidth sharing: A study of congestion at flow level," Proc. ACM SIGCOMM 2001, 2001.
[6] D.P. Heyman, T.V. Lakshman, and A.L. Neidhardt, "A new method for analysing feedback-based protocols with applications to engineering web traffic over the Internet," Proc. ACM SIGMETRICS'97, 1997.
[7] A.W. Berger and Y. Kogan, "Dimensioning bandwidth for elastic traffic in high-speed data networks," IEEE/ACM Trans. Netw., vol.8, no.5, pp.643–654, 2000.
[8] A. Riedl, T. Bauschert, M. Perske, and A. Probst, "Investigation of the M/G/R processor sharing model for dimensioning of IP access networks with elastic traffic," Proc. First Polish-German Teletraffic Symposium, 2000.
[9] H. Lee, S. Lee, and V. Choi, "The influence of the large bandwidth-delay product on TCP Reno, NewReno, and SACK," Proc. IEEE ICOIN 2001, 2001.
[10] R. Kawahara, K. Ishibashi, T. Asaka, S. Sumita, and T. Abe, "A method of bandwidth dimensioning and management using flow statistics," IEICE Trans. Commun., vol.E88-B, no.2, pp.643–653, Feb. 2005.
[11] R. Morris and D. Lin, "Variance of aggregated Web traffic," Proc. IEEE INFOCOM2000, 2000.
[12] 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, 2001.
[13] L. Kleinrock, Queuing Systems Volume 1, John Wiley & Sons, 1976.
[14] R.W. Wolff, Stochastic Modeling and the Theory of Queues, Prentice Hall, 1988.
[15] UCB/LBNL/VINT Network Simulator -ns (version 2). http://www.isi.edu/nsnam/ns
[16] S. Abe, T. Fujino, J. Yusheng, J. Matsukata, and S. Asano, "Long-range dependent traffic analysis and approximation method of queue-length probability for an academic information network," IEICE Trans. Commun. (Japanese Edition), vol.J86-B, no.12, pp.2487–2500, Dec. 2003.
[17] R. Kawahara, K. Ishibashi, T. Mori, T. Ozawa, and T. Abe, "Method of bandwidth dimensioning and management for aggregated TCP flows with heterogeneous access links," IEICE Trans. Commun., vol.E88-B, no.12, pp.4605–4615, Dec. 2005.
[18] V. Paxson and S. Floyd, "Wide-area. traffic: The failure of Poisson modeling," IEEE/ACM Trans. Netw., vol.3, no.3, pp.226–244, 1995.
[19] D.R. Cox and V. Isham, Point Processes, Chapman & Hall, 1980.

Appendix A: Proof of Eq. (4)

Preparations

First, we calculate the mean square of link utilization averaged over packet transfer time t_0, which will be used later. This is written as $\mathrm{E}[U_0^2]$, where we shorten $U_{\tau_0}(t)$ to $U_0(t)$, and calculated as

$$\begin{aligned}\mathrm{E}[U_0^2] &= \mathrm{E}\left[\left(\frac{1}{\tau_0}\int_0^{\tau_0} U(t)dt\right)^2\right] \\ &= \frac{1}{\tau_0^2}\int_0^{\tau_0}\int_0^{\tau_0} \mathrm{E}[U(t)\,U(s)]dt\,ds \\ &= \frac{2}{\tau_0^2}\int_0^{\tau_0}(\tau_0 - t)\mathrm{E}[U(t)\,U(0)]dt. \qquad (\mathrm{A}\cdot 1)\end{aligned}$$

See [19] for the last equation. Here,

$$\begin{aligned}&\mathrm{E}[U(0)\,U(t)] \\ &\quad= \Pr[U(t) = 1|U(0) = 1]\Pr[U(0) = 1] \\ &\quad= \Pr[U(t) = 1|U(0) = 1]\rho. \qquad (\mathrm{A}\cdot 2)\end{aligned}$$

Let x be the remaining transfer time for a packet when the packet is being transmitted at time 0. If we assume random packet arrival and assume that the probability of another packet arriving after the transmission of the current packet has finished is ρ, then

$$\Pr[U(t) = 1|U(0) = 1, x] = \begin{cases} 1 & t < x \\ \rho & t \geq x. \end{cases} \qquad (\mathrm{A}\cdot 3)$$

Thus, by integrating (A· 3) over x from 0 to τ_0, we have

$$\begin{aligned}&\Pr[U(t) = 1|U(0) = 1] \\ &\quad= \int_0^{\tau_0} \Pr[U(t) = 1|U(0) = 1, x]\,dP(x) \qquad (\mathrm{A}\cdot 4) \\ &\quad= (\rho t + (\tau_0 - t))\frac{1}{\tau_0}, \qquad (\mathrm{A}\cdot 5)\end{aligned}$$

where $P(x)$ is the distribution function of x, which we set as $P(x) = x/\tau_0$ $(0 \leq x \leq \tau_0)$.

Therefore, by substituting (A· 5) into (A· 2) and substituting the results into (A· 1), we have

$$\mathrm{E}[U_0^2] = \frac{1}{3}(\rho^2 + 2\rho). \qquad (\mathrm{A}\cdot 6)$$

Next, we consider the relationship between the link-utilization ratio and the number of flows. If we can assume that the probability of a packet being transmitted at time t is proportional to $S(t)$, that is, whichever is smaller of the

number of flows at time t or R_V, then because $\mathrm{E}[U(t)] = \rho = \mathrm{E}[S(t)]/R_V$, we have $\mathrm{E}[U(t)|F(t)] = \frac{S(t)}{R_V}$. If we can also assume that packets in a flow arrive at random and that $\mathrm{E}[U(t)]$ and $\mathrm{E}[U(t+\tau)]$ are independent and depend only on $F(t)$ and $F(t+\tau)$, respectively, then, for $\tau > \tau_0$, we also have $\mathrm{E}[U(t)\,U(t+\tau)|F(t)\,F(t+\tau)] = \frac{S(t)\,S(t+\tau)}{R_V^2}$. Taking the expectation of $F(t)$ and $F(t+\tau)$ in this equation, we obtain

$$\mathrm{E}[U(t)\,U(t+\tau)] = \frac{\mathrm{E}[S(t)\,S(t+\tau)]}{R_V^2}. \quad (\mathrm{A}\cdot 7)$$

We can replace $U(\cdot)$ with $U_0(\cdot)$ in (A· 7) if the number of flows hardly changes during time τ_0.

With these preparations, we can obtain the relationship between the variance of link utilization and the number of flows measured on the k-packet time-scale $k\tau_0$ for $k \geq 1$ (In the following calculation, we repeatedly use the equation $\rho R_V = \mathrm{E}[S(t)]$ without specific mention).

$$\begin{aligned}
&V_U(k\tau_0)\\
&= \mathrm{Var}\left[\frac{1}{k}\sum_{i=0}^{k-1} U_0(i\tau_0)\right]\\
&= \frac{1}{k^2}\sum_{i=0}^{k-1}\sum_{j=0}^{k-1}\left[\mathrm{E}\left[U_0(i\tau_0)\,U_0(j\tau_0)\right] - \mathrm{E}[U_0(0)]^2\right]\\
&= \frac{1}{k^2}\left[k\mathrm{E}[U_0(0)^2] + 2\sum_{j=0}^{k-1}(k-i)\mathrm{E}\left[U_0(i\tau_0)U_0(0)\right] - k^2\rho^2\right]\\
&\quad \text{(Ref. [19])}\\
&= \frac{1}{k^2}\left[\frac{k}{3}(\rho^2+2\rho) + 2\sum_{j=0}^{k-1}(k-i)\mathrm{E}\left[U_0(i\tau_0)U_0(0)\right]\right] - \rho^2\\
&= \frac{1}{k^2}\left[\frac{k}{3}(\rho^2+2\rho) + 2\sum_{j=0}^{k-1}(k-i)\frac{1}{R_V^2}\mathrm{E}\left[S(i\tau_0)S(0)\right]\right]\\
&\quad -\frac{E[S(0)]^2}{R_V^2}\\
&= \frac{1}{R_V^2k^2}\left[\frac{R_V^2k}{3}(\rho^2+2\rho) + 2\sum_{j=0}^{k-1}(k-i)\mathrm{E}\left[S(i\tau_0)S(0)\right]\right]\\
&\quad -\frac{E[S(0)]^2}{R_V^2}\\
&= \frac{1}{R_V^2k^2}\left[R_V^2k\rho^2 + 2\sum_{j=0}^{k-1}(k-i)\mathrm{E}\left[S(i\tau_0)S(0)\right] + k\mathrm{Var}[S(0)]\right]\\
&\quad +\frac{2}{3}\frac{\rho}{k} - \frac{2}{3}\frac{\rho^2}{k} - \frac{\mathrm{Var}[S(0)]}{R_V^2k} - \frac{\mathrm{E}[S(0)]^2}{R_V^2}\\
&= \frac{1}{R_V^2k^2}\left[k\mathrm{E}[S(0)^2] + 2\sum_{j=0}^{k-1}(k-i)\mathrm{E}\left[S(i\tau_0)S(0)\right]\right]\\
&\quad +\frac{2}{3}\frac{\rho}{k}(1-\rho) - \frac{\mathrm{Var}[S(0)]}{R_V^2k} - \frac{\mathrm{E}[S(0)]^2}{R_V^2}\\
&= \frac{1}{R_V^2k^2}\left[\mathrm{E}\left[\left(\sum_{i=0}^{k-1} S(i\tau_0)\right)^2\right]\right] - \frac{\mathrm{E}[S(0)]^2}{R_V^2}\\
&\quad +\frac{2}{3}\frac{\rho}{k}(1-\rho) - \frac{\mathrm{Var}[S(0)]}{R_V^2k}\\
&= \frac{V_S(k\tau_0)}{R_V^2} + \frac{1}{k}\left(\frac{2\rho}{3}(1-\rho) - \frac{\mathrm{Var}[S(0)]}{R_V^2}\right).
\end{aligned}$$

Appendix B: Proof of Eq. (6)

When $\Pr[F(t) > R_V] \simeq 0$ (i.e. $\mathbf{C}(R_V,\rho) \simeq 0$), the number of flows follows a Poisson distribution [14], and $\mathrm{Var}[S(t)] \simeq \mathrm{Var}[F(t)] \simeq \mathrm{E}[F(t)] = \rho R_V$. However, when $\Pr[F(t) > R_V] > 0$ and performance is degraded, the number of flows is no longer the same as $S(t)$, and its variance differs from its mean. However, we can obtain the following derivation for the variance of $S(t)$, which is closely related to the variance of link utilization.

$$\begin{aligned}
&\mathrm{Var}[S(t)]\\
&= \sum_{n=1}^{R_V-1} n^2\Pr[S(t)=n] + {R_V}^2\mathbf{C}(R_V,\rho) - \mathrm{E}[S(t)]^2\\
&= \sum_{n=1}^{R_V-1} n^2 P_0\frac{(\rho R_V)^n}{n!} + {R_V}^2\mathbf{C}(R_V,\rho) - (\rho R_V)^2\\
&= \rho R_V\sum_{n=1}^{R_V-1} nP_0\frac{(\rho R_V)^{n-1}}{(n-1)!} + {R_V}^2\mathbf{C}(R_V,\rho) - (\rho R_V)^2\\
&\quad +{R_V}^2\mathbf{C}(R_V,\rho) - (\rho R_V)^2\\
&= \rho R_V\left(\mathrm{E}[S(t)] - (R_V-1)P_0\frac{(\rho R_V)^{R_V-1}}{(R_V-1)!} - R_V\mathbf{C}(R_V,\rho)\right)\\
&\quad +\rho R_V\left(1 - P_0\frac{(\rho R_V)^{R_V-1}}{(R_V-1)!} - \mathbf{C}(R_V,\rho)\right)\\
&\quad +{R_V}^2\mathbf{C}(R_V,\rho) - (\rho R_V)^2\\
&= \rho R_V(1-\mathbf{C}(R_V,\rho)),
\end{aligned}$$

where $P_0 := \Pr[F(t) = 0]$. The final equation is obtained through straight-forward calculation using the relation $\mathbf{C}(R_V,\rho) = P_0\frac{(\rho R_V)^{R_V}}{R_V!}\frac{1}{(1-\rho)}$.

Keisuke Ishibashi received his B.S. and M.S. in mathematics from Tohoku University, in 1993 and 1995, respectively and received his Ph.D. in information science and technology from the University of Tokyo, in 2005. Since joining NTT in 1995, he has been engaged in research on traffic issues in computer communication networks. Dr. Ishibashi received IEICE's Young Researcher's Award and Information Network Research Award in 2002. He is a member of the IEEE and the Operations Research Society of Japan.

Ryoichi Kawahara received a B.E. in Automatic Control, an M.E. in Automatic Control, and a Ph.D. in Telecommunication Engineering from Waseda University, Tokyo, Japan, in 1990, 1992, and 2001, respectively. Since joining NTT in 1992, he has been engaged in research on traffic control for telecommunication networks. He is currently working on teletraffic issues in computer communication networks at the NTT Service Integration Laboratories. Dr. Kawahara received IEICE's Young Researcher's Award in 1999, Best Paper Award in 2003, and Telecommunication Management Research Award in 2005. He is a member of the Operations Research Society of Japan.

Takuya Asaka received his B.E. and M.E. in industrial and management system engineering from Waseda University in 1988 and 1990 and his Ph.D. in global information and telecommunication from the same institution in 2001. He joined NTT Laboratories in 1990. He was also a research fellow at the Telecommunications Advancement Organization of Japan and a visiting researcher at the Global Information and Telecommunication Institute of Waseda University from 1998 to 2000. His research interests include performance evaluations of communication networks and traffic control. He is currently an associate professor at the Graduate School of Informatics of Kyoto University. He is a member of IEEE and the Operations Research Society of Japan.

Masaki Aida received his B.S. and M.S. in Theoretical Physics from St. Paul's University, Tokyo, Japan, in 1987 and 1989 respectively, and received the Ph.D. in Telecommunications Engineering from the University of Tokyo, Japan, in 1999. Since joining NTT Laboratories in 1989, he has been mainly engaged in research on traffic issues in computer communication networks. From 1998 to 2001, he was a manager at the Traffic Research Center, NTT Advanced Technology Corporation (NTT-AT). Since April 2005, he has been an Associate Professor at the Faculty of System Design, Tokyo Metropolitan University. His current interests include traffic issues in communication systems. Dr. Aida received the IEICEYoung Researcher's Award in 1996. He is a member of the IEEE and the Operations Research Society of Japan.

Satoshi Ono received the B.E. and M.E. degrees in electronics engineering from the University of Tokyo in 1977 and 1979, respectively. He received the Dr.Eng. degree in electrical engineering from the University of Tokyo in 1982. From 1982 to 1991, he worked for NTT Basic Research Laboratories and NTT Software Laboratories as a researcher in parallel processing and program analysis. During the period 1991 to 1992, he was with the engineering department of the University of Tokyo, as a visiting associate professor in parallel processing and program verification. From 1992 to 2005, he had been with the Global Computing Laboratory in NTT Software Laboratories and a research group leader of NTT Information Sharing Platform Laboratories. He is currently a professor at the Center for Continuing Professional Development, Kogakuin University. His current fields of interest include a traceable time distribution system, timestamping systems, and advanced mobile data services. He is a member of IEEE and ACM.

Shoichiro Asano had graduated at Electronic Engineering, Faculty of Engineering, University of Tokyo, in 1970. He received the ME and DE, both from University of Tokyo, in 1972 and 1975, respectively. Dr. Asano is a Director and Professor of the Infrastructure System Research Division, National Institute of Informatics (NII) and a Professor in the Graduate School of Information Science and Technology, University of Tokyo. His current researches are mainly focused on Optical Network Architectures and development of SuperSINET in Japan.