Research Paper

A link utilization estimation scheme for nodes with multiple-queues in output ports

Kiyofumi IGAI[1] and Eiji OKI[2]

[1,2]*Department of Communication Engineering and Informatics, Graduate School of Informatics and Engineering*

ABSTRACT

This paper proposes a delay-based link utilization estimation scheme that assumes each output port in each node follows the multiple-queue model. The conventional alternative assumes only a single-queue model, where link utilization is estimated by using the probability of that the round-trip time of the packet exhibits the minimum delay. However, nodes in an actual network have multiple queues in each output port. Due to the multiple-queue model, the conventional scheme is unable to determine link utilization. This is because the probability that a packet has minimum delay is not directly related to the link utilization. The proposed scheme measures the probability of packet delay and the ratios of probe packets waiting in the queues; it solves simultaneous equations that include the probability of packet delay and queue utilization. Our simulation results show that it can estimate link utilization with error under 0.1.

KEYWORDS

Link utilization, estimation, active measurement, multiple queues

1 Introduction

Link utilization is the ratio of traffic rate to link capacity. Link utilization, U, is expressed by $U = \frac{T}{C}$, where C is link capacity and T is traffic rate passing through the link. Link utilization is used to judge if a network is congested, or link capacity is not satisfying traffic demands. To keep the quality of service while better utilizing network resources, network operators should control traffic, control routing, or enhance link capacity. For these goals, network operators should estimate the link utilization frequently and correctly.

Techniques that estimate network conditions by measuring one or more states in the network lie in the field of network tomography. There are two approaches: passive measurement and active measurement.

Passive measurement does not need to use probe packets and instead directly collects information from network nodes. The Simple Network Management Protocol (SNMP) [1] is one such protocol. This approach is able to collect link utilization frequently, and it has an advantage that the network states are not affected by the measurement because probe packets are not used. However, due to a limitation of administration authority, the network operators cannot always access all the nodes.

In active measurement, an observer injects probe packets from an observer host into the network. Active measurements schemes include the Train Of Packet Pair (TOPP) scheme and the Self Loading Periodic Stream (SLoPS) [2], [4]–[7]. In TOPP, there are two hosts, the transfer and receiving hosts. The transfer host sends probe packets to the receiving host. The receiving host analyzes the interval of packet arrivals. By analyzing the results, the observer estimates the available bandwidth on a path [2], [3], or the minimum link capacity on a path [4].

In SLoPS, an observer sends constant rate packet streams from an observer host to the network [5]–[7]. If the transfer rate of the packet stream is higher than

Received September 26, 2013; Revised December 2, 2013; Accepted December 5, 2013.

This work was supported in part by Strategic Information and Communication R&D Promotion Program of the Ministry of Internal Affairs and the Support Center for Advanced Telecommunications Technology Research (SCAT).

DOI: 10.2201/NiiPi.2014.11.7

the available bandwidth on the path, the one-way delay variation can be used to estimate the available bandwidth. The advantage of active measurement is that the observer is able to estimate the available bandwidth without administration authority in the network. However, this scheme must send streams of probe packets that have higher rate than the available bandwidth on the path. This influences the communication quality of users. For this reason, frequent probing is not suitable.

Studies to estimate link utilization based on round-trip time (RTT) measurements have been presented [8], [9], [14]. Link utilization is obtained from the probability of the minimum RTTs on a link, in which the RTT variation is due to only queuing delay. This scheme measures two RTTs from an observer host to [[both end nodes of the target link]] to measure the probability of minimum RTTs. The advantage of this scheme is that the measurements do not cause any link overload situation, because the transfer rate of the probe packets is much smaller than the available bandwidth. However, this scheme assumes that nodes have single queues in their output ports.

In a network that provides multiple service qualities, nodes process packets according their packet priority, which is associated with service quality. To this end, nodes have multiple queues on each output port. When each node receives a packet, it checks the packet's priority and the packet is stored in a queue corresponding to the priority. The priority flag in the packet header is used to determine priority [10], [11]. For example, Internet Protocol (IP) telephony requires lower latency than other applications, so nodes preferentially transmit packets of IP telephony.

Weighted Round-Robin (WRR) queuing is one of the packet scheduling algorithms [12]. WRR transmits each packet in a round-robin manner considering weights; the opportunity that the weighted round-robin pointer visits a queue is determined by the weight assigned to the queue. For example, consider a node that has three queues for high, middle and low priorities. If all queues are occupied by packets, the head-of-line packet in each queue is transmitted with a probability that is proportional to the ratio associated with the queue's weight.

The conventional RTT-based scheme assumes a single queue per port [8], [9], and so is not suitable for multiple-queue nodes. The conventional scheme judges that a link is utilized when a packet is delayed at the output port to the link. However, a head-of-line packet in a queue has to wait for its transmission chance in the multiple-queue node, because another queue may be selected for transmission. The observer is not able to judge the cause of packet delay in the multiple-queue node. Therefore, a scheme that can estimate link utilization in multiple-queue nodes is needed.

This paper proposes a delay-based link utilization estimation scheme that assumes the multiple-queue model for output ports. The proposed scheme measures the probability of packet delay and the ratio of the probe packets waiting in each queue, and solves simultaneous equations that include the probability of packet delay and queue utilization. Our simulation results show that its estimation error of link utilization is within 0.1.

The remainder of this paper is organized as follows. Section 2 describes two node models, which are a single queue model and a multiple queue model. Section 3 introduces our proposed scheme. Section 4 presents performance evaluations of the proposed scheme. Finally, section 5 concludes this study.

2 Node models

2.1 Single queue model

Fig. 1 shows the single queue model. When a packet passes through a node it experiences packet delay. One-way delay from a one node to the next-hop node consists of variable delay and fixed delay. Delay D_i for packet i is expressed by,

$$D_i = T_f + T_q(i). \tag{1}$$

T_f includes the fixed delay for forwarding and switching, serialization/de-serialization, and propagation. T_f is a constant value, i.e. independent of i, as long as the route of each packet is not changed. $T_q(i)$ is caused by queuing at the ingress node of the link, where packet i has to wait before being transmitted to the output port.

The queuing delay is measured by sending probe packets from the ingress node to the next hop node, or the egress node through the link. When the queue at the ingress node is empty, there is no queuing delay, i.e. $T_q(i) = 0$. If no queuing delay is observed, in other words $T_q(i) = 0$, the link is not utilized. On the other hand, if the queue is not empty, the delay for a packet passing through the link is varied by queuing. The measured delay is larger than the minimum delay, where $T_q(i) > 0$. In this case, the observe judges that the link is utilized.

In the single queue model, link utilization, U, is expressed by N_{min} and N_{other}. Let N_{min} be the number of probe packets whose delay is the minimum delay, and N_{other} be the number of probe packets whose delay is

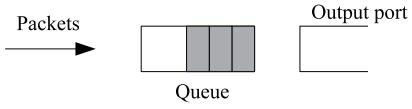


Fig. 1 Single-queue model.

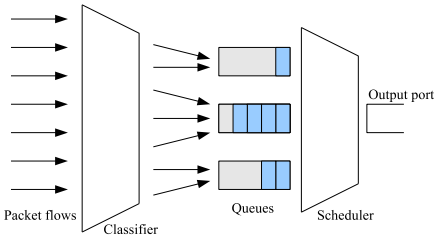


Fig. 2 Multiple-queue model.

larger than the minimum delay [8], [9].

$$U = 1 - \frac{N_{min}}{N_{min} + N_{other}} \tag{2}$$

2.2 Multiple queue model

Fig. 2 shows the multiple queue model. In a network that provides multiple service qualities, nodes have multiple queues in each output port. A packet arriving at the node is stored in the queue that corresponds to the packet's priority. Packets are transferred to the output port by a packet scheduling algorithm. For example, Priority Queuing (PQ) and Weighted Round Robin (WRR) Queuing scheduling algorithms are employed in the multiple queue model [12], [13].

In PQ, packets in a priority queue are transferred only if all higher-priority queues are empty. The benefit is that higher-priority packets are not influenced by lower-priority packets. On the other hand, lower-priority packets are influenced by higher-priority packets. Lower-priority packets are not transferred as long as even one higher-priority packet exists in a queue. This increases the delay of lower-priority packets.

In WRR, packets are transmitted in weighted round-robin manner according to a weight associated with each queue. For example, consider a node that has two queues in an output port; the weight of the high-priority queue is two and the weight of the low priority queue is one. When the two queues are full, the transfer ratio between the high-priority and low-priority queues is 2 : 1. The advantage of WRR is that the low-priority packets experience less delay than with PQ.

In the WRR queuing process, a weighted round-robin pointer visits queues in turn. The opportunity that the weighted round-robin pointer visits each queue is proportional to a weight associated with each queue. The total amount of traffic transmitted through an output port, T_{total}, is the sum of transmitted traffic rate T_i from queue i ($1 \leq i \leq n$), where n is the number of queues. T_{total} is expressed by,

$$T_{total} = \sum_{i=1}^{n} T_i. \tag{3}$$

Link utilization U is expressed by

$$U = \frac{T_{total}}{C} = \sum_{i=1}^{n} \frac{T_i}{C} = \frac{T_1}{C} + \frac{T_2}{C} + \cdots + \frac{T_n}{C}, \tag{4}$$

where C is link capacity. In Eq. (4), $\frac{T_i}{C}$ is replaced by U_i, which is defined as the queue utilization of queue i. Eq. (4) is expressed by,

$$U = \sum_{i=1}^{n} U_i. \tag{5}$$

An observer can measure the packet delay at queue i. The observer sends probe packets with the target priority, which are then stored at the corresponding queue. Let N_i^{min} be the number of packets with minimum delay in queue i, and N_i^{other} be the number of packets with non-minimum delay in the same queue. The packet delay ratio of queue i, which is denoted by X_i, is the ratio of the number of probe packets with the non-minimum delay to the sum of probe packets. X_i is defined by,

$$X_i = 1 - \frac{N_i^{min}}{N_i^{min} + N_i^{other}}. \tag{6}$$

The conventional scheme with single queue model is unable to handle nodes with WRR. The WRR scheduler does not always immediately transmit packets that are stored at the head of line of the queue, as another queue may be selected to transmit its head-of-line packet. Even if the queue is empty, delay variation is caused. Therefore, X_i is not equal to U_i. Note that, in PQ, link utilization can be estimated by sending probe packets to the lowest priority queue by using the conventional scheme.

3 Proposed scheme

The proposed scheme estimates link utilization under the multiple-queue model. The proposed scheme focuses on $U - X_i$, which is the probability of probe packets being processed immediately when there are some packets at other queues. Link utilization is calculated by solving simultaneous equations that are associated with link utilization U, queue utilization U_i and packet delay ratio X_i. We assume that w_i, which is the weight of queue i in WRR, is known. Customers may be informed by network operators of their service classes including priorities and how each service is provided together with its associated weight in the case of WRR, or may be able to guess the weight from the network operators' disclosed information.

3.1 Formulation for link utilization estimation

3.1.1 Cases with $n = 2, 3$

The queue indices are denoted by $i = 1, 2, \cdots, n$. μ_i is the processing rate of queue i per unit time. U_i is the utilization of queue i. P_i is the probability that the scheduler processes queue i. The processing rate per unit time of the output port, which is denoted by μ, is constant. WRR apportions the processing rate μ to μ_i, which is the processing rate for queue i. μ_i is expressed by,

$$\mu_i = \frac{w_i}{\sum_{k=1}^{n} w_k}\mu. \tag{7}$$

Note that μ and μ_i are introduced for convenience in order to derive formulations that can estimate link utilization, but they disappear in the process of obtaining the variables that should be determined.

We consider the case where $n = 2$, i.e. an output port has two queues. The relationships of U, U_i, X_i, and P_i are obtained as follows.

$$U = X_1 + P_1(1 - U_1)U_2 \tag{8}$$

$$U = X_2 + P_2U_1(1 - U_2) \tag{9}$$

$$U = U_1 + U_2 \tag{10}$$

In Eq. (8), we consider that a probe packet enters queue 1. A link is utilized, either when the probe packet has non-minimum delay, or when the probe packet has the minimum delay under the condition that queue 2 has at least one packet and the scheduler allows the probe packet in queue 1 to be transmitted immediately. Therefore, Eq. (8) indicates that, when we consider queue 1, U is the sum of two probabilities, probability X_1 (a probe packet in queue 1 has non-minimum delay), and probability, $P_1(1 - U_1)U_2$ (a probe packet in queue 1 is processed immediately under the condition that no packet is stored in queue 1 and at least one packet is stored in queue 2). In the same way, Eq. (9) indicates that, when we consider queue 2, U is the sum of two probabilities, probability, X_2 (a probe packet in queue 2 has non-minimum delay), and probability $P_2U_1(1-U_2)$ (a probe packet in queue 2 is processed immediately under the condition that no packet is stored in queue 2 and at least one packet is stored in queue 1). Eq. (10) indicates that the link utilization is equal to the sum of all queue utilizations. X_1 and X_2 are observable, but U, U_1, and U_2 are not known. An observer solves the simultaneous equations in Eqs. (8)–(10) by using X_1 and X_2.

We show that P_1 and P_2 are expressed by U_i and w_i. A state transition diagram for the multiple queue model with $n = 2$ is illustrated in Fig. 3. There are two states. State i, where $i = 1, 2$, means that the WRR scheduler processes queue i. The number with the arrow from state i to state j, $i \neq j$, indicates the transition

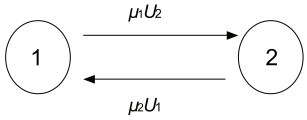


Fig. 3 State transition diagram for two queues. The number in each circle indicates the state of the queue that a scheduler processes.

probability that state i is changed to state j. State 1 is changed to state 2 if the scheduler finishes processing queue 1 and queue 2 has at least one packet. The transition probability that state 1 is changed to state 2 is expressed by μ_1U_2. State 2 is changed to state 1 if the scheduler finishes processing queue 2 and queue 1 has at least one packet. The transition probability that state 2 is changed to state 1 is expressed by μ_2U_1.

The equilibrium state in Fig. 3 gives the following relationship, given the probability of the scheduler processing each queue.

$$\mu_1U_2P_1 = \mu_2U_1P_2 \tag{11}$$

The left side of Eq. (11) is the transition rate from state 1 to state 2. The right side of Eq. (11) is the transition rate from state 2 to state 1.

By using the relationship of $P_1 + P_2 = 1$ and Eq. (7), P_1 and P_2 are given by,

$$P_1 = \frac{\mu_2U_1}{\mu_1U_2 + \mu_2U_1} = \frac{w_2U_1}{w_1U_2 + w_2U_1} \tag{12}$$

$$P_2 = \frac{\mu_1U_2}{\mu_1U_2 + \mu_2U_1} = \frac{w_1U_2}{w_1U_2 + w_2U_1}. \tag{13}$$

Eqs. (12) and (13) are used to solve the simultaneous equations in Eqs. (8)–(10).

Next, we consider the case where $n = 3$, the output port has three queues. In the same way as $n = 2$, the relationships of U, U_i, X_i, and P_i with $n = 3$ are given as follows.

$$\begin{aligned} U = {} & X_1 + P_1(1 - U_1)U_2U_3 \\ & + \frac{P_1}{P_1 + P_3}(1 - U_1)(1 - U_2)U_3 \\ & + \frac{P_1}{P_1 + P_2}(1 - U_1)U_2(1 - U_3) \end{aligned} \tag{14}$$

$$\begin{aligned} U = {} & X_2 + P_2U_1(1 - U_2)U_3 \\ & + \frac{P_2}{P_2 + P_3}(1 - U_1)(1 - U_2)U_3 \\ & + \frac{P_2}{P_1 + P_2}U_1(1 - U_2)(1 - U_3) \end{aligned} \tag{15}$$

$$\begin{aligned} U = {} & X_3 + P_3U_1U_2(1 - U_3) \\ & + \frac{P_3}{P_2 + P_3}(1 - U_1)U_2(1 - U_3) \end{aligned}$$

$$+ \frac{P_3}{P_1+P_3} U_1(1-U_2)(1-U_3) \quad (16)$$

$$U = U_1 + U_2 + U_3, \quad (17)$$

where $P_1 + P_2 + P_3 = 1$. Eq. (14) indicates that, for queue 1, U is the sum of three probabilities; X_1, (a probe packet in queue 1 has non-minimum delay), probability $\frac{P_1}{P_1+P_3}(1-U_1)(1-U_2)U_3$ (a probe packet in queue 1 is processed immediately under the condition that no packet is stored in queues 1 or 2 and at least one packet is stored in queue 3), and probability $\frac{P_1}{P_1+P_2}(1-U_1)U_2(1-U_3)$ (a probe packet in queue 1 is processed immediately under the condition that no packet is stored in queues 1 or 3 and at least one packet is stored in queue 2). Eqs. (15) and (16) are also obtained in the same way as Eq. (16).

P_i is expressed by U_i and w_i by considering a state transition diagram for the multiple queue model with $n = 3$, as illustrated in Fig. 4. There are three states. State i, where $i = 1, 2, 3$, means that the WRR scheduler processes queue i. The scheduler checks if the next queue is empty when the current process for a queue finishes. First, consider the case that state i is changed to state $M(i+1, 3)$. $M(k, n)$ is defined as $M(k, n) = n$ if ($k \mod n$) is zero and $M(k, n) = (k \mod n)$ otherwise, where ($x \mod y$) expresses the remainder of x divided by y. State i is changed to state $M(i+1, 3)$ if the scheduler finishes processing queue i and queue $i+1$ mod n has at least one packet. The transition probability that state i is changed to state $M(i+1, 3)$ is expressed by $\mu_i U_i U_{M(i+1,3)}$. Next, consider the case that state i is changed to state $M(i+2, 3)$. State i is changed to state $M(i+2, 3)$ if the scheduler finishes processing queue i, queue $M(i+1, 3)$ has no packet, and queue $M(i+2, 3)$ has at least one packet. The transition probability that state i is changed to state $M(i+2, 3)$ is expressed by $\mu_i U_i (1 - U_{M(i+1,3)}) U_{M(i+2,3)}$. The relationships of U, U_i, X_i, and P_i are obtained as follows.

$$(\mu_1 U_2 + \mu_1(1-U_2)U_3)P_1 = \mu_2 U_1(1-U_3)P_2 + \mu_3 U_1 P_3 \quad (18)$$

$$(\mu_2 U_3 + \mu_2 U_1(1-U_3))P_2 = \mu_1 U_2 P_1 + \mu_3(1-U_1)U_2 P_3 \quad (19)$$

$$(\mu_3 U_1 + \mu_3(1-U_1)U_2)P_3 = \mu_1(1-U_2)U_3 P_1 + \mu_2 U_3 P_2 \quad (20)$$

By using the relationship of $P_1 + P_2 + P_3 = 1$ and Eqs. (18)–(20), the simultaneous equations expressed by Eqs. (14)–(16) can be solved.

3.1.2 Generalization of formulation

Let us consider the general case, where the number of queues is n. The simultaneous equations are given by,

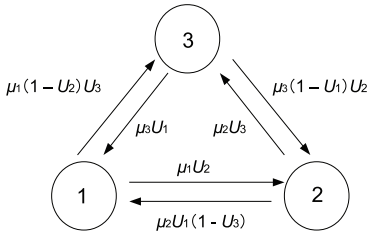


Fig. 4 State transition diagram for three queues. The number in each circle indicates the state of the queue that the scheduler processes.

$$U = X_i + (1-U_i)\sum_{k=1}^{n-1}\sum_{r=1}^{{}_{n-1}C_k}\left\{\frac{P_i}{P_i + \sum_{s\in E_{num}(i,k,r)} P_s} \prod_{s\in E_{num}(i,k,r)} U_s \prod_{t\in E_{num}(i,n-1,1)\backslash E_{num}(i,k,r)} (1-U_t)\right\}, \quad \forall i (1 \le i \le n) \quad (21)$$

$$U = \sum_{i=1}^{n} U_i, \quad (22)$$

where $\sum_{i=1}^{n} P_i = 1$. $E_{num}(i, q, r)$ is denoted by an enumeration set that excludes i, where $1 \le i \le n$. $E_{num}(i, q, r)$ enumerates all combinations of k ($1 \le k \le n-1$) queues from $n-1$ queues and returns the r-th combination set ($1 \le r \le {}_{n-1}C_k$). We present examples of $E_{num}(i, k, r)$ with $n = 2, 3, 4, 5$ in Appendix 5. We note that, if $\prod$ has no element in Eq. (21), the value of $\prod$ is specially treated as 1. For example, if $E_{num}(i, n-1, 1)\backslash E_{num}(i, k, r) = \emptyset$, where $\emptyset$ denotes an empty set, $\prod_{t\in E_{num}(i,n-1,1)\backslash E_{num}(i,k,r)}(1-U_t)$ is set to 1.

The relationship on P_i is given by,

$$\sum_{p(\neq i)} \alpha_{ip} P_i = \sum_{q(\neq i)} \beta_{qi} P_q. \quad (23)$$

α_{ip} is the transition rate moving from state i to state p. α_{ip} is expressed by,

$$\alpha_{ip} = \begin{cases} \mu_i U_p \prod_{k=i+1}^{p-1}(1-U_k), & i < p \\ \mu_i U_p \prod_{k=1}^{p-1}(1-U_k) \prod_{l=i+1}^{n}(1-U_l), & p < i. \end{cases} \quad (24)$$

β_{ip} is the transition rate moving from state q to state i. β_{ip} is expressed by,

$$\beta_{qi} = \begin{cases} \mu_q U_i \prod_{k=q+1}^{i-1}(1-U_k), & q < i \\ \mu_q U_i \prod_{k=1}^{i-1}(1-U_k) \prod_{l=q+1}^{n}(1-U_l), & i < q. \end{cases} \quad (25)$$

The same as for Eq. (21), if $\prod$ has no element in Eqs. (24) or (25), the value of $\prod$ is specially treated as 1.

3.2 Estimation for number of queues

The proposed scheme assumes that the number of queues of an output port in a node is known. The number of queues may be different in each node. In addition, an observer cannot determine the number of queues directly. The proposed scheme estimates the number of queues by checking probability distributions of probe packet delays. The proposed scheme estimates probability distributions for a target link by sending probe packets, in the same way as [8], [9], with various priority streams. Estimating probability distribution of delays for a target link is called restoration of the delay probability distribution. If there are multiple queues in an output port on the target link, the characteristics of probability distributions of the delays are different. The proposed scheme is able to judge how many different delay probability distributions are observed. This is called matching. Details of the procedures are described below.

3.2.1 Maximum number of probe streams

The maximum number of priority queues in an output port of a node, i.e. maximum number of priority classes, is defined as [11]. An observer sends probe packet streams, whose number is limited by the maximum number of priority classes, to both end nodes of the target link. If different packet streams belong to the same priority class, they are stored in the same queue. Next, we check if the delay properties of different packet streams, which includes packet-delay rate X_i, the average delay, and the delay distribution, match each other.

3.2.2 Restoration of delay probability distribution in target link

Let two probability distributions be F_x and F_y. F_x has a random variable X, the delay from observer host s to node p. F_y has a random variable Y, the delay from node p to node q. Fig. 5 shows a network model with observer host s, node p, and node q. F_{x+y} of the delay from the observer host s to node q is expressed by F_x and F_y [8], [9].

$$F_{x+y}(t) = \int_{-\infty}^{+\infty} F_x(t-z)F_y(z)dz \tag{26}$$

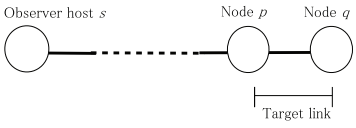


Fig. 5 Network model with observer host and target link between two nodes.

The observer measures the relative delay frequency distribution. Let $F_{sq}(i)$ be the relative delay frequency distribution from observer host s to node q, $F_{sp}(i)$ the relative delay frequency distribution from observer host s to node p, and $F_{pq}(k)$ the relative delay frequency distribution from node p to node q, where $k = 0, 1, \cdots, m$. When $k = 0$, the relative frequency of the measurable minimum delay is stored. Eq. (26) is expressed in discrete form by,

$$F_{sq}(m) = \sum_{k=0}^{m} F_{sp}(m-k)F_{pq}(m) \tag{27}$$

$F_{pq}(m)$ is obtained by,

$$F_{pq}(0) = \frac{F_{sq}(0)}{F_{sp}(0)} \quad (F_{sp}(0) \neq 0) \tag{28}$$

$$F_{pq}(m) = \frac{F_{sq}(m) - \sum_{k=0}^{m-1} F_{sp}(m-k)F_{pq}(m)}{F_{sp}(0)} \quad (m \neq 0, F_{sp}(0) \neq 0). \tag{29}$$

3.2.3 Estimation of the number of queues by checking the degree of coincidence

The number of queues is estimated from the number of matched estimated probability distributions in the target link. The number is determined by the normalization of the standard deviation and cross-correlation function. When there are relative frequency distributions $x(k)$ and $y(k)$ $(k = 0, \cdots, n)$, the degree of coincidence is calculated by,

$$R = \sum_{k=0}^{n} x(k)y(k) \tag{30}$$

$$S = \frac{R}{\sqrt{\sum_{k=0}^{n}(x(k))^2}\sqrt{\sum_{k=0}^{n}(y(k))^2}}. \tag{31}$$

When $S = 1$, $x(k)$ is equal to $y(k)$. In case of $S \geq 0.99$, we judge that both relative frequency distributions match. This matching procedure yields the number of queues in the output ports along the target link.

3.3 Estimation of scheduling algorithm

The scheduling algorithm used, PQ or WRR, may need to be estimated by measuring X_i $(1 \leq i \leq n)$. When the scheduling algorithm is PQ, the following relationship is satisfied.

$$X_1 \leq \cdots \leq X_{n-1} \leq X_n = U \tag{32}$$

Packets stored in higher-priority queues than the lowest-priority queue always affect lowest-priority packets. In PQ, X_n, which is measured by sending lowest-priority probe packets, is equal to U. When Eq. (32) is satisfied, the existence of PQ is possible.

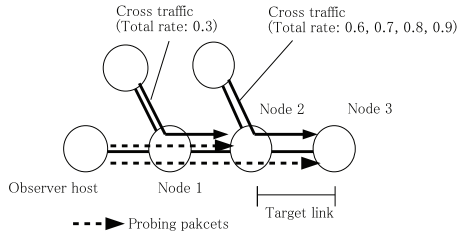


Fig. 6 Multi-hop network model 1.

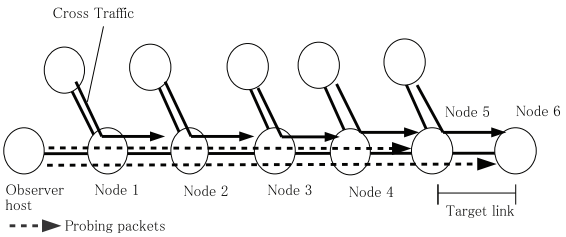


Fig. 7 Multi-hop network model 2.

However, in WRR, there is also the possibility that Eq. (32) is satisfied. Therefore, there is no assurance that the algorithm can be determined. When Eq. (32) is not satisfied, packets are not processed by the PQ algorithm.

4 Performance evaluation

We evaluated the proposed scheme in networks whose nodes used the WRR algorithm to transmit, by using a simulator that we developed. Examined networks are multiple hop networks, as shown in Figs. 6 and 7.

We define Δ as the estimation error; it is the absolute difference between an estimated value and an actual value. The actual utilization value is obtained by using byte counters of transmitted packets in each queue in this evaluation.

4.1 Evaluation settings

4.1.1 Settings for nodes and queues

Only queuing delay is considered at each node when evaluating the proposed scheme. Arriving packets are processed immediately and stored in the corresponding queue in an output port. Processing delay and serialization delay are ignored [15][1)] The weight of each queue is assumed to be known. When the number of queues is 2, $(w_1, w_2) = (2, 1)$ is set, and, when the number of queues is 3, $(w_1, w_2, w_3) = (3, 2, 1)$ is set.

[1)] To investigate the applicability of the proposed scheme for the multiple-queue network model at the first step, this evaluation considers only the queuing delay. In [14], how to incorporate factors of processing delay jitter and practical timer granularity values for link utilization estimation based on delay measurements is presented. The work in [14] can be also applied to our evaluation in the multiple-queue network model.

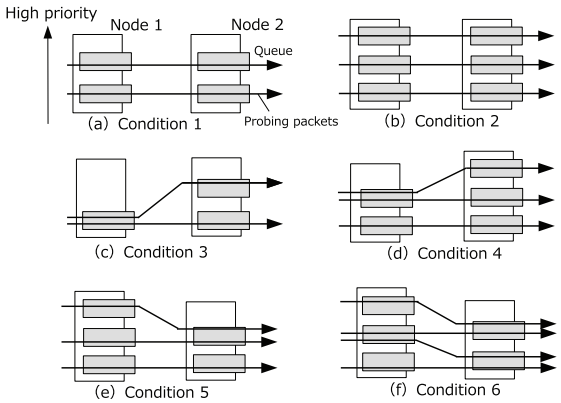


Fig. 8 Conditions of probe packet stream in network model 1.

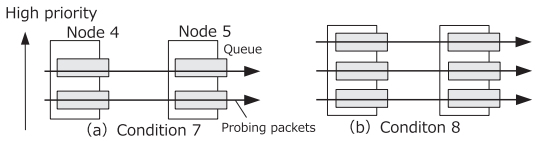


Fig. 9 Conditions of probe packet stream in network model 2.

Flows of probe packets streams are shown in Fig. 8, and the number of queues in the transit nodes is shown in Fig. 6 in network model 1. In addition, for network model 2, the flows are shown in Fig. 7, and the number of queues is shown in Fig. 9. In condition 7, the number of queues at the transit nodes is fixed to two, and, in condition 8, the number of queues at the middle nodes is fixed to three. These conditions are adopted to ireflect practical implementations [16].

4.1.2 Settings for cross traffic and links

Each link bandwidth is set to 1 [Gbps], and propagation delay is set to 1.6 [msec]. Cross traffic rate is set to 0.6, 0.7, 0.8, and 0.9 in a target link. The total rate of cross traffic is fixed to 0.3 in a non-target link. To evaluate the basic characteristics of the proposed scheme, cross traffic is not set in the backward direction. Furthermore, the rate of cross traffic is distributed to each queue randomly.

User Datagram Protocol (UDP) packets are generated with the sizes of 500 [bytes] and 1,200 [bytes] at the same probability. In actual networks, packet size distribution is biased, and several hundred byte packets and over 1,000 byte packets are often observed [17], our assumption of packet length follows this observation. The packet arrival interval distribution follows the geometric distribution.

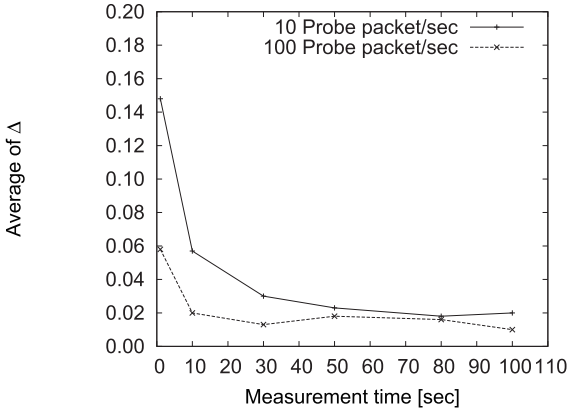


Fig. 10 Relationship between measurement time and estimated value.

4.1.3 Settings for probe packets

All probe packets are 32 [bytes] long; this is the default size of ICMP (Internet Control Message Protocol) request/response packets in Windows® Operating System.

4.1.4 Simulation granularity

Simulation granularity is set to 100 bytes per step. This value means that each node transmits 100 bytes from output ports per step. Smaller granularity than this increases the simulation time. We determined the simulation granularity by balancing evaluation accuracy against practical simulation times.

4.2 Relationship between measurement time and estimated value

Using condition 1 of Fig. 8, we investigate the relationship between the estimation time and the estimated value. Fig. 10 plots the relationship. To check the estimated value, estimation error Δ is used. When the number of probe packets loaded into each queue per second is 100, the estimated value is less than 0.02. For measurement times of 10 or more seconds, there is no significant change of the estimated value. In the case of 100 second probing periods with 10 packets per second, the estimated value nearly equals to that estimated under 10 second probing periods with 100 packets per second. This result shows that the estimated values are stable with over 1000 probe packets for each queue.

4.3 Estimation accuracy of link utilization

We use conditions 1-8 to investigate the estimation accuracy of link utilization. It is assumed that queue weights and the number of queues are known. The measurement time is set to 30 seconds and the packet per second value is set to 100, so the total number of probe packets is 3000 ($= 100 \times 30$) per probe packet stream.

Table 1 Estimation results of link utilization inder conditions 1-3.

	Estimation errors $\Delta \pm \sigma$		
Actual	Condition 1	Condition 2	Condition 3
0.6	0.036 ± 0.021	0.056 ± 0.026	0.041 ± 0.029
0.7	0.033 ± 0.028	0.045 ± 0.034	0.045 ± 0.029
0.8	0.032 ± 0.025	0.054 ± 0.042	0.032 ± 0.022
0.9	0.051 ± 0.027	0.064 ± 0.028	0.018 ± 0.015

Table 2 Estimation results of link utilization in conditions 4-6.

	Estimation errors $\Delta \pm \sigma$		
Actual	Condition 4	Condition 5	Condition 6
0.6	0.053 ± 0.026	0.052 ± 0.033	0.055 ± 0.034
0.7	0.060 ± 0.035	0.046 ± 0.039	0.061 ± 0.040
0.8	0.038 ± 0.022	0.085 ± 0.069	0.045 ± 0.028
0.9	0.044 ± 0.030	0.050 ± 0.039	0.035 ± 0.023

Tables 1 and 2 plot estimation results under network model 1 (Fig. 6). Estimation error Δ with standard deviation σ is expressed by $\Delta \pm \sigma$. In all cases, estimation errors are under 0.1. The correlation coefficient between actual utilization and estimation errors is investigated. The correlation coefficients under conditions 1, 2, 3, 4, 5, and 6 are -0.22, -0.08, -0.35, 0.08, 0.10, and -0.26, respectively. These results show that the estimation accuracy of the proposed scheme does not depend on link utilization in the target link.

We also investigate the coefficient of the correlation between the numbers of queues at node 5 and the estimation errors under conditions 5 and 6. The correlation coefficient is 0.35, which indicates a positive correlation. The estimation accuracy is impacted by the number of queues. If the number of queues increases, the number of variables in the non-linear simultaneous equations in Eqs. (21)–(25) increases which worsens the accuracy of the solutions.

Estimation results of link utilization in network model 2 (Fig. 7) are shown in Table 3. In all cases, the estimation errors are also under 0.1. The correlation coefficient between actual utilization and estimation errors is -0.30 under condition 7 and 0.61 under condition 8. These results indicate that, as the load of a target link rises, the estimation error increases. The reason is as follows. As the hop count increases, fewer probe packets have no queuing delay. The proposed scheme needs to obtain the exact percentage of probe packets with the minimum delay. However, since the hop counts are larger than those of conditions 1-6 and

Table 3 Estimation results of link utilization under conditions 7-8.

	Estimation errors $\Delta \pm \sigma$	
Actual	Condition 7	Condition 8
0.6	0.064 ± 0.049	0.036 ± 0.031
0.7	0.029 ± 0.032	0.041 ± 0.031
0.8	0.030 ± 0.045	0.083 ± 0.040
0.9	0.030 ± 0.027	0.091 ± 0.015

Table 4 Comparison of probing loads between proposal and conventional.

Load of proposed scheme [%]	Load of conventional scheme [%]
0.0512 ~ 0.2048	0.0256

the load of the target link is also high, the number of probe packets with the minimum delay becomes small, which degrades the estimation accuracy.

4.4 Comparison of probing loads between proposed and conventional schemes

We compare the probing loads created by the proposed scheme and the conventional scheme presented in [8], [9]. In both schemes, the probing load is proportional to the number of probe packets sent per second and their length. In addition, the load of the proposed scheme is proportional to the number of priority classes. The probing loads of both schemes are shown in Table. 4, where link bandwidth is set to 1 [Gbps], probe packet length is set to 32 [bytes], and probe packets per second is set to 1000. In the proposed scheme, the number of priority classes, or the number of probe packet streams, was set from 2 to 8.

Since the conventional scheme does not set any priority class, the load is a constant 0.0256%. In contrast, the proposed scheme creates larger loads because it sends several probe packet streams into a path. However, the maximum load of the proposed scheme is only 0.2%, which has an insignificant impact on user traffic.

5 Conclusions

This paper has proposed a scheme that can estimate link utilization where it is assumed that each output port at each node has multiple queues. Since the conventional scheme assumes that nodes have single queue in each output port, the conventional scheme does not support the multiple queue condition. The proposed scheme measures the non-minimum delay probability of probe packets by observing the ratios of probe packets waiting in the queues. It estimates link utilization by solving simultaneous equations that include the non-minimum delay probability of probe packets in each queue. Evaluation results showed that the estimation error of link utilization is under 0.1 in a multi-hop network with multiple-queue nodes.

References

[1] J. Case, M. Fedor, M. Schoffstall, and J. Davin. "A Simple Network Management Protocol (SNMP)," *RFC1157*, May 1990.

[2] B. Melander, M. Bjorkman, and P. Gunningberg, "A New End-to-End Probing and Analysis Method for Estimating Bandwidth Bottlenecks," *Proc. GLOBECOM'00*, pp.415–420, San Francisco, USA, 2000.

[3] N. Hu and P. Steenkiste, "Estimating Available Bandwidth Using Packet Pair Probing," *Computer Science Technical Reports*, Carnegie Mellon Univ. Pittsburgh, USA, 2002.

[4] C. Dorolis, "Packet-Dispersion Techniques and a Capacity-Estimation Methodology," *IEEE/ACM TRANSACTINOS*, vol.12, no.6, pp.963–977, 2004.

[5] M. Jain and C. Dovrolis, "End-to-End Available Bandwidth: Measurement Methodology, Dynamics, and Relation with TCP Through-put," *Proc. the 2002 SIGCOMM*, pp.295–308, Pennsylvania, USA, 2002.

[6] V. Ribeiro, R. Riedi, R. Baraniuk, J. Navratil, and L. Cottrell, "pathChirp: Efficient Available Bandwidth Estimation for Network Paths," *Proc. Passive and Active Measurement Workshop*, San Diego, USA, 2003.

[7] M. Jain and C. Dovrolis, "End-to-end Estimation of the Available Bandwidth Variation Range," *Proc. the 2005 ACM SIGMETRICS*, pp.265–276, Banff, Alberta, Canada, 2005.

[8] K. Igai and E. Oki, "A Simple Link-Utilization Estimation Scheme Based on RTT Measurement," *Proc. of the IEEE-ISAS 2011*, pp.266–270, Yokohama, Japan, 2011.

[9] K. Igai and E. Oki, "A Simple Estimation Scheme for Upper Bound of Link Utilization Based on RTT Measurementt," *Cyber Journals: Journal of Selected Areas in Telecommunications (JSAT)*, August 2011 Edition, pp.10–16, Aug. 2011.

[10] P. Almquist, "Type of Service in the Internet Protocol Suite," *RFC 1349*, July 1992.

[11] K. Nichols, S. Blake, F. Baker, and D. Black. "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers," *RFC2464*, Dec. 1998.

[12] C. Semeria, "Supporting Differentiated Service Classes: Queue Scheduling Disciplines," *White Paper*, Juniper Networks, Inc. Sunnyvale, USA, 2001.

[13] N. Yamanaka, K. Shiomoto, and E. Oki, *GMPLS Technologies: Broadband Backbone Networks and Systems*, Boca Raton, CRC Press, 2005.

[14] K. Igai and E. Oki, "Scheme for Estimating Upper-Bound of Link Utilization Based on RTT Measurements with Consideration of Processing Delay Jitter and Practical Timer Granularity," *Cyber Journals: Multidisciplinary Journals in Science and Technology, Journal of*

Selected Areas in Telecommunications (JSAT), July Edition, 2013. vol.3, issue 7, pp.17–24, 2013.

[15] "[Ns-developers] Google SoC 2009 - Modeling routers," http://mailman.isi.edu/pipermail/ns-developers/2009-March/005601.html (Accessed in Aug. 2010)

[16] "Configuring Priority Queueing," *Cisco IOS Quality of Service Solutions Configuration Guide, Release 12.2*, Cisco Systems, Inc., 2012.

[17] R. Sinha, C. Papadopoulos, and J. Heidemann, "Internet Packet Size Distributions: Some Observations," *Technical Report ISI-TR-2007-643*, USC/Information Sciences Institute, California, USA, 2007.

Appendix

A. Examples of $E_{num}(i,k,r)$

Tables 5-8 present examples of $E_{num}(i,k,r)$ and $E_{num}(i,n-1,1)\backslash E_{num}(i,k,r)$, which are used in Eq. (21), with $n = 2,3,4,5$ respectively, where $1 \le i \le n$, $1 \le k \le n-1$, and $1 \le r \le {}_{n-1}\mathrm{C}_k$.

Table 5 Example of $E_{num}(i,k,r)$ with $n = 2$.

i	k	r	$E_{num}(i,k,r)$	$E_{num}(i,n-1,1)\backslash E_{num}(i,k,r)$
1	1	1	{2}	∅
2	1	1	{1}	∅

Table 6 Example of $E_{num}(i,k,r)$ with $n = 3$.

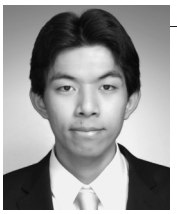
i	k	r	$E_{num}(i,k,r)$	$E_{num}(i,n-1,1)\backslash E_{num}(i,k,r)$
1	1	1	{2}	{3}
1	1	2	{3}	{2}
1	2	1	{2,3}	∅
2	1	1	{1}	{3}
2	1	2	{3}	{1}
2	2	1	{1,3}	∅
3	1	1	{1}	{2}
3	1	2	{2}	{1}
3	2	1	{1,2}	∅

Table 7 Example of $E_{num}(i,k,r)$ with $n = 4$.

i	k	r	$E_{num}(i,k,r)$	$E_{num}(i,n-1,1)\backslash E_{num}(i,k,r)$
1	1	1	{2}	{3,4}
1	1	2	{3}	{2,4}
1	1	3	{4}	{2,3}
1	2	1	{2,3}	{4}
1	2	2	{2,4}	{3}
1	2	3	{3,4}	{2}
1	3	1	{2,3,4}	∅
2	1	1	{1}	{3,4}
2	1	2	{3}	{1,4}
2	1	3	{4}	{1,3}
2	2	1	{1,3}	{4}
2	2	2	{1,4}	{3}
2	2	3	{3,4}	{1}
2	3	1	{1,3,4}	∅
3	1	1	{1}	{2,4}
3	1	2	{2}	{1,4}
3	1	3	{4}	{1,2}
3	2	1	{1,2}	{4}
3	2	2	{1,4}	{2}
3	2	3	{2,4}	{1}
3	3	1	{1,2,4}	∅
4	1	1	{1}	{2,3}
4	1	2	{2}	{1,3}
4	1	3	{3}	{1,2}
4	2	1	{1,2}	{3}
4	2	2	{1,3}	{2}
4	2	3	{2,3}	{1}
4	3	1	{1,2,3}	∅

Table 8 Example of $E_{num}(i,k,r)$ with $n = 5$.

i	k	r	$E_{num}(i,k,r)$	$E_{num}(i,n-1,1)\backslash E_{num}(i,k,r)$
1	1	1	{2}	{3,4,5}
1	1	2	{3}	{2,4,5}
1	1	3	{4}	{2,3,5}
1	1	4	{5}	{2,3,4}
1	2	1	{2,3}	{4,5}
1	2	2	{2,4}	{3,5}
1	2	3	{2,5}	{3,4}
1	2	4	{3,4}	{2,5}
1	2	5	{3,5}	{2,4}
1	2	6	{4,5}	{2,3}
1	3	1	{2,3,4}	{5}
1	3	2	{2,3,5}	{4}
1	3	3	{2,4,5}	{3}
1	3	4	{3,4,5}	{2}
1	4	1	{2,3,4,5}	∅
2	1	1	{1}	{3,4,5}
2	1	2	{3}	{2,4,5}
2	1	3	{4}	{2,3,5}
2	1	4	{5}	{2,3,4}
2	2	1	{1,3}	{4,5}
2	2	2	{1,4}	{3,5}
2	2	3	{1,5}	{3,4}
2	2	4	{3,4}	{1,5}
2	2	5	{3,5}	{1,4}
2	2	6	{4,5}	{1,3}
2	3	1	{1,3,4}	{5}
2	3	2	{1,3,5}	{4}
2	3	3	{1,4,5}	{3}
2	3	4	{3,4,5}	{1}
2	4	1	{1,3,4,5}	∅
3	1	1	{1}	{2,4,5}
3	1	2	{2}	{1,4,5}
3	1	3	{4}	{1,2,5}
3	1	4	{5}	{2,3,4}
3	2	1	{1,2}	{4,5}
3	2	2	{1,4}	{2,5}
3	2	3	{1,5}	{2,4}
3	2	4	{2,4}	{1,5}
3	2	5	{2,5}	{1,4}
3	2	6	{4,5}	{1,2}
3	3	1	{1,2,4}	{5}
3	3	2	{1,2,5}	{4}
3	3	3	{1,4,5}	{2}
3	3	4	{2,4,5}	{1}
3	4	1	{1,2,4,5}	∅
4	1	1	{1}	{2,3,5}
4	1	2	{2}	{1,3,5}
4	1	3	{3}	{1,2,5}
4	1	4	{5}	{1,2,3}
4	2	1	{1,2}	{3,5}
4	2	2	{1,3}	{2,5}
4	2	3	{1,5}	{2,3}
4	2	4	{2,3}	{1,5}
4	2	5	{2,5}	{1,3}
4	2	6	{3,5}	{1,2}
4	3	1	{1,2,3}	{5}
4	3	2	{1,2,5}	{3}
4	3	3	{1,3,5}	{2}
4	3	4	{2,3,5}	{1}
4	4	1	{1,2,3,5}	∅
5	1	1	{1}	{2,3,4}
5	1	2	{2}	{1,3,4}
5	1	3	{3}	{1,2,4}
5	1	4	{4}	{1,2,3}
5	2	1	{1,2}	{3,4}
5	2	2	{1,3}	{2,4}
5	2	3	{1,4}	{2,3}
5	2	4	{2,3}	{1,4}
5	2	5	{2,4}	{1,3}
5	2	6	{3,4}	{1,2}
5	3	1	{1,2,3}	{4}
5	3	2	{1,2,4}	{3}
5	3	3	{1,3,4}	{2}
5	3	4	{2,3,4}	{1}
5	4	1	{1,2,3,4}	∅



Kiyofumi IGAI

Kiyofumi IGAI received the B.E. and M.E. degrees from the University of Electro-Communications, Tokyo, Japan, 2011 and 2013, respectively. His research interests include network measurement and network tomography, particularly estimation schemes for available bandwidth or link utilization.

Eiji OKI

Eiji OKI is a Professor at The University of Electro-Communications, Tokyo, Japan. He received the B.E. and M.E. degrees in instrumentation engineering and a Ph.D. degree in electrical engineering from Keio University, Yokohama, Japan, in 1991, 1993, and 1999, respectively. In 1993, he joined Nippon Telegraph and Telephone Corporation (NTT) Communication Switching Laboratories, Tokyo, Japan. He has been researching network design and control, traffic-control methods, and high-speed switching systems. From 2000 to 2001, he was a Visiting Scholar at the Polytechnic Institute of New York University, Brooklyn, New York, where he was involved in designing terabit switch/router systems. He was engaged in researching and developing high-speed optical IP backbone networks with NTT Laboratories. He joined The University of Electro-Communications, Tokyo, Japan, in July 2008. He has been active in standardization of path computation element (PCE) and GMPLS in the IETF. He wrote more than ten IETF RFCs. Prof. Oki was the recipient of the 1998 Switching System Research Award and the 1999 Excellent Paper Award presented by IEICE, the 2001 Asia-Pacific Outstanding Young Researcher Award presented by IEEE Communications Society for his contribution to broadband network, ATM, and optical IP technologies, and the 2010 Telecom System Technology Prize by the Telecommunications Advanced Foundation. He has authored/co-authored four books, *Broadband Packet Switching Technologies*, published by John Wiley, New York, in 2001, *GMPLS Technologies*, published by CRC Press, Boca Raton, FL, in 2005, *Advanced Internet Protocols, Services, and Applications*, published by Wiley, New York, in 2012, and *Linear Programming and Algorithms for Communication Networks*, CRC Press, Boca Raton, FL, in 2012. He is a Fellow of IEEE and a Fellow of IEICE.