Inferring Link Loss Rates from Unicast-Based End-to-End Measurement

<table>
<thead>
<tr>
<th>著者</th>
<th>鈴木 清敏, 高坂 謙, 大江 由紀</th>
</tr>
</thead>
<tbody>
<tr>
<td>出版物名</td>
<td>IEICE Transactions on Communications</td>
</tr>
<tr>
<td>巻</td>
<td>85</td>
</tr>
<tr>
<td>号</td>
<td>1</td>
</tr>
<tr>
<td>頁</td>
<td>70-78</td>
</tr>
<tr>
<td>年</td>
<td>2002-01-01</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://hdl.handle.net/10228/00006307">http://hdl.handle.net/10228/00006307</a></td>
</tr>
</tbody>
</table>
Inferring Link Loss Rates from Unicast-Based End-to-End Measurement

Masato TSURU††††, Tetsuya TAKINE†††*, and Yuji OIE††††∗∗, Regular Members

SUMMARY In the Internet, because of huge scale and distributed administration, it is of practical importance to infer network-internal characteristics that cannot be measured directly. In this paper, based on a general framework we proposed previously, we present a feasible method of inferring packet loss rates of individual links from end-to-end measurement of unicast probe packets. Compared with methods using multicast probes, unicast-based inference methods are more flexible and widely applicable, whereas they have a problem with imperfect correlation in concurrent events on paths. Our method can infer link loss rates under this problem, and is applicable to various path-topologies including trees, inverse trees and their combinations. We also show simulation results which indicate potential of our unicast-based method.

key words: link loss rate, end-to-end measurement, statistical inference

1. Introduction

The Internet is currently shifting towards a social and economical infrastructure, which needs to be operated in a reliable and efficient way, and thus its characteristics should be measurable. However, because of huge scale and distributed administration, it is difficult to measure its internal states and performance on the Internet. Therefore, it is of practical importance to develop statistical methods to infer network-internal characteristics that cannot be measured directly from end-to-end path measurement.

In [1], we have studied a general principle of inferring various characteristics (i.e., occurrence probabilities of some states) of links from given characteristics of paths with an arbitrary “path-topology” (by which we mean a topological structure of measurable paths).

In this paper, based on our general framework, we present a feasible method of inferring packet loss rates of individual links from end-to-end measurement of unicast probe packets among several senders’ and receivers’ nodes. Our ultimate goal is to infer characteristics of individual (directed) links in a network with an arbitrary topology from end-to-end path measurement. We consider a set of paths covering all links whose characteristics should be inferred. These paths can be regarded as an appropriate combination of tree and inverse tree path-topologies under certain conditions. Thus, if we can infer characteristics of links from end-to-end path measurement on a tree and an inverse tree, then it is expected that we can infer characteristics of links on a general path-topology. Indeed, as mentioned later (Sect. 4), for most of path-topologies in actual networks, loss rates of individual links can be inferred by combining inference of loss rates on trees and inverse trees. Therefore, we develop a technique to infer link loss rates on both trees and inverse trees, which are essential to inference on general topologies.

For tree path-topologies, extensive researches related to multicast-based inference of network-internal characteristics have been done in MINC project. In [2], they employed end-to-end multicast probe packets from a root sender to many leaf receivers, and they utilized correlation in losses on end-to-end paths measured by receivers to infer loss rates on each link.

Compared with methods using multicast probes, unicast-based inference methods are more flexible and widely applicable, and thus, are of practical importance. For example, such methods are applicable to networks where multicast communication is not available or path-topologies are not limited to trees. Furthermore, they can be combined with passive measurement (monitoring) of real traffic generated by unicast communication. Note that, since multicast-based methods cannot treat inverse tree path-topologies, a unicast-based method is essential to us.

Nonetheless, unicast-based methods have some pitfalls. The most significant problem is imperfect correla-
tion in concurrent events on paths. As mentioned later (Sec. 2), a basic inference method (based on correlation among observations of packets) requires the assumption that if a link is shared by a set of paths then the packets along the paths experience the same event on the shared link within one atomic trial. Unlike multicast-based methods, unicast-based methods cannot realize such perfect correlation, and thus the above assumption is not always true. The other problem is the bandwidth inefficiency. Indeed, unicast-based methods need to send more probe packets than multicast-based methods if applicable. The focus of this paper is mainly on the former problem on inference under imperfect correlation of concurrent event on paths.

There exist several related works. For tree path-topologies, recent researches propose some techniques for this problem [4]–[6]. In particular, the method in [5] is similar to ours. On the other hand, our method can treat not only tree but also inverse tree path-topologies, and their combinations. For inverse tree path-topologies, a recent research proposes how to infer whether or not a pair of flows experiencing congestion are congested at the same shared link [7]. On the other hand, our method infers more quantitative characteristics, i.e., packet loss rates of individual links.

In the remainder of this paper, we propose a technique to infer link loss rates on both trees (Sec. 2) and inverse trees (Sec. 3). We explain how to use it on general path-topologies (Sec. 4). And we show simulation results which indicate potential of our unicast-based inference (Sec. 5).

2. Tree Path-Topology

2.1 Basic Model and Problem

Let us consider a single-level binary tree ((I) of Fig. 1). We denote the path from node 0 to 1 by $a$, and the path from 0 to 2 by $b$. We label each link by paths including the link (e.g., $l_a$, $l_b$, and $l_{ab}$).

We regard an event that a probe packet successfully passes through a link as an occurrence of “no loss” on the link. Let $x_R$ ($R \in \{a, b, ab\}$) denote the occurrence probability of “no loss” on link $l_R$, which is one minus the loss rate of $l_R$. Our goal is to determine $x_R$ from end-to-end measurement using unicast probes, i.e., observations of unicast probes at the receivers’ nodes. We assume that each $x_R$ is not equal to 0 (i.e., positive).

Consider that unicast probe packets can be sent from node 0 to nodes 1 and 2 along paths $a$ and $b$, respectively. Let $P_r$ ($r \in \{a, b\}$) denote a probe packet on path $r$. Let $y_r$ be the occurrence probabilities of “no loss” on path $r$, and $y_{ab}$ be the occurrence probability of “no loss” on both path $a$ and $b$ concurrently.

To obtain those occurrence probabilities, we dispatch a series of trials where each trial consists of sending $P_a$ and $P_b$, and the trials can be regarded as independent of each other. We denote the number of all trials by $N$, the number of trials in which $P_r$ reaches the destination (i.e., “no loss” occurs on path $r$) by $N_r$, and the number of trials in which both $P_a$ and $P_b$ reach the destinations by $N_{ab}$. Then, by “Law of large numbers,” we can estimate $y_r$ as $N_r/N$ and $y_{ab}$ as $N_{ab}/N$, respectively.

Then if $P_a$ and $P_b$ in a trial are perfectly correlated on the shared link $l_{ab}$, we have simple equations:

$$y_a = x_{ab}x_a, \quad y_b = x_{ab}x_b, \quad y_{ab} = x_{ab}x_a x_b$$  \hspace{1cm} (1)

where, if $y_a$, $y_b$ and $y_{ab}$ are given, $x_{ab}$, $x_a$ and $x_b$ are uniquely determined as follows.

$$x_a = y_{ab}/y_b, \quad x_b = y_{ab}/y_a, \quad x_{ab} = y_a y_b / y_{ab}$$  \hspace{1cm} (2)

However, the following problems arise in this naive inference:

(∗) Concurrent events on paths—In (1), we assume that $P_a$ and $P_b$ experience the same event on the shared link $l_{ab}$ within one atomic trial (observation). Nevertheless, even when $P_a$ passes through $l_{ab}$, $P_b$ may be dropped on $l_{ab}$, and vice versa. In general, the above assumption is not always true in unicast-based inference methods.

(**) Correlation among links—This problem is common for both unicast-based and multicast-based inference methods. In (1), we assume independence of losses among links. In actual networks, however, states of links caused by background traffic flows are not usually independent of each other.

For the latter issue (**), note that if the correlation among sibling links is very weak, the error will be negligible. Furthermore, if we estimate the degree of correlation in some way, we can correct the inference error. Some analytical and simulation results regarding losses on tree path-topologies have been presented in [2] and [3]. In actual networks, we can expect such correlations are non-negative and small. Moreover, it can be shown that, for each intermediate link (i.e., each link except for the root and the leaf links in a tree), the effect of correlation among sibling links on the inference error of the link is counterbalanced by the effect of correlation among child links on it (e.g., [2]). Therefore, inference errors of intermediate links are expected to be small.
2.2 Inference Method Using Unicast Probes

We continue to consider a single-level binary tree. Let $M_{ab}$ be a series of trials where each trial sends an ordered pair $(P_a, P_b)$ of probes, in which sending $P_b$ follows immediately sending $P_a$. Let $M_b$ be a series of trials where each trial sends probe $P_b$. We perform both $M_{ab}$ and $M_b$ in which trials can be regarded as independent of each other, and observe arrivals of the probes at the destination nodes.

Let $Pr[X]$ denote an occurrence probability of event $X$, and $X^c$ denote a co-event of event $X$. For event $X$ which can be defined in both $M_{ab}$ and $M_b$, let $X_D$ denote the event in measurement $D (D \in \{a, b\})$. Let $X_{R,r}$ denote the event that probe $P_r (r \in \{a, b\})$ passes through link $l_R$ successfully, i.e., no loss occurs on link $l_R$ in probe $P_r$. Let $V_r$ denote the event that the $r$-th probe entering the shared link $l_{ab}$ in a trial passes through the link.

For each trial in $M_{ab}$, the event occurring on the shared link $l_{ab}$ is one of the followings: (e1) $X_{ab,a} \cap X_{ab,b} = V_1 \cap V_2$, (e2) $X_{ab,a} \cap X_{ab,b}^c = V_1 \cap (V_2)^c$, (e3) $X_{ab,a}^c \cap X_{ab,b} = (V_1)^c \cap V_2$, (e4) $X_{ab,a}^c \cap X_{ab,b}^c = (V_1)^c \cap (V_2)^c$, where (e2) means that $P_a$ passes but $P_b$ is dropped, and (e3) means that $P_a$ is dropped but $P_b$ passes. The occurrence of (e2) and (e3) causes the imperfect correlation in concurrent events on paths. If the first packet entering a link is dropped then the second packet entering the link immediately after the first one is likely to be dropped because of the nature of a FIFO queue. Thus, we assume later that the conditional probability of $(V_1)^c$ given that $V_2$ occurs is small. This assumption implies (e3) can be negligible, and is essential to our technique. On the other hand, the occurrence of (e2) cannot be negligible.

We define the following unknown probabilities in $M_{ab}$ and $M_b$. Our goal is to determine $x_a$, $x_b$, and $x_{ab}$ from measurable probabilities.

$$x_{ab}^* \overset{\text{def}}{=} Pr[V_1(ab) \cap V_2(ab)]$$

$$x_a \overset{\text{def}}{=} Pr[V_1(b)] = Pr[V_1(ab)]$$

$$x_b \overset{\text{def}}{=} Pr[X_{a,a}^c|V_1(ab)]$$

$$x_b^* \overset{\text{def}}{=} Pr[X_{b,b}^c|V_1(ab) \cap V_2(ab)] = Pr[X_{b,b}^c|V_2(ab)]$$

where we introduce an approximation $Pr[V_1(ab)] = Pr[V_1(ab)]$, which means that loss of the first probe packet in a trial on the shared link $l_{ab}$ is independent of the destination of the probe. This approximation allows us to regard $x_{ab}$ as the “general” no loss rate of probes entering link $l_{ab}$ without interference from other probes. Similarly, $x_r$ can be regarded as the general no loss rate of probes entering $l_r$ given that the probes have passed through the previous link $l_{ab}$ without interference. On the other hand, $x_b^*$ is regarded as the no loss rate of probes entering $l_b$ given that the probes have passed through $l_{ab}$ in spite of interference from preceding probes, where we introduce an approximation $Pr[X_{b,b}^c|V_1(ab) \cap V_2(ab)] = Pr[X_{b,b}^c|V_2(ab)]$.

Let $Y_r$ be the event that probe $P_r$ reaches its destination successfully. Then we can define the following probabilities, which can be obtained from observations of $M_b$ and $M_{ab}$.

$$y_b(b) \overset{\text{def}}{=} Pr[Y_b(b)] = Pr[X_{b,b}^c \cap V_1(ab)]$$

$$y_a \overset{\text{def}}{=} Pr[Y_a(ab)] = Pr[X_{a,a}^c \cap V_1(ab)]$$

$$y_b \overset{\text{def}}{=} Pr[Y_b(ab)] = Pr[X_{b,b}^c \cap V_2(ab)]$$

$$y_{ab} \overset{\text{def}}{=} Pr[Y_a(ab) \cap Y_b(ab)]$$

$$= Pr[X_{a,a}^c \cap X_{b,b}^c \cap V_1(ab) \cap V_2(ab)]$$

Furthermore, we define $\delta$ as a degree of correlation between sibling links $l_a$ and $l_b$. In what follows, we regard $\delta$ as a given small non-negative constant value, which depends on the network considered here.

$$Pr[X_{a,a}^c \cap X_{b,b}^c | V_1(ab) \cap V_2(ab)] = x_a x_b^*(1+\delta)$$

Then we have the relation between $x$ and $y$: $y_b(b) = x_a x_b, y_a = x_a x_b x_b^*(1-\varepsilon), y_{ab} = x_a x_b x_b^*(1+\delta)$ with six unknown variables $x_a$, $x_b$, $x_b^*$, $x_{ab}$ and $\varepsilon$. Note that, if we regard $\varepsilon$ as a known value (parameter), $x_a$, $x_b$ and $x_{ab}$ are uniquely solved as follows, although $x_{ab}^*$ and $x_b^*$ still cannot be determined.

$$x_a = \frac{y_{ab}}{y_b(b) (1-\varepsilon)(1+\delta)}, x_b = \frac{y_b(b) y_{ab}}{y_a y_b(1-\varepsilon)(1+\delta)}$$

$$x_{ab} = \frac{y_a y_b(1-\varepsilon)(1+\delta)}{y_{ab}}$$

Since we assume both $\varepsilon$ and $\delta$ are small, we can obtain approximate inference values ($\hat{x}_a$, $\hat{x}_b$ and $\hat{x}_{ab}$) by letting $\varepsilon = 0$ and $\delta = 0$ as follows.

$$\hat{x}_a \overset{\text{def}}{=} \frac{y_{ab}}{y_b(b)} \cdot \hat{x}_b \overset{\text{def}}{=} \frac{y_b(b) y_{ab}}{y_a y_b} \cdot \hat{x}_{ab} \overset{\text{def}}{=} \frac{y_a y_b}{y_{ab}}$$

We estimate errors between $x$ in (3) and $\hat{x}$ in (4) as follows ($r \in \{a, b\}$):

$$\frac{\hat{x}_{ab}}{x_{ab}} = \frac{1}{(1-\varepsilon)(1+\delta)}, \frac{\hat{x}_r}{x_r} = (1-\varepsilon)(1+\delta)$$
where $\varepsilon$ and $\delta$ are counterbalanced because both of them are non-negative.

The above idea can be extended to multi-level tree path-topologies. For example, we consider a two-level binary tree ((II) of Fig. 1). Unicast probe packets can be sent from node 0 to nodes 1, 2, 3 and 4 along paths $a$, $b$, $c$ and $d$, respectively. In measurement $M_D$, let $P_r(D)$ denote probe $P_r$ along path $r$, and $y_r(D)$ denote probability $y_r$ ($r \in \{a, b\}$).

We consider four independent measurements, $M_{a,b}$, $M_{a,c}$, $M_{c,d}$ and $M_d$, where $M_{r,r'} ((r, r') \in \{(a, b), (b, c), (c, d), (d, d)\})$ consists of a series of ordered pairs $(P_r(r, r'), P_{r'}(r, r'))$ of probes, and $M_d$ consists of a series of probes $P_d(D)$. We perform these measurements independently, and then estimate probabilities $y_{ab}$, $y_{bc}$, $y_{cd}$, and $y_d$.

According to the above method for single-level binary trees, we can obtain $x_{abcd}$ and $x_a$ from $M_{a,b}$, $x_{abcd}$ and $x_{ab}$ from $M_{a,c}$, $x_{abcd}$ and $x_c$ from $M_{c,d}$, and $x_{abcd}$ and $x_{ab}$ from $M_d$. Combining them, we can infer $x_{ab}$, $x_b$, $x_c$, $x_d$, $x_{ab}$, and $x_{abcd}$. Note that redundant information may be used for detecting unexpected correlation among trials which should be independent. For example, we show $\hat{x}_{ab}$ (an inferred value of $x_{ab}$) and its error estimation:

$$\frac{\hat{x}_{ab}}{x_{ab}} = \frac{1 - \varepsilon_{abcd} + \varepsilon_{abcd}}{1 - \varepsilon_{ab}} = 1 + \varepsilon_{ab}$$

where $\varepsilon_{R,R'}$ is a degree of correlation between links $l_R$ and $l_{R'}$, $\varepsilon_R$ is $\Pr[V_1 | V_2]$ on a shared link $l_R$, and $\rho$ is $\Pr[V_2 | V_1]$ on $l_{ab} = \{x_{ab}, x_{abcd}\}$. Note that $\rho \leq 1$ and $\rho$ is expected to be close to 1. If we assume that $\delta_{a,b} = \delta_{ab,cd}$ and $\delta_{b,c} = 0$ (because $l_b$ and $l_c$ are not directly connected to the same link along paths), then we have:

$$\frac{\hat{x}_{ab}}{x_{ab}} = 1 - \varepsilon_{abcd} + \varepsilon_{abcd} = 1 + \varepsilon_{ab}$$

Unlike $\delta$, the effect of $\varepsilon$ on inference errors for intermediate links are still first order.

3. Inverse Tree Path-Topology

We first consider a single-level inverse binary tree ((I) of Fig. 2). Unicast probe packets can be sent from nodes 1 and 2 to 0 along paths $a$ and $b$, respectively.

Let $M_{a,b}$ be a series of trials where each trial sends an unordered pair $(P_a, P_b)$ of probes. In $M_{a,b}$, $P_a$ and $P_b$ should be sent so that they are likely to enter the shared link $l_{ab}$ closely in a trial. Let $M_r$ be a series of trials where each trial sends probe $P_r$ ($r \in \{a, b\}$). We perform $M_{a,b}$, $M_a$ and $M_b$ in which trials can be regarded as independent of each other, and observe arrivals of the probes at the destination node.

In addition to the notations previously defined, we define some events related to probes entering the shared link. Let $H_r$ denote the event that probe $P_r$ enters $l_{ab}$ first in a trial in $M_{a,b}$. Let $S_{ab}$ denote the event that probes $P_a$ and $P_b$ enter $l_{ab}$ so closely in a trial in $M_{a,b}$ that the second probe is interfered, and thus more liable to be dropped on $l_{ab}$ than the first one. For convenience, we also define $W_{ab}$ as $X_{a,a} \cap X_{b,b}$.

We define the following unknown probabilities in $M_{a,b}$, $M_a$ and $M_b$. Our goal is to determine $x_a$, $x_b$ and $x_{ab}$ from measurable probabilities.

$$h_a = \Pr[H_a \cap S_{ab} | W_{ab}]$$

$$h_b = \Pr[H_b \cap S_{ab} | W_{ab}]$$

$$x_{a,b} = \Pr[V_1(x_{a,b}) \cap V_2(x_{a,b}) | H_a \cap S_{ab} \cap W_{ab}]$$

$$\varepsilon_{a} = \Pr[(V_1(x_{a,b}) \cap V_2(x_{a,b}) | H_a \cap S_{ab} \cap W_{ab}]$$

$$\varepsilon_{a} = \Pr[(V_1(x_{a,b}) \cap V_2(x_{a,b}) | H_a \cap S_{ab} \cap W_{ab}]$$

$$x_a = \Pr[V_1(x_{a,b}) | W_{ab}]$$

$$x_b = \Pr[V_1(x_{a,b}) | W_{ab}]$$

where we introduce approximations $\Pr[V_1(x_{a,b}) | W_{ab}] = \Pr[V_1(x_{a,b}) | X_{a,a}] = \Pr[V_1(x_{a,b}) | X_{b,b}]$.
probes have passed the previous link successfully. Similarly, $x_r$ can be regarded as the general no loss rate of probes entering $l_r$.

We also define the following probabilities which can be obtained from observations of $M_a$, $M_b$ and $M_{ab}$. For example, $y_{ab}^a$ (resp. $y_{ab}^b$) is the probability that two probes in a trial in $M_{ab}$ enter the shared link $l_{ab}$ closely (resp. separately) and then both of them reach their common destination. (Note: hereafter, we use “resp.” as an abbreviation for “respectively.”)

$$
y_{a}^{(a)} \equiv \Pr [Y_a^{(a)}] = \Pr [V_1^{(a)} \cap X_{a,a}^{(a)}]
$$

$$
y_{b}^{(b)} \equiv \Pr [Y_b^{(b)}] = \Pr [V_1^{(b)} \cap X_{b,b}^{(b)}]
$$

$$
y_{a}^{(ab)} \equiv \Pr [Y_a^{(ab)}] = \Pr [X_{ab,a}^{(ab)} \cap X_{a,a}^{(ab)}]
$$

$$
y_{b}^{(ab)} \equiv \Pr [Y_b^{(ab)}] = \Pr [X_{ab,b}^{(ab)} \cap X_{b,b}^{(ab)}]
$$

$$
y_{ab}^* \equiv \Pr [V_1^{(ab)} \cap V_2^{(ab)} \cap S_{ab} \cap W_{ab}]
$$

$$
y_{ab}' \equiv \Pr [V_1^{(ab)} \cap V_2^{(ab)} \cap (S_{ab})^c \cap W_{ab}]
$$

where we assume that, in measurement $M_{ab}$, we can detect whether event $S_{ab}$ occurs or not by observing arrival probes and other traffic (if needed) pass through $l_{ab}$ at the destination node 0. In other words, we can determine whether two probes in a trial reaching the destination have passed through $l_{ab}$ very closely or not. This assumption allows us to estimate $y_{ab}^*$ and $y_{ab}'$ from end-to-end observations.

Furthermore, we define the degree of correlation between sibling links $l_a$ and $l_b$ as a parameter:

$$
\Pr [W_{ab}] = x_a x_b (1 + \delta)
$$

Then we have the relation between $x$ and $y$:

$$
y_{a} = x_a (1 - x_b (1 + \delta)) x_{ab}
$$

$$
y_{a} = (1 + x_a x_b (1 + \delta)) (h_a x_{ab} + h_b x_{ab,b}) (1 - \varepsilon_b)
$$

$$
= x_a x_b (1 + \delta) (1 - h_a) x_{ab}
$$

$$
y_{a} = x_a x_b (1 + \delta) (h_a x_{ab,a} + h_b x_{ab,b})
$$

$$
y_{a} = x_a x_b (1 + \delta) (1 - h_a - h_b) x_{ab}^2
$$

$$
y_{b} = y_{x_a x_b - x_a x_b (1 + \delta) x_{ab}}
$$

$$
y_{b} = x_b x_{ab} (1 + \delta) (1 - h_a - h_b) x_{ab}^2
$$

$$
y_{b} = x_a x_b (1 + \delta) (h_a x_{ab,a} + h_b x_{ab,b})
$$

$$
y_{b} = x_a x_b (1 + \delta) (1 - h_a - h_b) x_{ab}^2
$$

To solve $x_a$, $x_b$ and $x_{ab}$, we define the following:

$$
h \equiv h_a + h_b,
$$

$$
\rho_r \equiv \frac{x_{ab}^*}{x_{ab,r} x_{ab}} (r \in \{a, b\}),
$$

$$
\varepsilon \equiv \frac{1}{h} \left( \rho_a \varepsilon_a h_a + \rho_b \varepsilon_b h_b \right)
$$

$$
z_1 \equiv (y_{a}^{(a)} - (1 + \delta)) x_{ab} (1 - \varepsilon)
$$

$$
z_2 = y_{a}^{(a)} y_{b}^{(b)} - y_{ab}'
$$

Consequently we have the following equations.

$$
y_{a}^a = x_a x_{ab},
$$

$$
y_{b}^b = x_b x_{ab},
$$

$$
z_1 = x_a x_b (1 + \delta) x_{ab} (1 - \varepsilon) h,
$$

$$
z_2 = x_a x_b (1 + \delta) x_{ab}^2 h - x_a x_b \delta x_{ab}^2
$$

Since $\rho_a, \rho_b \leq 1$, $\varepsilon$ can be estimated as $\varepsilon \leq \max(\varepsilon_a / (1 - \varepsilon_a), \varepsilon_b / (1 - \varepsilon_b))$. Similarly to the previous case of tree path-topologies, if we regard $\varepsilon$ as a parameter, then $x_a$, $x_b$ and $x_{ab}$ are uniquely solved: $x_{ab} = (y_{a}^{(a)} y_{b}^{(b)} (1 - \varepsilon) / (1 - \varepsilon_a), x_a = y_{a}^{(a)} / (y_{a}^{(a)} + (y_{b}^{(b)} + (y_{b}^{(b)} + y_{a}^{(a)}) (1 - \varepsilon)))$, and $x_b = y_{b}^{(b)} (1 - \varepsilon) / (y_{b}^{(b)} + y_{a}^{(a)}) (1 - \varepsilon))$. Since we assume $\varepsilon_a, \varepsilon_b$ and $\delta$ are small, we can obtain approximate inference values ($\hat{x}_a, \hat{x}_b$ and $\hat{x}_{ab}$) by letting $\varepsilon = 0$ and $\delta = 0$ as follows.

$$
\hat{x}_a \equiv \frac{y_{a}^{(a)} z_1}{z_2} = \frac{(y_{a}^{(a)} - y_{a}) + (y_{b}^{(b)} - y_{b}) + y_{ab}^*}{y_{b}^{(b)} - y_{ab}/y_{a}^{(a)}},
$$

$$
\hat{x}_b \equiv \frac{y_{b}^{(b)} z_1}{z_2} = \frac{(y_{b}^{(b)} - y_{b}) + (y_{a}^{(a)} - y_{a}) + y_{ab}^*}{y_{a}^{(a)} - y_{ab}/y_{b}^{(b)}},
$$

$$
\hat{x}_{ab} \equiv \frac{z_2}{z_1} = \frac{y_{a}^{(a)} y_{b}^{(b)} - y_{ab}'}{(y_{a}^{(a)} - y_{a}) + (y_{b}^{(b)} - y_{b}) + y_{ab}^*}
$$

We estimate errors between $\hat{x}$ in (5) and $\hat{x}$ in (6) from $z_2 \leq y_{a}^{(a)} y_{b}^{(b)} \leq 1$:

$$
\frac{1}{(1 + \delta) z_2 (1 - \varepsilon)} \leq \frac{\hat{x}_{ab}}{x_{ab}} \leq \frac{1}{(1 + \delta) (1 - \varepsilon)}
$$

$$
(1 + \delta) (1 - \varepsilon) \leq \frac{\hat{x}_r}{x_r} \leq \frac{1}{(1 + \delta) z_2 (1 - \varepsilon)}
$$

which indicate that errors may increase in inverse proportion to $z_2$ (i.e., $y_{a}^{(a)} y_{b}^{(b)} - y_{ab}')$. However, in an adequate measurement in which $S_{ab}$ is likely to occur, $y_{ab}'$ and (thus) errors are expected to be small.

The above idea can be extended to multi-level inverse tree path-topologies. For example, we consider a two-level inverse tree (II of Fig. 2). Unicast probe packets can be sent from node 1, 2, 3 and 4 to 0 along
dependent measurements, we can infer:

\[
\hat{x}_{ab} = \frac{z_2^{(ab)} (z_2^{(bc)})}{(z_2^{(ab)} z_2^{(bc)})}
\]

\[
z_1^{(ab)} = (y_a^{(a)} - y_a^{(b)}) + (y_b^{(b)} - y_b^{(ab)}) + y_{ab}
\]

\[
z_2^{(ab)} = y_a^{(a)} - y_a^{(b)}
\]

\[
z_1^{(bc)} = (y_b^{(b)} - y_b^{(bc)}) + (y_c^{(c)} - y_c^{(bc)}) + y_{bc}
\]

\[
z_2^{(bc)} = y_b^{(b)} - y_b^{(bc)}
\]

\[
\hat{\varepsilon}_a = \frac{x_{abcd}/x_{ab}}{1 - x_{abcd}/x_{ab}} = \frac{x_{abc}/x_{ab}}{1 - x_{abc}/x_{ab}}
\]

\[
\hat{\varepsilon}_b = \frac{(y_a^{(a)} - y_a^{(b)}) + (y_b^{(b)} - y_b^{(ab)}) + y_{ab}}{(y_a^{(a)} - y_a^{(bc)}) + (y_b^{(b)} - y_b^{(bc)}) + y_{bc}}
\]

In what follows, we explain how this inference works in general, based on the principle and notations in [1]. First, we consider a set of measurable paths, and assume each (directed) link is uniquely identified by the set of paths including the link, so that we can label each link by paths including the link. We denote all labels for links by \(\Delta \subset 2^\text{Path}\), thus the set of all labels are denoted by \(\{R \mid R \in \Delta\}\). For conciseness, we use \(abc\) instead of \(\{a, b, c\}\) as an expression of set \(R \subset \text{Path}\) consisting of paths \(a, b, c\) for example.

Let \(L_r\) be a set of links included by path \(r \in \text{Path}\). Thus, \(\{R \mid R \subset L_r\}\) is the “no loss” rate of link \(l_R (R \in \Delta)\), and also assume each \(x_R\) is positive (non-zero).

Then it can be shown that \(\{x_R \mid R \in \Delta\}\) is uniquely determined by \(A(\mathbf{L}(R)) = \sum_{R \in \Delta} x_R\) under the above assumptions. Therefore, our goal is to infer \(A(\mathbf{L}(R))\) for each link \(l_R\) from end-to-end path measurement on a

4. General Path-Topology

In the previous sections, we show how to infer loss rates of individual links on a “general” tree (resp. inverse tree) by combining inference of loss rates on some single-level binary trees (resp. inverse trees).

This can be extended to more general path-topologies. For example, in (I) of Fig.3, we assume unicast probe packets can be sent from node 0 to 2, 1 to 2 and 1 to 3 along paths \(a, b, c\) and \(a, b, c, d\), respectively.

Consider a tree traversed by \(P_a\) and \(P_b\), and an inverse tree traversed by \(P_a\) and \(P_b\), and four independent measurements on them, i.e., \(M_a, M_ab, M_b, M_bc\). Combining both inference for a tree and an inverse tree, we can infer \(x_a, x_b, x_c, x_d, x_{abcd}\) and \(x_{abcd}\). For example, \(x_b\) is inferred as follows:

\[
\hat{x}_b = \frac{(y_a^{(a)} - y_a^{(ab)}) + (y_b^{(b)} - y_b^{(ab)}) + y_{ab}}{(y_a^{(a)} - y_a^{(bc)}) + (y_b^{(b)} - y_b^{(bc)}) + y_{bc}}
\]

Fig. 3 General path-topologies.
general path-topology. If \( \mathbf{L}(R) = \mathbf{L}_a \) for path \( \exists a \in R \), then \( A(\mathbf{L}(R)) = A(\mathbf{L}_a) = y_a \) is simply obtained from measurement of path \( a \).

Otherwise, we consider whether every pair of paths in \( R \) has a shared link besides (outside) \( \mathbf{L}(R) \) or not. The cases in which all paths in \( R \) are mutually overlapped outside \( \mathbf{L}(R) \), like \( a \), \( b \), \( c \) and \( d \) for \( l_{abcd} \) in (II) of Fig. 3, are unusual in actual networks. Thus, we can assume, for each \( l_R \), there exists (at least) one pair of paths in \( R \) having no shared link outside \( \mathbf{L}(R) \). For such a pair \( (a, b) \), it can be shown that \( \mathbf{L}(R) = \mathbf{L}(ab) \) and \( (\mathbf{L}_a - \mathbf{L}(ab)) \cap (\mathbf{L}_b - \mathbf{L}(ab)) = \emptyset \). For example, for \( l_{abcd} \) in (II) of Fig. 3, we see \( \mathbf{L}(ab) = \{l_{abcd}, l_{abc} \} = \mathbf{L}(ab) \), \( \mathbf{L}_a - \mathbf{L}(ab) = \{l_a, l_{acd} \} \) and \( \mathbf{L}_b - \mathbf{L}(ab) = \{l_b, l_{bcd} \} \). Therefore, what we should do is to obtain \( A(\mathbf{L}(ab)) \) for an appropriate pair \( (a, b) \) of paths in \( R \).

Let us consider topological relations among \( \mathbf{L}_a \), \( \mathbf{L}_b \) and \( \mathbf{L}(ab) \). In general, there exist four cases: (A), (B), (C), (D) of Fig. 4. (A) can hardly appear in most of actual networks with usual (link-cost based) routing schemes. (B) (resp. (C)) is a single-level binary tree (resp. inverse tree), in which \( A(\mathbf{L}(ab)) \) can be inferred as mentioned in the previous sections.

Finally, in case (D), it is difficult to detect accurately whether the two probes enter the shared part \( \mathbf{L}(ab) \) “closely” or not in end-to-end measurement of paths \( a \) and \( b \). This also makes it difficult to infer \( A(\mathbf{L}(ab)) \) directly in general. However, if there exists a set \( L \) of links including \( \mathbf{L}(ab) \) satisfying that both \( A(L) \) and \( A(L - \mathbf{L}(ab)) \) are inferred, then \( A(\mathbf{L}(ab)) \) can be obtained as \( A(L)/A(L - \mathbf{L}(ab)) \). For example, for \( l_{abc} \), in (III) of Fig. 3, we can infer \( x_{ac} \) from measurements \( M_{bc} \) on a tree, and \( x_{ac} \) from \( (M_a, M_c, M_{ac}) \) on an inverse tree, and consequently \( x_{ab} \) can be obtained.

5. Simulation

We examine four path-topologies shown in Fig. 1 and Fig. 2 by using the network simulator ns [8]. Each probe uses a 64-byte packet (ICMP echo request). The bandwidth of each link is 1.5 Mbps with 10 ms of propagation delay, and a FIFO queue with 6-packet capacity. We generate the background traffic by 1500-byte packet TCP flows on each link, between edge nodes of the link, with an infinite data source. The direction of the TCP flow is same as the flow of the probes on the link. The number of TCP flows on links \( l_a, l_c, l_{ab} \) and \( l_{abcd} \) (resp. on \( l_b, l_d \) and \( l_{cd} \)) is one (resp. two). These background flows make queues of some links full, causing packet losses on the links.

Let \( \{M_1, M_2, ..., M_n\} \) be a set of measurements needed for an inference scenario, and \( T_j \) be the \( i \)-th trial in \( M_j \). In this simulation, we choose a simple configuration of measurement. We just dispatch each trial in \( M_1, ..., M_n \) in turn in order to perform these measurements independently. Time intervals between executing adjacent trials (i.e., \( T_j \) and \( T_{j+1} \), or \( T_n \) and \( T_{n+1} \)) change randomly within some range. The range of time intervals we choose are \([8,24]\) msec for Fig. 1, \([50,80]\) msec for (I) of Fig. 2, and \([70,100]\) msec for (II) of Fig. 2, respectively.

The mean value and the minimum one of the time interval between adjacent trials are significant. To avoid change of network states (e.g., routes of paths), it is important to complete the whole measurement in an adequate term. Thus, since inference requires a number of trials, short time intervals between trials are preferable. On the other hand, to avoid unexpected correlation in different trials, time intervals should not be so short. Especially for inverse tree path-topologies, short intervals may cause two probes in different trials to enter a shared link closely, which can thus make one.
probe interfere with another. As the minimum value of the time intervals, we employ 8 msec (the transmission time of one TCP packet) for tree, 50 msec (greater than 8 msec × 6 packet) for single-level inverse tree, and 70 msec for two-level inverse tree path-topologies, respectively.

Time intervals between sending two probes in a trial are fixed values. For measurement $M_{r,r'}$, the value is so short that $P_r$ and $P_{r'}$ are sent immediately after $P_r$ ($r, r' \in \{a, b, c, d\}$). For $M_{rr'}$, the value is chosen so that probes $P_r$ and $P_{r'}$ are likely to enter a shared link closely.

In $M_{rr'}$, we expect that if $S_{rr'}$ occurs (i.e., two probes enter a shared link closely) and none of them are dropped on the link, then the inter-arrival time at the destination node 0 between the two probes is likely to be less than some threshold value. We choose 24 msec as this value.

For inverse tree path-topologies, we also examine the cases on high bandwidth links with a large number of TCP flows starting randomly. The bandwidth of each link is 150 Mbps with 1 ms of propagation delay. As the number of TCP flows, we choose 60 for $l_{ab}$, 80 for $l_a$, and 110 for $l_b$ in (I) of Fig. 2; 13 for $l_a$, $l_c$ and $l_{ab}$, 17 for $l_{abcd}$, and 25 for $l_b$, $l_d$ and $l_{cd}$ in (II) of Fig. 2, respectively.

In Fig. 5, Fig. 6 and Fig. 7, the upper (resp. lower) shows comparison between the inferred loss rates and the real loss rates of probes on some links in an example of the single-level (resp. two-level) tree or inverse tree path-topology. On tree path-topologies in Fig. 5, inference seems quite stable and accurate, although there exists a certain bias in some cases. On the other hand, on inverse tree path-topologies in Fig. 6, we see inaccuracy and slow convergence with instability. In Fig. 7, however, we can see better accuracy and stability in high bandwidth networks (with many background TCP flows).

In general, the above errors are mainly due to 1) correlation between links in a trial, 2) correlation between trials in different measurements, and 3) measurement procedure itself. Both 1) and 2) come from the interaction among probes and background TCP flows exhibiting non-smooth behaviors. As examples of 3), in measurement $M_{ab}$, $M_{bc}$ and $M_{cd}$, we sometimes find that two probes in a trial do not enter a shared link so closely, and that it is not so accurate to detect whether the two probes enter the shared link closely or not.

Therefore, adequate control of the timing of dispatching trials and adequate criteria to detect interference between two received probes are required for accurate inference. To improve them, we may need to introduce more randomness and adaptability (e.g., feedback of probes’ arrival time information from receivers to senders) in control of measurement.
6. Concluding Remarks

In this paper, we have presented a method of inferring packet loss rates of individual links from end-to-end unicast probe measurement, which is applicable to various path-topologies including trees, inverse trees and their combinations. Simulation results have indicated potential of our method.

To establish a reliable inference method which can be widely usable in the Internet, we are going to examine our method in more complex scenarios and clarify the reliability and limitations in practical use.

References


Masato Tsuru received B.E. and M.E. degrees from Kyoto University, Kyoto, Japan in 1983 and 1985, respectively. From 1985 to 1990, he worked at Oki Electric Industry Co., Ltd., Tokyo. From 1990 to 2000, he was with the Information Science Center, Nagasaki University, Nagasaki. Since April 2000, he has been working at Japan Telecom Information Service Co., Ltd., Kitakyushu, and the GENESIS project, Telecommunications Advancement Organization of Japan. His research interests include performance measurement, modeling and analysis of computer communication networks. He is a member of the IPSJ and ISSST.

Tetsuya Takine was born in Kyoto, Japan, on November 28, 1961. He received B.Eng., M.Eng. and Dr.Eng. degrees in applied mathematics and physics from Kyoto University, Kyoto, Japan, in 1984, 1986 and 1989, respectively. In April 1989, he joined the Department of Applied Mathematics and Physics, Faculty of Engineering, Kyoto University, as an Assistant Professor. Beginning in November 1991, he spent one year at the Department of Information and Computer Science, University of California, Irvine, on leave of absence from Kyoto University. In April 1994, he joined the Department of Information Systems Engineering, Faculty of Engineering, Osaka University as a Lecturer, and from December 1994 to March 1998, he was an Associate Professor in the same department. Since April 1998, he has been an Associate Professor in the Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University. His research interests include queueing theory and performance analysis of computer/communication systems. Dr. Takine is a member of the Operations Research Society of Japan, the Information Processing Society of Japan, the Institute of Systems, Control and Information Engineers, and IEEE. He received the 1996 Best Paper Award from the Operations Research Society of Japan.

Yuji Oie received B.E., M.E. and D.E. degrees from Kyoto University, Kyoto, Japan in 1978, 1980 and 1987, respectively. From 1980 to 1983, he worked at Nippon Denso Company Ltd., Kariya. From 1983 to 1990, he was with the Department of Electrical Engineering, Sasebo College of Technology, Sasebo. From 1990 to 1995, he was an Associate Professor in the Department of Computer Science and Electronics, Faculty of Computer Science and Systems Engineering, Kyushu Institute of Technology, Iizuka. From 1995 to 1997, he was a Professor in the Information Technology Center, Nara Institute of Science and Technology. Since April 1997, he has been a Professor in the Department of Computer Science and Electronics, Faculty of Computer Science and Systems Engineering, Kyushu Institute of Technology. His research interests include performance evaluation of computer communication networks, high speed networks, and queueing systems. He is a member of the IEEE and IPSJ.