Studies on Characteristics Measurement and Inference in Network Tomography

Masato TSURU

Doctor Thesis
submitted to Graduate School of Computer Science and Systems Engineering,
Kyushu Institute of Technology, in partial fullfillment of requirements
for the degree of DOCTOR

Thesis title:	Studies on Characteristics Measurement and Inference in Network Tomography
Author:	Masato TSURU
Supervisor:	Professor Yuji OIE
Submission Date:	December 10, 2001

Preface

As the Internet is shifting towards a social and economical infrastructure, it is requited to be operated in a reliable and efficient way, and thus should be measurable in terms of its characteristics. Nowadays, characteristics measurement in the Internet is an important issue for all participants, e.g., service providers, end-users, hardware or application developers, and researchers [1]. However, since the Internet is characterized by its huge scale, diversity, and distributed administration, it is sometime difficult to directly measure the dynamic states and performance of the network. Therefore, it is of practical importance to develop a statistical and indirect way to infer several characteristics that are expensive, or impossible in some cases, to be measured directly in a large network.

In this thesis, I have studied on characteristics measurement and inference in statistical and indirect ways categorized as the "network tomography", which means a scheme of statistically inferring some unobservable characteristics in a network by measuring other (easily observable) characteristics simultaneously, maybe at multiple points, and especially by using correlation among those measured characteristics. The goal of this thesis, in short, is to develop a general framework and its applications of the "inverse function" approach, which is one of major approaches to the network tomography, and to show the potential of this approach.

First of all, Chapter 1 briefly explains, as motivation of my research theme, why the network tomography has become important in characteristics measurements on a large network such as the Internet.

Chapter 2 introduces the inference problems regarded as the network tomography, and presents an overview for this area including my contributions. I give a rough definition of the network tomography, and address that the network tomography includes two typical forms, "inference of flow characteristics based on aggregated-flow measurements" and "inference of network-internal characteristics based on end-to-end path measurements".

In the former form, since what we want to know is a statistical perspective of global traffic, a "flow" is regarded as a series of (some kind of) packets between nodes, where nodes do not correspond to a single host but to a large set of hosts (i.e., a network or a set of networks). In the latter form, since what we want to know is statistical characteristics of an internal-portion along which traffic traverses, a "link" is regarded as not only a physical link but also a set of networks (a network cloud) on an end-to-end path.

I also address that two typical approaches, "calculating an inverse function from measurable parameters to target parameters" and "calculating an approximate solution of MLE (Maximum Likelihood Estimator) of target parameters with respect to observed data", have been employed in order to solve the above problems. Historically, first, the latter (the "MLE solver") approach was applied to the OD (origin-destination) traffic matrix inference problem, as the former form of the network tomography, by Vanderbe [2], Vardi [3] (in the middle of '90), and the succeeding studies. On the other hand, the former (the "inverse function") approach was applied to multicast-based inference of network-internal characteristics, as the latter form, by the MINC (Multicast-based Inference of Network-internal Characteristics) project ([4] and the succeeding studies from the end of '90).

Next, while the latter approach was applied to unicast-based inference of network-internal characteristics (as the latter form) by Coates and Nowak [5] and the succeeding studies, my study (as shown in Chapter 5) applied the former approach to a variant of the OD traffic matrix inference (as the former form), which was based on a generalization and extension of the former approach presented in Chapter 3. Furthermore, several studies have applied the former approach to unicast-based inference of network-internal characteristics. Among them, my study (as shown in Chapter 4) presented inference of link loss rates using the former approach with a certain extension that enables us to deal with inverse tree path topologies, while preceding studies dealt with only tree path topologies. Note that a combination of both approaches can be also seen (e.g., [6]).

Consequently, as similarities and differences between two forms of the inference problems are recognized, a common framework of the network tomography has been roughly established, which can give a unified viewpoint to various practical methods in the network tomography, and thus can give useful insights into advantages and disadvantages of individual inference methods.

Chapter 3, an extension of my paper [7], presents a principle of determining characteristics (i.e., occurrence probabilities of some states) of network-internal links from given characteristics of end-to-end paths with an arbitrary path topology, which can be regarded as a general framework for the "inverse function" approach. A whole measurement (including inference) process in the inverse function approach can be divided into two stages: 1) infer characteristics of end-to-end paths by monitoring traffic on the paths, and 2) determine characteristics of links from the given (by 1)) characteristics of the paths related to the links. I have focused on the latter part, and established a fundamental model and calculations for it, which is applicable to an arbitrary path-topology. Packet loss and queuing delay can be treated in this model, for example. As mentioned before, this generalization indicates that the "inverse function" approach is also applicable to the "inference of flow characteristics" based on aggregated-flow measurements, which is demonstrated in Chapter 5.

Chapter 4, an extension of my paper [8], presents a method, based on the "inverse function" approach with some extension, inferring packet loss rates on individual links from end-to-end measurement of unicast probe packets among several senders' and receivers' nodes. Suppose a set of observable paths covering all links whose characteristics should be inferred. Since these paths can be regarded as an appropriate combination of tree and inverse tree path-topologies under certain conditions, I have presented how to infer loss rates on each link (by using correlation among paths) both on trees and on inverse trees, which enable us to infer link loss rates on almost general path topologies.

Chapter 5, an extension of my paper [9], presents a method, based on the "inverse function" approach, inferring arrival rates of (some kind of) packets on individual flows from measurement of aggregated-flows at several links (e.g., routers' interfaces). Suppose a set of observable links being passed through by flows whose characteristics should be inferred. I have presented how to infer arrival rates on each flow (by using correlation among aggregated-flows at the links) on general aggregated-flow topologies. Although this method requires some condition on dynamics of arrivals due to the principle of the "inverse function" approach, it is applicable to general (irregular) distributions that cannot be captured by preceding methods based on an MLE of normal-based parametric models.

Finally, Chapter 6 concludes this thesis by clarifying the contributions of my studies. I also remark the future work to be done in this research area, which includes challenges to deployment in the real Internet, application to new problems, and novel methodologies that can deal with temporal and spatial dependence.

Acknowledgements

This thesis was not made without support and encouragement from several people. I must thank all of them on this occasion.

First of all, I shall be grateful to Yuji Oie, Professor of Kyushu Institute of Technology, as my supervisor and mentor, for guiding me to the research on the Internet technologies and patiently supporting through this research process. Without his guidance, my research has not been accomplished. I shall also be grateful to Professor Tetsuya Takine at Kyoto University, as the other mentor for me, for his careful and invaluable advice on my research.

I would like to thank Professor Kyoki Imamura, Professor Takaichi Yoshida, and Professor Kenji Kawahara, at Kyushu Institute of Technology, for their careful comments making this thesis better.

I am much indebted to Professor Hideo Miyahara and Professor Shinji Shimojo at Osaka University, Dr. Fumito Kubota and Dr. Shinichi Nakagawa at Communications Research Laboratory, Mr. Isao Kaneko at Kitakyushu GENESIS Research Center, and the rest of the members of the GENESIS project in the Telecommunications Advancement Organization of Japan (TAO), for fruitful discussions inspiring me.

I am also highly indebted to Tadanori Hongou, President of Japan Telecom Information Service Co., Ltd., for providing me this research opportunity. Special thanks to Mr. Hiroshi Yamamura at Japan Telecom Co., Ltd., for his kindly help when I started this research opportunity.

Many thanks to all colleagues and students in Oie's Laboratory and Kawahara's Laboratory, in particular, Assistant Professor Takeshi Ikenaga, for their kindly help and friendship.

Finally, my gratitude goes to my parents Daisuke and Eiko Tsuru, my wife Akiko Tsuru, and my daughters Yukiko, Mayuko, and Yuriko Tsuru, who all have encouraged me at all times.

Contents

List of Figures

List of Tables

Chapter 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 should be measurable in terms of its characteristics. As network users are becoming very interested in quality of services (QoS) provided by Internet Service Providers (ISP) as well as networks' connectivity, ISP should be aware of QoS they provide and show actual QoS by use of measured data. Therefore, nowadays, characteristics measurement in the Internet is an important issue for all participants, e.g., ISP, end-users, hardware or application developers, and researchers [1].

What to be known (measured) mainly consists of two types of characteristics: one is quality (such as performance) of end-to-end communications, and the other is network-internal statistics, which include local states of network-internal portions and global behaviors of traffic flows passing through those network-internal portions. Knowledge of such dynamic characteristics on a network is essential not only to reliable, efficient and QoS-aware network operations (in a static or a dynamic way), but also to development and researches for new network technologies.

How to know (measure) characteristics on a network is categorized in two types of measurements. One is sending controlled probe traffic along one or more paths (routes) on a target network and observing it at receivers, i.e., active methods. The other is collecting (capturing) real traffic or its statistics at one or more points on a target network and analyzing the data, i.e., passive methods. In typical (and simple) cases, the former can be used for direct measurement of quality of end-to-end communications, while the latter for direct measurement of network-internal statistics.

However, since the Internet is characterized by its huge scale, diversity, and distributed administration, such simple scenarios are not always feasible. Because of diversity, i.e., no typical configurations / behaviors in the Internet, it is difficult to predict needed characteristics without enough information, and thus, we need to know (measure) individual characteristics on the network in each case and each portion. On the other hand, because of a huge scale and distributed administration, it is expensive, or impossible in some cases, to directly measure such individual characteristics. For example, end-users cannot directly measure ISP's internal states, and even ISP cannot directly measure users' LAN or other ISP's internal states.

Therefore, it is of practical importance to develop a statistical and indirect way to infer several characteristics that are difficult to be measured directly in a large network. In general, there exist possibilities of indirect measurements in each combination of target characteristic types and measurement types (Table. 1.1).

Table 1.1: Direct and indirect measurements

	active methods	passive methods
end-to-end quality	Direct, Indirect	Direct, Indirect
local states	Indirect	Direct, Indirect
global traffic	(Indirect)	Direct, Indirect

In this thesis, I have focused on the "network tomography", which is raised in such inference problems. A term "network tomography", although not strictly defined, often means a scheme of statistically inferring some unobservable characteristics (or characteristics on a large fraction) in a network by measuring other, easily observable, characteristics (or characteristics on a small fraction) simultaneously, maybe at multiple points, and especially by using correlation among those measured characteristics.

The network tomography includes two typical forms, "inference of flow characteristics based on aggregated-flow measurements" and "inference of network-internal characteristics based on end-to-end path measurements". In the former form, since what we want to know is a statistical perspective of global traffic, a "flow" is regarded as a series of (some kind of) packets between nodes, where nodes do not correspond to a single host but to a large set of hosts (i.e., a network or a set of networks). In the latter form, since what we want to know is statistical characteristics of an internal-portion along which traffic traverses, a "link" is regarded as not only a physical link but also a set of networks (a network cloud) on an end-to-end path. I have studied on characteristics measurement and inference especially in the "inverse function" approach to both of the above forms in the network tomography.

The remainder of this thesis is organized as follows. Chapter 2 introduces the inference problems regarded as the network tomography, and presents an overview and a rough framework for this area. Chapter 3 presents a principle of determining characteristics of network-internal links from given characteristics of end-to-end paths with an arbitrary path topology, as a general framework for the "inverse function" approach. Chapter 4 presents a method, based on the "inverse function" approach with a certain extension, inferring packet loss rates on individual links from end-to-end measurement of unicast probe packets among several senders' and receivers' nodes. Chapter 5 presents a method, based on the "inverse function" approach, inferring arrival rates of (some kind of) packets on individual flows from measurement of aggregated-flows at several links (e.g., routers' interfaces). Finally, Chapter 6 concludes this thesis, by clarifying the contributions of my studies and remarking the future work to be done in this research area.

Chapter 2

Related researches on the Network Tomography

1 Introduction

As the Internet is shifting towards a social and economical infrastructure, it is requited to be operated in a reliable and efficient way, and thus should be measurable in terms of its characteristics. However, since the Internet is characterized by its huge scale, diversity, and distributed administration, it is sometime difficult to directly measure the dynamic states and performance of the network. Therefore, it is of practical importance to develop a statistical and indirect way to infer several characteristics that are expensive, or impossible in some cases, to be measured directly in a large network.

In this work, we present a brief introduction to the "network tomography", which is raised in such inference problems. A term "network tomography", although not strictly defined, often means a scheme of statistically inferring some unobservable characteristics (or characteristics on a large fraction) in a network by measuring other, easily observable, characteristics (or characteristics on a small fraction) simultaneously, maybe at multiple points, and especially by using correlation among those measured characteristics. Two typical forms of problems are known; "inference of flow characteristics based on aggregated-flow measurements" and "inference of network-internal characteristics based on end-to-end path measurements".

Although there already exist other survey papers for this area (e.g., [10], [11]), the viewpoint of this work is somewhat different from those of them. In [10], the authors introduced their proposed methodology using multicast active probing for inference of network-internal characteristics, while the authors in [11] emphasized their proposed approach of "approximately solving an maximum likelihood estimator" in both cases for inference of network-internal characteristics and for inference of flow characteristics. On the other hand, in this work, we intend to compare those two forms of problems in order to notice the similarities and the differences between them by applying two different approaches to toy examples in both forms, and to show that a common framework of the network tomography has been roughly established. This can give a unified viewpoint to various practical methods in the network tomography, and thus may

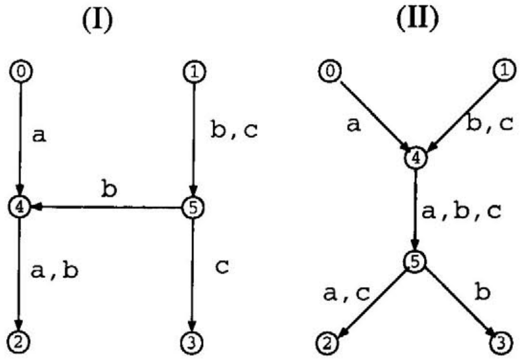

Figure 2.1: Path topologies

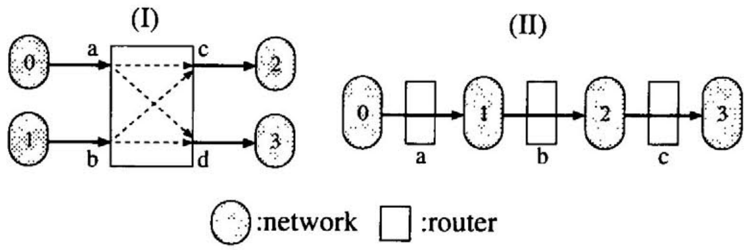

Figure 2.2: Aggregated-flow topologies

give useful insights into advantages and disadvantages of individual inference methods.

The remainder of this chapter is organized as follows. Section 2 and Section 3 describe two typical forms of the network tomography problems with simple examples, and two typical approaches to the problems, respectively. Finally, Section 4 presents a brief overview of recent researches and future work in this area.

2 Two forms of the tomography

Network tomography can be regarded as a statistical inverse problem in general. There exist two typical forms of the network tomography in recent researches.

One is inferring dynamic characteristics (statistical behavior of states) of each link in a network based on end-to-end measurement of traffic along paths. A "link" is regarded as any network-internal portion along which traffic traverses, which is distinguished by a set of observable paths passing through the link. Thus, it can be just a physical link, or a set of networks (a network cloud). This form is regarded as a case for indirect measurements of local states by active or passive methods. Several works have been presented to infer loss rates, queueing delay distributions and moments, bandwidths on individual links, or the existence of shared congestion in a network (See Section 4). In general cases, we consider a set of paths covering all links whose characteristics should be obtained, and try to infer some characteristics on those (unobservable) links from end-to-end measurements of traffic along those paths. For example, in Fig. 2.1, there exist three observable paths a, b, and c, and five unobservable links, i.e., l_a, l_b, l_c, l_{ab}, and l_{bc} in (I), and l_a, l_b, l_{ac}, l_{bc}, and l_{abc} in (II), respectively.

The other is inferring dynamic characteristics of each traffic flow based on measurement of aggregated-flow traffic at some links in a network. A "flow" is regarded as a series of (some kind of) packets from an origin node to a destination node, which is distinguished by a set of observable links being passed by the flow. Here we intend that a "node" does not correspond to a single host but to a large set of hosts (i.e., a network or a set of networks). This form is regarded as a case for indirect measurements of global traffic behaviors by passive methods. Several works have been presented to infer intensity of relative large traffic or occurrence rates of small traffic on individual flows. (See Section 4). In general cases, we consider a set of links (as observation points) being passed through by flows whose characteristics should be obtained. and try to infer some characteristics on those (unobservable) individual flows from measurements of traffic (aggregated-flows) passing those links. For example, in Fig. 2.2 (I), there exist four observation points a, b, c, and d, and four unobservable flows, i.e., f_{ac}, f_{bc}, f_{ad}, and f_{bd}, while in Fig. 2.2 (II), there exist three observation points a, b, and c, and six unobservable flows, i.e., f_a, f_b, f_c, f_{ab}, f_{bc}, and f_{abc}.

Knowledge of dynamic characteristics of individual network-internal portions is essential to network management (e.g., detection of failures/congestion, provisioning, and traffic engineering like QoS routing or server selections). However, because of a huge scale and distributed administration, it is expensive (sometime impossible) to measure such characteristics directly. Similarly, although knowledge of statistical perspective of global traffic flows, i.e., characteristics of individual traffic flows across a network, is also essential to network management (e.g., configuration, provisioning, traffic engineering, and detection of anomalous/malicious activities), it is also expensive (sometime impossible) to measure such characteristics of each flow separately. Therefore, both kind of inference problems are of practical importance.

Let us show five simple but intrinsic examples of the above inference problems.

2.1 Packet loss rates on links

Let us consider the most simple path topology Fig. 2.3 (a relation between links and observable paths passing through the links), and dispatch a number of multicast probe packets from root sender's node 0 to leaf receivers' nodes 1 and 2. Each link is labeled by the end-node 0, 1 or 2 connected to the link.

We regard an event that a probe is dropped at a link as an occurrence of the "loss" on the link. Let X_i denote a random variable that takes value 1 if the loss state occurs on link i, 0 otherwise.

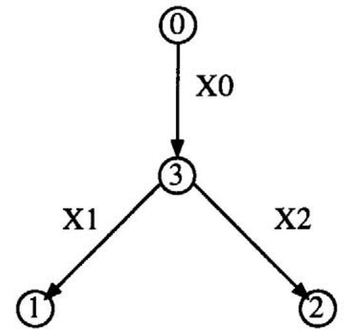

Figure 2.3: A simple path topology

We assume each X_i has a stationary distribution, and thus, let $x_i \stackrel{\text{def}}{=} \Pr[X_i = 0]$, which is one minus the loss rate of the link.

The goal is to infer the no-loss rates $\{x_i | 0 \leq i \leq 2\}$ from end-to-end measurement using multicast probes, i.e., observing whether each probe reaches the receivers' nodes or not. Note that "No loss along a path" $\Leftrightarrow$ "No loss on every link in a path".

Let Y_j denote a random variable that takes value 1 if a probe is not observed at the destination node j (i.e., it is dropped on the way), 0 otherwise. Let n be the number of all probes (trials) in a measurement, and c_{pq} be the number of trials in which event $(Y_1 = p \cap Y_2 = q)$ occurs for $p, q = 0, 1$. For convenient, we also define $m_1 \stackrel{\text{def}}{=} c_{00} + c_{01}$, $m_2 \stackrel{\text{def}}{=} c_{00} + c_{10}$, and $m_{12} \stackrel{\text{def}}{=} c_{00}$.

2.2 Queueing delay distributions on links

Let us consider the same network as the previous example (Fig. 2.3), and dispatch multicast probe packets in the same manner.

We regard an event that a probe packet experiences a queueing delay of d on a link as an occurrence of the delay of d on the link. Let X_i denote a random variable that takes value d if the delay of d occurs on link i. We assume the delays are quantized (as multiples of a unit time) in the range $\{0, 1, ..., d_{max}\}$. A too large delay and an infinite delay (i.e., packet loss) are regarded as the delay of d_{max}. We also assume each X_i has a stationary distribution, and thus, let $x_i(d) \stackrel{\text{def}}{=} \Pr[X_i = d]$, which is the probability of an occurrence of the delay of d on link i.

The goal is to infer the (discrete) queueing delay distribution $\{x_i(d) | 0 \leq i \leq 2, 0 \leq d \leq d_{max}\}$ from end-to-end measurement using multicast probes, i.e., observing the queueing delay that each probe experiences totally along a path. Note that "A queueing delay along a path" = "A sum of queueing delays on every link in a path".

Suppose the sender records a timestamp in each prove just before sending the probe, and each receiver can read the timestamp information in receiving probes. Furthermore, we assume the queueing delay along each path sometime takes 0. Since the directly-observable delay (i.e.,

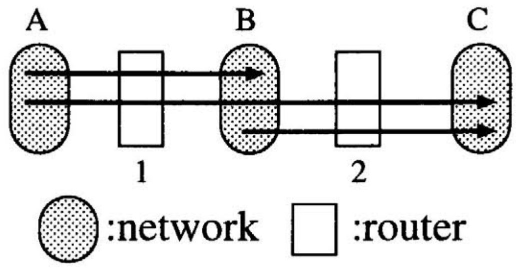

Figure 2.4: A simple aggregated-flow topology

difference between the timestamp in a received probe and the receiving time of the prove at the receiver) mainly consists of *propagation time* + (*packet size* / *bandwidth*) + *queuing delay* along the path, we can extract the queuing delay of each probe by using the minimum value of the directly-observable delays along the path.

Let Y_j denote a random variable that takes the value of the queueing delay of a probe along the path to the destination node j. Let n be the number of all probes (trials) in a measurement, and c_{pq} be the number of trials in which event $(Y_1 = p \cap Y_2 = q)$ occurs for $p, q = 0, 1, ..., d_{max}$. For convenient, we also define $m_1(d) \stackrel{\text{def}}{=} \sum_{p=0}^{d} \sum_{q=0}^{d_{max}} c_{pq}$, $m_2(d) \stackrel{\text{def}}{=} \sum_{p=0}^{d_{max}} \sum_{q=0}^{d} c_{pq}$, and $m_{12}(d) \stackrel{\text{def}}{=} \sum_{p=0}^{d} \sum_{q=0}^{d} c_{pq}$.

2.3 Occurrence rates of the anomalous packets on flows

Let us consider the most simple aggregated-flow topology Fig. 2.4 (a relation between flows and observable links passed through by the flows), and observe packets passing through the routers (as observable links) 1 and 2. There exist three flows from network A to C, from A to B, and from B to C, which are labeled by 0, 1 and 2, respectively.

We regard an event that some kind of packets (e.g., a kind of ICMP packets) pass through a flow in an observation time-interval as an occurrence of the "anomalous" state on the flow. Let X_i denote a random variable that takes value 1 if the anomalous state occurs on flow i, 0 otherwise. We assume each X_i has a stationary distribution, and thus, let $x_i \stackrel{\text{def}}{=} \Pr[X_i = 0]$, which is regarded as the occurrence rate of the "silent" state, i.e., one minus that of the anomalous state on flow i.

The goal is to infer the silent state rates $\{x_i | 0 \leq i \leq 2\}$ by observing whether anomalous packets pass through the routers or not, in each observation time-interval. Note that "No anomalous packets at a router" $\Leftrightarrow$ "No anomalous packets in every flow through a router".

Let Y_j denote a random variable that takes value 1 if anomalous packets are observed at router j in an observation time-interval, 0 otherwise, Let n be the number of observation time-intervals (trials) in a measurement, and c_{pq} be the number of trials in which event $(Y_1 = p \cap Y_2 = q)$

occurs for $p, q = 0, 1$. For convenient, we also define $m_1 \stackrel{\text{def}}{=} c_{00} + c_{01}$, $m_2 \stackrel{\text{def}}{=} c_{00} + c_{10}$, and $m_{12} \stackrel{\text{def}}{=} c_{00}$. For conciseness, we ignore the problem related to timing of observations at different routers (points), e.g., the influence of the difference between the time of observations for the same packet on a flow, or the clock synchronization between observers.

2.4 Arrival rates on flows (a discrete model)

Let us consider the same network as the previous example (Fig. 2.4), and observe arrival rates of packets passing through the routers, where "arrival rate" means the number of some kind of packets (or bytes) arrived in a unit time-interval. We regard an event that the number of (some kind of) packets passing through the routers on a flow in an observation time-interval is d as an occurrence of the arrival rate of d on the flow. Let X_i denote a random variable that takes value d if the arrival rate of d occurs on flow i. Note that "The number of arrival packets on an aggregated-flow" = "A sum of the number of arrival packets on every flow in an aggregated-flow". First, in this subsection, a discrete model is shown, and then, in the next subsection, a normal-based model is also shown.

In a discrete model, we assume the arrival rates are quantized (as multiples of a unit bin) in the range $\{0, 1, ..., d_{max}\}$. A too high rate is regarded as the rate of d_{max}. We also assume each X_i has a stationary distribution, and thus, let $x_i(d) \stackrel{\text{def}}{=} \Pr[X_i = d]$, which is the probability of an occurrence of the arrival rate of d on flow i.

The goal is to infer the (discrete) arrival rates distribution $\{x_i(d) | 0 \leq i \leq 2, 0 \leq d \leq d_{max}\}$ by observing the number of arrival packets on aggregated-flows at the routers in each observation time-interval.

Let Y_j denote a random variable that takes the value of the arrival rate on an aggregated-flow at router j. Let n be the number of observation time-intervals (trials) in a measurement, and c_{pq} be the number of trials in which event $(Y_1 = p \cap Y_2 = q)$ occurs for $p, q = 0, 1, ..., d_{max}$. For convenient, we also define $m_1(d) \stackrel{\text{def}}{=} \sum_{p=0}^{d} \sum_{q=0}^{d_{max}} c_{pq}$, $m_2(d) \stackrel{\text{def}}{=} \sum_{p=0}^{d_{max}} \sum_{q=0}^{d} c_{pq}$, and $m_{12}(d) \stackrel{\text{def}}{=} \sum_{p=0}^{d} \sum_{q=0}^{d} c_{pq}$.

2.5 Arrival rates on flows (a normal-based model)

In a normal-based model, in contrast to the previous subsection, we assume the arrival rates of individual flows are modeled by independent stationary normal distributions with a special relation between means and variances. Note that it is also implicitly assumed that arrival rates are so large that they are suitable to be modeled by normal distributions.

Let $m_i \stackrel{\text{def}}{=} E[X_i]$ and $v_i \stackrel{\text{def}}{=} Var[X_i]$ $(i \in \{0, 1, 2\})$, and let us assume: $v_i = \rho m_i$ for an unknown constant (scale parameter) ρ, for example.

The goal is to infer the parameters governing the arrival rates distribution, i.e., $\theta \stackrel{\text{def}}{=} \{m_0, m_1, m_2, \rho$

by observing the number of arrival packets on aggregated-flows at the routers in each observation time-interval.

Let Y_j denote a random variable that takes the value of the arrival rate on an aggregated-flow at router j. Since X_is are independent each other, the probability density function $p_{\boldsymbol{X}}(\boldsymbol{\xi};\theta)$ of $\boldsymbol{X} = \{X_i\}$ for a given θ is:

$$\frac{1}{(2\pi)^{\frac{3}{2}}|\boldsymbol{V}|^{\frac{1}{2}}} \exp(-\frac{1}{2}{}^t(\boldsymbol{\xi} - \boldsymbol{m})\boldsymbol{V}^{-1}(\boldsymbol{\xi} - \boldsymbol{m})) \tag{2.1}$$

where mean vector $\boldsymbol{m}$ and covariance matrix $\boldsymbol{V}$ are as follows.

$$\boldsymbol{m} \stackrel{\mathrm{def}}{=} \begin{pmatrix} m_0 \\ m_1 \\ m_2 \end{pmatrix}, \boldsymbol{V} \stackrel{\mathrm{def}}{=} \begin{pmatrix} v_0 & 0 & 0 \\ 0 & v_1 & 0 \\ 0 & 0 & v_2 \end{pmatrix}$$

Thus, $\boldsymbol{Y} = \{Y_j\} = \boldsymbol{A}\boldsymbol{X}$ also has a multivariate normal distribution with mean vector $\boldsymbol{A}\boldsymbol{m}$ and covariance matrix $\boldsymbol{A}\boldsymbol{V}{}^t\boldsymbol{A}$, i.e., the probability density function $p_{\boldsymbol{Y}}(\boldsymbol{\eta};\theta)$ for a given θ is:

$$\frac{1}{(2\pi)D^{\frac{1}{2}}} \exp(-\frac{1}{2}{}^t(\boldsymbol{\eta} - \boldsymbol{A}\boldsymbol{m})(\boldsymbol{A}\boldsymbol{V}{}^t\boldsymbol{A})^{-1}(\boldsymbol{\eta} - \boldsymbol{A}\boldsymbol{m})) \tag{2.2}$$

where

$$\boldsymbol{A} \stackrel{\mathrm{def}}{=} \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}, \ D \stackrel{\mathrm{def}}{=} |\boldsymbol{A}\boldsymbol{V}{}^t\boldsymbol{A}|$$

.

Let n be the number of observation time-intervals (trials) in a measurement, and $\eta_j^{(i)}$ be the observed arrival rate on an aggregated-flow at router j in the i-th trials ($1 \leq i \leq n$).

3 Two approaches to the tomography

The above inference problems are included in a very general class of statistical inverse problem as follows. Under an appropriate stochastic model, let $\boldsymbol{X}$ be a multidimensional random variable that is unobservable, $\boldsymbol{Y}$ be a multidimensional random variable that is observable (i.e., whose statistics can be estimated in a straight-forward way from observations), and R is a relation between $\boldsymbol{X}$ and $\boldsymbol{Y}$, which is known by us: $R(\boldsymbol{X}, \boldsymbol{Y})$. Note that we assume R is a deterministic relation. Furthermore, we are interested in cases that there exists the unique map from $\boldsymbol{X}$ to $\boldsymbol{Y}$, but the inverse (from $\boldsymbol{Y}$ to $\boldsymbol{X}$) is not unique (Fig. 2.5).

Suppose we perform a number of trials (observations) and collect data $\mathcal{M}$ related to observation of $\boldsymbol{Y}$. Our goal is to infer some interested parameter $\boldsymbol{x}$ governing statistics of $\boldsymbol{X}$ from $\mathcal{M}$.

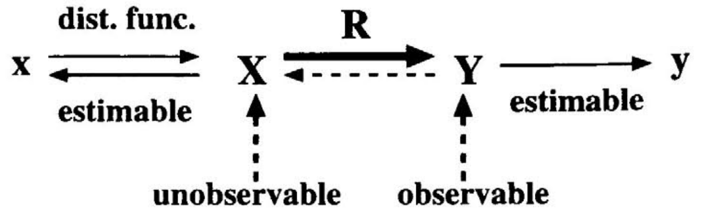

Figure 2.5: Relation between X and Y

3.1 A Maximum Likelihood Estimator

Let $\mathcal{M}_c$ be data as a termination condition of a measurement (related to Y) where we finish the measurement when $\mathcal{M}_c$ occurs. Then let us divide measured data $\mathcal{M}$ into $\mathcal{M}_c$ and $\mathcal{M}_o$, i.e., $\mathcal{M} = (\mathcal{M}_c, \mathcal{M}_o)$. The likelihood $L[\mathcal{M}_o|\mathcal{M}_c]$ of $(\mathcal{M}_o|\mathcal{M}_c)$ is defined as the conditional probability (or probability density in cases for continuous distributions) of the occurrence of $\mathcal{M}_o$ given that $\mathcal{M}_c$ occurs. We assume that the likelihood is expressed using parameter $\boldsymbol{x}$ based on relation R between X and Y (i.e., $L[\mathcal{M}_o|\mathcal{M}_c; \boldsymbol{x}]$).

The MLE (maximum likelihood estimator) of $\boldsymbol{x}$ with respect to $(\mathcal{M}_o|\mathcal{M}_c)$ is a parameter maximizing the likelihood function, or equivalently, maximizing the log likelihood function $\log L[\mathcal{M}_o|\mathcal{M}_c; \boldsymbol{x}]$ (i.e., $\arg\max_{\boldsymbol{x}} \log L[\mathcal{M}_o|\mathcal{M}_c; \boldsymbol{x}]$).

If this maximization has a unique solution and can be solved exactly (analytically), the MLE is always a good choice for the inference of $\boldsymbol{x}$ because of its preferable properties, e.g., consistency, efficiency, and asymptotic normality. In general, however, the above maximization problem is often difficult to solve exactly, and thus we need sophisticated methods; two typical approaches often appear in recent researches of the network tomography. One is calculating an inverse function mapping from some parameter $\boldsymbol{y}$ (with respect to statistics of Y) to $\boldsymbol{x}$. The other is solving an MLE approximately and numerically. Let us explain those two approaches and how to apply them to the previous five examples. In what follows, we employ the following common assumptions:

- $\{X_i | 0 \le i \le 2\}$ are independent each other, i.e., spatial independence.
- For each i, X_i is *iid* over successive trials (probes or time-intervals), i.e., temporal independence.
- For discrete models, $c_{00} > 0$. This implies that each x_i in Section 2.1 and Section 2.3 (each $x_i(0)$ in Section 2.2 and Section 2.4) is not equal to 0, which conditions are required for the use of an essential relation between values 0 of the random variables: $Y_j = 0 \Leftrightarrow X_0 = 0 \wedge X_j = 0$ for ($j \in \{1, 2\}$).

3.2 Calculation of an inverse function

The first approach calculates an inverse function from some parameter $\boldsymbol{y}$ (with respect to statistics of $\boldsymbol{Y}$) to $\boldsymbol{x}$, and uses it to infer $\boldsymbol{x}$. Although a map from $\boldsymbol{Y}$ to $\boldsymbol{X}$ is not unique, there often exists a unique map from $\boldsymbol{y}$ to $\boldsymbol{x}$ (Fig. 2.5). The widest definition may be given as follows.

1. Express parameter $\boldsymbol{y}$ by $\boldsymbol{x}$ deterministically, based on relation R, i.e., $\boldsymbol{y} = F(\boldsymbol{x})$.
2. Estimate $\boldsymbol{y}$ statistically from measured $\mathcal{M}$, i.e., $\hat{\boldsymbol{y}} \stackrel{\text{def}}{=} H(\mathcal{M})$.
3. Find the inverse $G = F^{-1}$ and infer $\boldsymbol{x}$ using $\hat{\boldsymbol{y}}$, i.e., $\hat{\boldsymbol{x}} \stackrel{\text{def}}{=} G(H(\mathcal{M}))$.

This approach seems to be promising especially in cases that X can be modeled by a discrete distribution under a certain condition (See [7]) like the following examples. The examples in Section 2.1 and Section 2.3 have the same model, and thus the "inverse function" approach can be applied in the same way.

Let $y_j \stackrel{\text{def}}{=} \Pr[Y_j = 0]$ (for $j \in \{1, 2\}$) and $y_{12} \stackrel{\text{def}}{=} \Pr[Y_1 = 0 \cap Y_2 = 0]$, respectively. Under spatial independence, we have a trivial map $\boldsymbol{y} = F(\boldsymbol{x})$:

$$y_1 = x_0 x_1,\ y_2 = x_0 x_2,\ y_{12} = x_0 x_1 x_2,$$

which has a trivial inverse map $\boldsymbol{x} = G(\boldsymbol{y})$, if $x_i > 0$, as follows.

$$x_0 = y_1 y_2 / y_{12},\ x_1 = y_{12}/y_2,\ x_2 = y_{12}/y_1.$$

On the other hand, under temporal independence, from obtained data $\mathcal{M}_c = \{n\}$ and $\mathcal{M}_o = \{c_{00}, c_{01}, c_{10}, c_{11}\}$ by observing Y_1 and Y_2, $\boldsymbol{y}$ is estimated in a straight-forward way, i.e., $\hat{y}_j \stackrel{\text{def}}{=} m_j/n$ (for $j \in \{1, 2\}$) and $\hat{y}_{12} \stackrel{\text{def}}{=} m_{12}/n$.

Consequently, since we assume $c_{00} > 0$, we infer $\boldsymbol{x}$ as

$$\hat{x}_0 \stackrel{\text{def}}{=} \frac{m_1 m_2}{n m_{12}},\ \hat{x}_1 \stackrel{\text{def}}{=} \frac{m_{12}}{m_2},\ \hat{x}_2 \stackrel{\text{def}}{=} \frac{m_{12}}{m_1}$$

A basic idea for the above inference can be explained as follows. Recall Fig. 2.1 for loss rates on a link, or Fig. 2.2 for anomalous packet rates on a flow. In a trial, if $Y_1 = 0$ (i.e., a probe packet reaches node 1, or no anomalous packet is observed at router 1) and $Y_2 = 1$ (i.e., a probe does not reach node 2, or an anomalous packet is observed at router 2), then we know $X_2 = 1$ (i.e., a probe is dropped on link 2, or an anomalous packet passes through flow 2). Thus, we can estimate statistics of X_2 after a number of trials.

Let us see the relation to the MLE of $\boldsymbol{x}$. We denote $L[\{m_1, m_1, m_{12}\}, \{n\}; \boldsymbol{x}]$ as the likelihood of $(\{m_1, m_2, m_{12}\}|\{n\})$. It is clear that $L \propto \eta_{12}^{m_{12}}(\eta_1 - \eta_{12})^{m_1 - m_{12}}(\eta_2 - \eta_{12})^{m_2 - m_{12}}(1 - \eta_1 - \eta_2 + \eta_{12})^{n - m_1 - m_2 + m_{12}}$, where $\eta_1 \stackrel{\text{def}}{=} x_0 x_1$, $\eta_2 \stackrel{\text{def}}{=} x_0 x_2$ and $\eta_{12} \stackrel{\text{def}}{=} x_0 x_1 x_2$. Thus, $(\eta_1, \eta_2, \eta_{12}) = (m_1/n, m_2/n, m_{12}/n)$ maximizes this likelihood, and thus , since we assume $c_{00} > 0$, so does

$$(x_0, x_1, x_2) = (\frac{m_1 m_2}{n m_{12}}, \frac{m_{12}}{m_2}, \frac{m_{12}}{m_1}) \tag{2.3}$$

Hence, the above $\hat{\boldsymbol{y}}$ and $\hat{\boldsymbol{x}}$ are equal to the MLE of $\boldsymbol{y}$ and $\boldsymbol{x}$ with respect to $(\{m_1, m_2, m_{12}\}|\{n\})$, respectively.

In this special case, it should be noted that, since $\{n, c_{00}, c_{01}, c_{10}, c_{11}\}$ and $\{m_1, m_2, m_{12}, n\}$ have an one-to-one corresponding, two likelihoods of those data are the same, and thus, the above $\hat{\boldsymbol{y}}$ and $\hat{\boldsymbol{x}}$ are also equal to the MLE of $\boldsymbol{y}$ and $\boldsymbol{x}$ w.r.t. $(\mathcal{M}_o|\mathcal{M}_c)$, respectively.

In cases for general tree path topologies, although there may exist several estimators of $\boldsymbol{x}$ derived from this approach, it can be shown that one of them is equal to the MLE of $\boldsymbol{x}$ w.r.t. $(\mathcal{M}_o|\mathcal{M}_c)$ for a sufficiently large n ([4]). In cases for general path or aggregated-flow topologies, however, those estimators are different from the MLE w.r.t. $(\mathcal{M}_o|\mathcal{M}_c)$, although it can be shown that, for each estimator, $\hat{\boldsymbol{x}} \to \boldsymbol{x}$ when $\hat{\boldsymbol{y}} \to \boldsymbol{y}$ as $n \to \infty$.

Next we consider the examples in Section 2.2 and Section 2.4, which have the same model and can be regarded as a generalization of the above. For $j \in \{1,2\}$ and $d \in \{0, 1, ...d_{max}\}$, let $y_j(d) \stackrel{\text{def}}{=} \Pr[Y_j \le d]$, $y_{12}(d) \stackrel{\text{def}}{=} \Pr[Y_1 \le d \cap Y_2 \le d]$, $\boldsymbol{y}(d) \stackrel{\text{def}}{=} \{y_{12}(k), y_1(k), y_2(k)|0 \le k \le d\}$, and $\boldsymbol{x}(d) \stackrel{\text{def}}{=} \{x_0(k), x_1(k), x_2(k)|0 \le k \le d\}$.

Under spatial independence, we have a map $\boldsymbol{y}(d) = F^{(d)}(\boldsymbol{x}(d))$ for each $d \in \{0, 1, ..d_{max}\}$:

$$\begin{aligned} y_j(d) &= \sum_{k=0}^{d} x_0(k)\underline{x}_j(d-k) \quad j = 1, 2 \\ y_{12}(d) &= \sum_{k=0}^{d} x_0(k)\underline{x}_1(d-k)\underline{x}_2(d-k) \end{aligned}$$

where $\underline{x}_j(d) \stackrel{\text{def}}{=} \sum_{k=0}^{d} x_j(d)$. It can be shown that, if $x_i(0) > 0$ ($i \in \{0, 1, 2\}$), the above system can be solved inductively with respect to d and has the inverse $G^{(d)} = (G_0^{(d)}, G_1^{(d)}, G_2^{(d)})$:

$$x_i(d) = G_i^{(d)}(y_{12}(k), y_1(k), y_2(k); 0 \le k \le d)$$

On the other hand, under temporal independence, from obtained data $\mathcal{M}_c = \{n\}$ and $\mathcal{M}_o = \{n, c_{pq}|0 \le p \le d_{max}, 0 \le q \le d_{max}\}$ by observing Y_1 and Y_2, $\boldsymbol{y}$ is estimated in a straightforward way, i.e., $\hat{y}_j(d) \stackrel{\text{def}}{=} m_j(d)/n$ (for $j \in \{1, 2\}$) and $\hat{y}_{12}(d) \stackrel{\text{def}}{=} m_{12}(d)/n$.

Consequently, since we assume $c_{00} > 0$, we infer $\boldsymbol{x}$ as $(0 \le i \le 2)$:

$$\hat{x}_i(d) \stackrel{\text{def}}{=} G_i^{(d)}(\frac{m_{12}(k)}{n}, \frac{m_1(k)}{n}, \frac{m_2(k)}{n}; 0 \le k \le d)$$

A basic idea for the above inference can be explained as follows. Recall Fig. 2.1 for queueing delays on a link, or Fig. 2.2 for arrival rates on a flow. In a trial, if $Y_1 = 0$ (i.e., no queueing delay is observed at node 1, or no packet is observed at router 1) and $Y_2 = d$ (i.e., a queueing delay observed at node 2 is d, or an arrival rate of packet observed at router 2 is d), then we know $X_2 = d$ (i.e., a queueing delay on link 2 is d, or an arrival rate of packet on flow 2 is d),). Thus, we can estimate statistics of X_2 after a number of trials. Note that such estimators are different from the MLE of $\boldsymbol{x}$ w.r.t. $(\mathcal{M}_o|\mathcal{M}_c)$, because they use only partial information of obtained data $\mathcal{M}$.

Finally we consider the example in Section 2.5, to which the "inverse function" approach in a broad sense is applicable. Let $m_j^* \stackrel{\text{def}}{=} E[Y_j]$, $v_1^* \stackrel{\text{def}}{=} Var[Y_1]$, and $v_{12}^* \stackrel{\text{def}}{=} Cov[Y_1, Y_2]$ (for $j \in \{1, 2\}$). By using $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X}$ and $v_i = \rho m_i$ (for $i \in \{0, 1, 2\}$), under spatial independence, we have a map from θ to $\{m_1^*, m_2^*, v_1^*, v_{12}^*\}$:

$$m_j^* = m_0 + m_j,\ v_{12}^* = \rho m_0,\ v_1^* = \rho(m_0 + m_1)$$

which has a trivial inverse map as follows.

$$\rho = \frac{v_1^*}{m_1^*},\ m_0 = \frac{m_1^* v_{12}^*}{v_1^*},\ m_j = m_j^* - \frac{m_1^* v_{12}^*}{v_1^*}$$

On the other hand, under temporal independence, from obtained data $\mathcal{M} = \{\boldsymbol{\eta}^{(i)} = (\eta_1^{(i)}, \eta_2^{(i)}) | 1 \le i \le n\}$, $\{m_1^*, m_2^*, v_1^*, v_{12}^*\}$ is estimated in a straight-forward way:

$$\begin{aligned}
\hat{m}_j^* &\stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{n} \eta_j^{(i)} \qquad (j \in \{1, 2\}) \\
\hat{v}_1^* &\stackrel{\text{def}}{=} \frac{1}{n-1}\sum_{i=1}^{n} (\eta_1^{(i)} - \hat{m}_1^*)^2 \\
\hat{v}_{12}^* &\stackrel{\text{def}}{=} \frac{1}{n-1}\left(\sum_{i=1}^{n} \eta_1^{(i)}\eta_2^{(i)} - \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} \eta_1^{(i)}\eta_2^{(j)}\right)
\end{aligned}$$

which can be used to infer θ.

However, since the above method uses little information to infer the parameters, it is not expected to give a good estimator.

3.3 The EM method

The second approach solves an MLE approximately and numerically. While there exist classical methods like the Newton-Raphson methods and its variants (e.g., Fisher's scoring method), the EM (expectation-maximization) algorithm is regarded as one of promising methods for this purpose. The EM method can finally find a value for $\boldsymbol{x}$ that maximizes likelihood $L[\mathcal{M}_o|\mathcal{M}_c; \boldsymbol{x}]$, but it dose so by making essential use of another likelihood $L[\mathcal{M}_o, \mathcal{M}'|\mathcal{M}_c; \boldsymbol{x}]$ where $\mathcal{M}'$ is some unobservable (hidden) data ([12], [13]).

Let $\boldsymbol{M}$ and $\boldsymbol{M'}$ be multidimensional random variables corresponding to observable data $\mathcal{M}$ and unobservable data $\mathcal{M}'$, respectively. Moreover, let $f(.,.)$ be the function satisfying that $f(\mathcal{M}, \boldsymbol{x}) = \log L[\mathcal{M}_o|\mathcal{M}_c; \boldsymbol{x}]$, and $g(.,.,.)$ be the function satisfying that $g(\mathcal{M}, \mathcal{M}', \boldsymbol{x}) = \log L[\mathcal{M}_o, \mathcal{M}'|\mathcal{M}_c; \boldsymbol{x}]$. Although the goal is to find a value for $\boldsymbol{\theta}$ maximizing $f(\mathcal{M}, \boldsymbol{\theta})$ for given $\mathcal{M}$, we employ the maximization of an expectation of $g(\mathcal{M}, \mathcal{M}', \boldsymbol{x})$, instead. Since $\mathcal{M}'$ cannot be measured, we consider the conditional probability (or probability density) of $\boldsymbol{M'}$ given that

$\boldsymbol{M} = \mathcal{M}$, and assume it can be computed under a given parameter $\boldsymbol{x}$. Thus we can take the conditional expectation of $g(\boldsymbol{M}, \boldsymbol{M}', \boldsymbol{\theta})$ given that $\boldsymbol{M} = \mathcal{M}$, as function Q:

$$Q(\mathcal{M}, \boldsymbol{\theta}, \boldsymbol{x}) \stackrel{\text{def}}{=} E[g(\boldsymbol{M}, \boldsymbol{M}', \boldsymbol{\theta}) | \boldsymbol{M} = \mathcal{M}; \boldsymbol{x}]$$

By using measured data $\mathcal{M}$, the EM method finds $\hat{\boldsymbol{x}}$ iteratively as follows.

1. Choose an appropriate initial parameter $\boldsymbol{x}_0$.
2. (E-step): For a given (k-th stage) parameter $\boldsymbol{x}_k$, construct $Q(\mathcal{M}, \boldsymbol{\theta}, \boldsymbol{x}_k)$ as a function with only free variable $\boldsymbol{\theta}$.
3. (M-step): Find a $\boldsymbol{\theta}$ maximizing $Q(\mathcal{M}, \boldsymbol{\theta}, \boldsymbol{x}_k)$ as the next stage parameter $\boldsymbol{x}_{k+1}$:
   $\boldsymbol{x}_{k+1} \stackrel{\text{def}}{=} \arg\max_{\boldsymbol{\theta}} Q(\mathcal{M}, \boldsymbol{\theta}, \boldsymbol{x}_k)$.
4. If $\boldsymbol{x}_k \approx \boldsymbol{x}_{k+1}$, then $\hat{\boldsymbol{x}} \stackrel{\text{def}}{=} \boldsymbol{x}_{k+1}$ and stop. Otherwise, go to (E-step) again.

Since it can be shown that $f(\mathcal{M}, \boldsymbol{x}_k) \leq f(\mathcal{M}, \boldsymbol{x}_{k+1})$ for each stage k, the above $\hat{\boldsymbol{x}}$ is expected to give us (at least) a local maximum of $f(\mathcal{M}, \boldsymbol{\theta})$, and thus is a candidate of the MLE with respect to given $\mathcal{M}$. It is clear that the EM method is useful when maximization (w.r.t. $\boldsymbol{\theta}$) of $Q(\mathcal{M}, \boldsymbol{\theta}, \boldsymbol{x})$ is easier than that of $f(\mathcal{M}, \boldsymbol{\theta})$.

For the examples in Section 2.1 and Section 2.3, as shown in Section 3.2, we can compute a trivial MLE (2.3) directly, so that no sophisticated method is necessary.

On the other hand, for the examples in Section 2.2 and Section 2.4, under spatial independence, we have $\Pr[Y_1 = p \bigcap Y_2 = q] = \sum_{k=0}^{K_{pq}} x_0(k)x_1(p-k)x_2(q-k)$. Moreover, under temporal independence, the likelihood $L[\{c_{pq}|0 \leq p, q \leq d_{max}\}, \{n\}; \{x_i(r)|0 \leq i \leq 2, 0 \leq r \leq d_{max}\}]$ can be expressed as:

$$\prod_{p,q}(\sum_{k=0}^{K_{pq}} x_0(k)x_1(p-k)x_2(q-k))^{c_{pq}} \tag{2.4}$$

where $K_{pq} \stackrel{\text{def}}{=} \min(p, q)$. Since it is difficult to find $\{x_i(r)\}$ maximizing the above likelihood (2.4) directly, let us employ the EM method. Note that, if $c_{00} = 0$, then the maximization of (2.4) is not uniquely solvable. Thus, we assume $c_{00} > 0$ as mentioned in Section 3.1.

Suppose $a_i(r)$ is the number of trials in which a probe experiences the queueing delay of r on link i (or in which the arrival rate of packets on flow i is r), and consider $\{a_i(r)|0 \leq i \leq 2, 0 \leq r \leq d_{max}\}$ as hidden data. Under spatial and temporal independence, the likelihood $L[\{c_{pq}\} \cup \{a_i(r)\}, \{n\}; x]$ can be expressed as

$$\prod_i \prod_r x_i(r)^{a_i(r)}$$

for any $\{c_{pq}\}$ that is consistent with $\{a_i(r)\}$.

Let $A_i(r)$ be a random variable corresponding to $a_i(r)$, $g(\{a_i(r)\}, \{x_i(r)\}) \stackrel{\text{def}}{=} \sum_i \sum_r a_i(r) \log x_i(r)$ as the function expressing the log likelihood, and $Q(\{c_{pq}\}, \boldsymbol{\theta}, \boldsymbol{x}) \stackrel{\text{def}}{=} E[g(\{A_i(r)\}, \boldsymbol{\theta}) | \{c_{pq}\}; \boldsymbol{x}] = \sum_i \sum_r \hat{a}_i(r) \log \theta_i(r)$ where

$$\begin{aligned} \hat{a}_i(r) &\stackrel{\text{def}}{=} E[A_i(r) | \{c_{pq}, n\}; \{x_i(r)\}] \\ &= \sum_{p,q} c_{pq} \Pr[X_i = r | Y_1 = p \cap Y_2 = q] \\ &= \sum_{p,q} c_{pq} \frac{\Pr[X_i = r \cap Y_1 = p \cap Y_2 = q]}{\Pr[Y_1 = p \cap Y_2 = q]} \end{aligned}$$

Note that $\hat{a}_i(r)$ can be computed using $\{c_{pq}\}$ and $\{x_i(r)\}$, and $\hat{\theta}_i(r) \stackrel{\text{def}}{=} \hat{a}_i(r)/\sum_r \hat{a}_i(r)$ maximizes $Q(\{c_{pq}\}, \{\theta_i(r)\}, \{x_i(r)\})$ because $\sum_r \theta_i(r) = 1$ for each i. Therefore, on the k-th stage of the EM method,

- (E-step): Compute $Q(\{c_{pq}\}, \{\theta_i(r)\}, \{x_i^{(k)}(r)\})$ as a function with free variables $\{\theta_i(r)\}$, by using $\{c_{pq}\}$ and $\{x_i^{(k)}(r)\}$.
- (M-step): Compute $\{\hat{a}_i(r)/\sum_{i,r} \hat{a}_i(r)\}$ as the next stage parameter $\{x_i^{(k+1)}(r)\}$.

Finally we consider the example in Section 2.5. Using the probability density function (2.2), i.e., $p_{\boldsymbol{Y}}(\boldsymbol{\eta}; \theta)$, of $\boldsymbol{Y}$ for a given $\theta = (m_0, m_1, m_2, \rho)$, the likelihood $L[\{\boldsymbol{\eta}^{(i)} | 1 \leq i \leq n\}; \theta]$ can be expresses as $\prod_{i=1}^n p_{\boldsymbol{Y}}(\boldsymbol{\eta}^{(i)}; \theta)$ under temporal independence. Since it is difficult to find θ maximizing the above likelihood directly, let us employ the EM method.

Suppose $\xi_j^{(i)}$ is the (unobservable) arrival rate on flow j in the i-th observation time-interval, and consider $\{\xi_j^{(i)} | 0 \leq j \leq 2, 1 \leq i \leq n\}$ as hidden data. Under temporal independence, the likelihood $L[\{\boldsymbol{\eta}^{(i)}\}, \{\boldsymbol{\xi}^{(i)}\}; \theta]$ can be expressed as $\prod_{i=1}^n p_{\boldsymbol{X}}(\boldsymbol{\xi}^{(i)}; \theta)$ where $p_{\boldsymbol{X}}(\boldsymbol{\xi}; \theta)$ is the probability density function (2.1) of $\boldsymbol{X}$ for a given θ.

Let $X_j^{(i)}$ and $Y_k^{(i)}$ be a random variable corresponding to $\xi_j^{(i)}$ and $\eta_k^{(i)}$, respectively. The Q-function with a given parameter $\hat{\theta}$ is defined as

$$\begin{aligned} Q(\{\boldsymbol{\eta}^{(i)}\}, \theta, \hat{\theta}) &\stackrel{\text{def}}{=} E[\log L | \boldsymbol{Y}^{(i)} = \boldsymbol{\eta}^{(i)}, 1 \leq i \leq n; \hat{\theta}] \\ &= \sum_{i=1}^n E[\log p_{\boldsymbol{X}}(\boldsymbol{X}^{(i)}; \theta) | \boldsymbol{Y}^{(i)} = \boldsymbol{\eta}^{(i)}; \hat{\theta}] \\ &= h(\boldsymbol{m}, \boldsymbol{V}, \sum_{i=1}^n \hat{\boldsymbol{m}}^{(i)}, \sum_{i=1}^n (\hat{\boldsymbol{m}}^{(i)})^2, \hat{\boldsymbol{v}}) \end{aligned}$$

where h is a (given) polynomial, $\hat{m}_j^{(i)} \stackrel{\text{def}}{=} E[\boldsymbol{X}_j^{(i)} | \boldsymbol{Y}^{(i)} = \boldsymbol{\eta}^{(i)}; \hat{\theta}]$, and $\hat{v}_j \stackrel{\text{def}}{=} Var[\boldsymbol{X}_j^{(i)} | \boldsymbol{Y}^{(i)} = \boldsymbol{\eta}^{(i)}; \hat{\theta}]$.

Since $\hat{m}_j^{(i)}$ and $\hat{v}_j$ can be computed using $\boldsymbol{\eta}^{(i)}$ and $\hat{\theta}$, the k-th stage of the EM method is as follows.

- (E-step): Compute $\tilde{Q}(\theta) \stackrel{\text{def}}{=} Q(\{\boldsymbol{\eta}^{(i)}\}, \theta, \hat{\theta}^{(k)})$ as a function with a free variable vector $\theta = (m_0, m_1, m_2, \rho)$, by using $\boldsymbol{\eta}^{(i)}$ and $\hat{\theta}^{(k)}$.

- (M-step): Compute (m_0, m_1, m_2, ρ) maximizing $\tilde{Q}(m_0, m_1, m_2, \rho)$ as the next stage parameter $\hat{\theta}^{(k+1)}$, by solving the unique positive solution of the following equations:

$$\frac{\partial \tilde{Q}}{\partial \rho} = 0, \ \frac{\partial \tilde{Q}}{\partial m_j} = 0 \ (j \in \{0, 1, 2\})$$

4 Overview and future work

As mentioned in the previous sections, the network tomography includes two typical forms; "inference of flow characteristics based on aggregated-flow measurements" and "inference of network-internal characteristics based on end-to-end path measurements". In order to solve them, two typical approaches; "an inverse function from measurable parameters to target parameters" and "an approximate solution of MLE of target parameters with respect to observed data", have been employed.

General statistical properties of both approaches should be noted. Since the "inverse function" approach uses only partial information in the whole observed data so that an inverse function is uniquely determined, several (different) estimators are derived from the same data. Although they have consistency, they are not always efficient estimators. On the other hand, although it uses the whole observed data (and the true MLE always has good properties), the "MLE solver" approach has general pitfalls of approximate solvers for maximization problems. For example, such solvers can find only local solutions, convergence of solvers is sometimes slow, and such behaviors often depend on the choice of an initial point of the iterations.

Researches for the former form (inference of flow characteristics) started from [2] and [3] (originated from a similar problem in the car transportation system) followed by [14] and [15] as "the OD (origin-destination) traffic matrix problem", which was to infer unobservable traffic volume (intensity) between each OD pair from the aggregated traffic volume measured at some routers' incoming and/or outgoing interfaces.

They assumed that all OD byte counts were modeled by independent normal (with a special relation between means and variances) or Poisson distributions, and were *iid* over successive observation time-intervals. Then they calculated the approximate MLE of parameters in the models numerically (e.g., using the EM method), which were expected to perform well for flows having relative large traffic volume. The time-varying nature was also treated by fitting the *iid* model locally using a moving data window [15].

On the other hand, under somewhat different conditions, a method based on the "inverse function" approach was presented for a variant of the OD traffic matrix inference [9], which came from a generalization of the "inverse function" approach for the latter form [7].

Researches for the latter form (inference of link characteristics) have been extensively done in the MINC (Multicast-based Inference of Network-internal Characteristics) project mainly based on the "inverse function" approach ([10]). They employed end-to-end multicast probe packets from a root sender to several leaf receivers in order to infer network-internal characteristics such as packet losses ([4], [16]), queueing delays ([17]), and queueing delay variances ([18]) on links. A network-internal topology (as a multicast tree structure from a sender to several receivers) was also inferred ([19], [20], [21]). Their method had rigorous and robust statistical properties due to the theoretical advantages of properties of multicast probes and tree structures of paths. The use of information gathered by RTCP (in RFC1886) was also discussed ([22], [23]).

The succeeding researches, however, have focused on the use of back-to-back packet pairs in unicast active probing or passive monitoring of unicast traffic, although unicast-based methods have a problem of imperfect correlation in concurrent events on paths. The reasons why they employed unicast-based methods are that multicast communication is not available in the major part of the current Internet, and, even in a network supporting multicast, the performance for multicast traffic often differs from that for unicast traffic, which is still major in the Internet. Inference of packet loss statistics ([5], [24], [25], [8]), queueing delay statistics ([26], [27], [6], [28]), physical bandwidths ([29]), and the existence of shared congestion ([30]) have been studied.

Among those works, the extensive use of the "MLE solver" approach (e.g., using the EM method) to this form of inference started from [5] followed by several researches ([24], [26], [27], [6], and [28]). On the other hand, the "inverse function" approach still has been employed ([25], [8], [6]). The use of passive measurements was shown in [24]. Benefit by using a packet stripe (more than two closely time-spaced packets) was demonstrated in [25]. Non-tree (e.g., inverse tree) path topologies were dealt with in [8]. The time-varying nature was treated by using a sequential Monte Carlo procedure in [27]. A combination of both approaches can be seen in [6] where the authors employed the "inverse function" approach to find an appropriate initial estimator for the "MLE solver" approach.

Consequently, in recent one or two years, as similarities and differences between two forms of the inference problems are recognized, a common framework of the network tomography has been roughly established. Table. 2.1 shows related researches, using the following notations: A - inference of link characteristics (unicast-based), A'- inference of link characteristics (multicast-based), B inference of flow characteristics, C generalization/formulation, X solving an MLE, Y - calculating an inverse function, Z - other specific approaches.

We remark that there exists much work to be done. First, more deployment: although a number of simulation results and experimental results have been presented, there exists little practical use of the network tomography in the real Internet. We have many issues to examine and solve in actual networks, such as, reliability (limitation), distributed simultaneous measurements, and scalability.

There exist several methods (algorithms) to approximately solve an MLE equation in the

Table 2.1: Related researches

	X	XY	Y	Z
A	[5], [26], [27], [24], [28]	[6]	[25], [8]	[30], [29]
A'			[4], [16], [17] [18], [23]	[22], [19], [20], [21],
B	[2], [3], [14], [15],		[9]	
C	[12](EM) etc.		[7]	

"MLE solver" approach as well as several methods (estimators) derived from the "inverse function" approach, in which we should consider a trade-off between computational and/or operational costs and accuracy of the inference. We should also choose a number of detailed choices (parameters) in measurement and inference phases of a method. Analysis and verification of accuracy and reliability of the inference by each concrete method in actual environments are of practical importance. Note that the accuracy in experiments (on simulations or real networks) reported in the above researches (Table. 2.1) ranges from under 1% to over 20% (in some sense of "relative errors"), which depends not only on the type of the inference method but on the situation for which the method is applied. Moreover, acceptable accuracy also depends on the situation.

Second, more applications: the framework and principle (behind two typical forms) of the network tomography may be applicable to other inference problems not only in the IP layer but also in various applications. We also expect that the third form of the network tomography will be found.

Third, beyond independence assumptions: we often make assumptions of spatial and/or temporal independence to solve a problem. While several studies have dealt with the time-varying nature (non-stationarity), spatial independence is often essential to identifiability. Several studies have employed Bayesian inference to deal with some dependence, but prior distributions were often uncertain. It is a big challenge to find novel methodologies that can deal with temporal and spatial dependence.

Chapter 3

Unified approach to inferring link characteristics from path characteristics

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 should be measurable in terms of its characteristics. Furthermore, network users become very interested in quality of services (QoS) provided by Internet Service Providers (ISP) as well as networks' connectivity. Thus ISP should be aware of QoS they provide and show actual QoS by use of measured data.

On the other hand, the Internet is characterized by its huge number of geometrically and administratively distributed routers and hosts, which makes it difficult to measure its internal states and performance [31].

Therefore, it is of practical importance to develop a statistical way to infer network-internal characteristics that cannot be measured directly. Our major concern is to study a general framework for this purpose. In order to infer characteristics of internal links (i.e. occurrence probabilities of states of internal links) from end-to-end path measurements, what we have to do is mainly to 1) infer characteristics of end-to-end paths by monitoring traffic on the paths, and to further 2) determine characteristics of links from the inferred characteristics (given by 1)) of the paths related to the links. The focus of this work is on the latter part. In other words, we intend to establish a fundamental model and calculations for the issue of 2), thus we assume that the characteristics of paths can be obtained in some way.

In this work, first, we show a general model and conditions that network-internal characteristics to be inferred must satisfy. They are satisfied by various performance parameters of practical interest such as packet loss rates and queuing delay time distribution. And then, we propose a general method of determining such characteristic of each internal-link from given characteristics of paths.

The related research have been extensively done in MINC project [10]. They use end-to-end multicast probe packets from a root sender to many leaf receivers in order to infer network-internal characteristics such as packet loss and delay statistics ([4], [16] and [17]). Their method has rigorous and robust statistical properties due to the advantages of multicast probes and tree structures of paths. However, it is restricted to networks in which multicast communication is available and to paths of a tree topology, whereas multicast is not supported by many ISPs at this time.

On the other hand, we treat general network topologies. Our inference method consists of two parts. One is an extension of MINC's calculation using occurrence probabilities of OR-events on a set of paths, and the other part is a new calculation using occurrence probabilities of AND-events on a set of paths. Since it is defined on a general characteristics model and on an arbitrary path-topology, our method has possibilities to be combined with various types of end-to-end measurements, which may include both unicast and multicast communications, and both active probes and passive monitoring.

Furthermore, our framework is applicable to such network design problems as determining minimal conditions (specifications) of characteristics of internal-links to achieve required characteristics of end-to-end paths.

The remainder of this chapter is organized as follows. Section II describes a model of network and some concepts on it. Section III describes a model of network-internal characteristics to be inferred. Section IV presents main results. Section V shows case studies where our method can be applied effectively. Section VI discusses some issues arising in applying our method to actual networks. Finally Section VII concludes this work. Appendix provides proofs of Lemmas and Theorems.

2 Network model

This section explains some concepts on our general network model. They include a "logical link" which is distinguishable by a set of paths which traverse the link, a lattice structure Δ with regard to a relation between paths and links, and the maximal number of independent path groups.

2.1 A set of paths including a link

Each packet starts at a source node, traverses a path and finally reaches a destination node. Assume we can know (identify) a path (i.e. a set of links) which an observed packet traverses.

Path is defined as the set of all paths whose characteristics can be obtained from observations. ***Link*** is defined as the set consisting of all physical links included by any path in ***Path***.

Fig. 3.1 shows some examples. ***Path*** $= \{a, b, c\}$. In (I) and (II) of Fig. 3.1, packets are sent from node 0 to 1,2,3 through path a, b, c respectively. In (III), packets are sent from node 0 to

2,3 and 1 to 4 through path a, b, c respectively. (II) can be regarded as static multi-routes under a policy routing, for example.

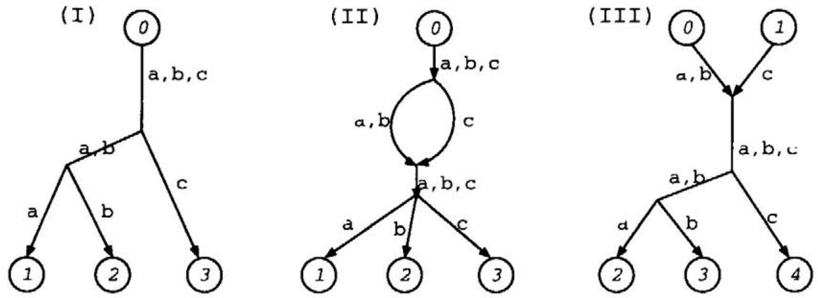

Figure 3.1: Examples of simple networks

For each physical link $e \in \boldsymbol{Link}$, we define an influence-path set $\boldsymbol{R}(e)$ as a set of paths which include the physical link e. Note that $\boldsymbol{R}(e) \neq \emptyset$.

We denote the set consisting of all influence-path sets by Δ: $\Delta \stackrel{\text{def}}{=} \{\boldsymbol{R}(e) | e \in \boldsymbol{Link}\}$. Then, for $R \in \Delta$, we define a "logical link" l_R as a set $\{e | \boldsymbol{R}(e) = R\}$ of physical links. We also denote the set consisting of all logical links by Δ^*: $\Delta^* \stackrel{\text{def}}{=} \{l_R | R \in \Delta\}$, thus $R \in \Delta \leftrightarrow l_R \in \Delta^*$. For example, in all three of Fig. 3.1, $\Delta^* = \{l_a, l_b, l_c, l_{ab}, l_{abc}\}$.

In what follows, we use a term "link" as a "logical link". A link may include more than one physical link. For example, in (II) of Fig. 3.1, there exists some logical link (l_c, l_{abc}) to which more than one physical link belong.

Note that a link of our model indicates a part of the network which packets traverse, and can correspond to various elements of an actual network. A link of our model can be just a link of the network, and can be a subnet of the network.

Δ (Δ^*) can be regarded as a lattice structure on which the inclusion relation among sets of paths is considered as semi-order. Fig. 3.2 shows examples of Δ. (I) of Fig. 3.2 corresponds to all three in Fig. 3.1, (II) corresponds to (I) of Fig. 3.3 in Section VI, and (III) corresponds to (II) of Fig. 3.3.

2.2 A set of links included in a path

For each path $r \in \boldsymbol{Path}$, we define $\boldsymbol{L}_r$ as a set of links which are included in the path r. We identify a path by its including links, thus $\boldsymbol{L}_r = \boldsymbol{L}_{r'} \Leftrightarrow r = r'$.

$$\boldsymbol{L}_r \stackrel{\text{def}}{=} \{l_R | r \in R,\ R \in \Delta\} \quad r \in \boldsymbol{Path}$$

For each non-empty set R of paths, we define $\underline{\boldsymbol{L}}(R)$ as a set of links which are included in at least one of all paths in R, $\overline{\boldsymbol{L}}(R)$ as a set of links which are included in all paths in R (namely

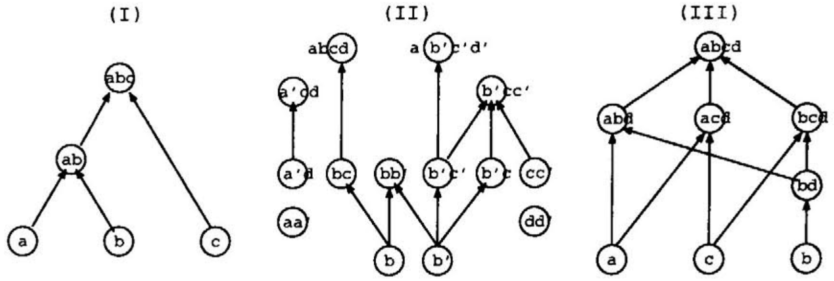

Figure 3.2: Examples of structure Δ

"shared-link set" of R).

$$\begin{aligned}\underline{L}(R) &\stackrel{\text{def}}{=} \{l_{R'} | R \cap R' \neq \emptyset,\, R' \in \Delta\} = \bigcup_{r \in R} L_r \\ \overline{L}(R) &\stackrel{\text{def}}{=} \{l_{R'} | R \subset R',\, R' \in \Delta\} = \bigcap_{r \in R} L_r\end{aligned}$$

We also define $L^R_{R'}$ as a set of links which are included in all paths in R' and are not included in any path in $R - R'$.

$$L^R_{R'} \stackrel{\text{def}}{=} \overline{L}(R') \setminus \underline{L}(R - R') \quad \text{for } R' \subset R$$

Note that

$$L^R_R = \overline{L}(R) \tag{3.1}$$

$$L^R_{R'} = L^{R+\{r\}}_{R'+\{r\}} + L^{R+\{r\}}_{R'} \quad \text{for } r \notin R \tag{3.2}$$

where (3.2) is shown as follows:

$$\begin{aligned}L^R_{R'} &= L^R_{R'} \cap L_r + L^R_{R'} \setminus L_r \\ &= \overline{L}(R' + \{r\}) \setminus \underline{L}(R - R') \;+\; \overline{L}(R') \setminus \underline{L}(R + \{r\} - R')\end{aligned}$$

2.3 Maximal number of independent path groups

We introduce the maximal number of independent path groups (called "the MIP number" hereafter) of a set R of paths satisfying that $\overline{L}(R) \neq \emptyset$. For such a set R, we decompose each L_r $(r \in R)$ to a shared part $(\overline{L}(R))$ and the remainder (K_r), i.e. $L_r = \overline{L}(R) + K_r$.

If no remainder part exists, i.e. $\underline{L}(R) = \overline{L}(R)$, then we define the MIP number of R as 0. Note that $|R| = 1 \Leftrightarrow$ the MIP number of R is equal to 0.

Otherwise there exists a path r $(\in R)$ satisfying that $K_r \neq \emptyset$. Let R_0 be the set consisting of all such paths, i.e. $R_0 = \{r \in R | K_r \neq \emptyset\}$. If the collection of "remainder" ($\bigcup_{r \in R_0} K_r$) can be divided to a maximum of N subsets with regard to paths ($\bigcup_{r \in R_i} K_r$, $i = 1, 2, .., N$) then we define the MIP number of R as N, and the MIP division as $\{R_1, R_2, .., R_N\}$. In other words, the MIP number of R is N if there exists a N-division $\{R_1, R_2, .., R_N\}$ of R_0 (i.e. $R_0 = \sum_{i=1}^{N} R_i$) satisfying that $(\bigcup_{r\in R_i} K_r) \cap (\bigcup_{r\in R_j} K_r) = \emptyset$ (for $i \neq j$) and no $(N+1)$-division has the above properties.

For example, in (I) of Fig. 3.2 where the root node indicates $\{a, b, c\}$, the MIP number of $\{a, b, c\}$ is 2, and the MIP division of $\{a, b, c\}$ is $\{\{a, b\}, \{c\}\}$ as shown there.

3 Network-internal characteristics model

This section presents a general model and conditions of network-internal characteristics to be inferred, and then explains how to apply the model to actual characteristics of links and paths such as packet loss and delay. We implicitly assume a suitable probability space $(\Omega, \mathcal{F}, \Pr)$.

3.1 A model and conditions of characteristics

- $M = \{0, 1, .., max\}$: A set of discrete numbers which represent states related to characteristic of a link or of a set of links.
- $X_l(m)$: An event that the state of a link l ($\in \Delta^*$) is $m \in M$.
- $V(L)(m)$: An event that the state of a set L ($\subset \Delta^*$) of links is $m \in M$. If $L = \{l\}$ then $V(\{l\})(m) = X_l(m)$.
- $Y_r(m) \stackrel{\text{def}}{=} V(L_r)(m)$: An event that the state of a path r ($\in$ ***Path***) is $m \in M$.
- $\underline{Z}(R)(m) \stackrel{\text{def}}{=} \bigcup_{r\in R} Y_r(m)$: An OR-event that the state of at least one path in a set R ($\subset$ ***Path***) of paths is $m \in M$.
- $\overline{Z}(R)(m) \stackrel{\text{def}}{=} \bigcap_{r\in R} Y_r(m)$: An AND-event that the state of every path in a set R ($\subset$ ***Path***) of paths is $m \in M$.

First we assume $\mathcal{V} \stackrel{\text{def}}{=} \{V(L)(m) | L \subset \Delta^*, L \neq \emptyset, m \in M\}$ satisfies the following conditions:

$$V(L)(i) \neq \emptyset \quad \text{and} \quad V(L)(i) \cap V(L)(j) = \emptyset \quad \text{if } i \neq j \tag{3.3}$$

Furthermore for $\mathcal{V}$, we assume that there exists $\{S(m) \subset M^2 | m \in M\}$ such that:

$$V(L + L')(m) = \sum_{s\in S(m)} V(L)(s_1) \cap V(L')(s_2) \qquad \text{if } L \cap L' = \emptyset,\ L, L' \neq \emptyset \tag{3.4}$$

$$(s_1, s_2) \in S(m) \quad \Rightarrow \quad s_1, s_2 \leq m \tag{3.5}$$

$$T(m, s) \in \{T(m', s') | 0 \leq m' < m, 0 \leq s' < m\} \qquad \text{if } 1 \leq s \leq m \tag{3.6}$$

where $T(m, s) \stackrel{\text{def}}{=} \{s' | (s, s') \in \sum_{k=0}^{m} S(k)\}$.

For each $k \in \{2, 3, ...\}$, and for each $m \in M$, we define a set $S_k(m)$ of k-tuples of states in the following way:

$$\begin{aligned} S_2(m) &\stackrel{\text{def}}{=} S(m) \\ S_k(m) &\stackrel{\text{def}}{=} \{(\boldsymbol{s}'', s'_2) | \boldsymbol{s}'' \in S_{k-1}(s'_1), (s'_1, s'_2) \in S(m)\} \\ &= \{(s'_1, \boldsymbol{s}'') | \boldsymbol{s}'' \in S_{k-1}(s'_2), (s'_1, s'_2) \in S(m)\} \end{aligned} \tag{3.7}$$

Then we have:

$$V(\sum_{j=1}^{k} L_j)(m) = \sum_{\boldsymbol{s} \in S_k(m)} \bigcap_{j=1}^{k} V(L_j)(s_j) \tag{3.8}$$

for any $\{L_j | 1 \leq j \leq k,\ L_j \neq \emptyset\}$ satisfying that $i \neq j \Rightarrow L_i \cap L_j = \emptyset$. Especially, for $L = \{l_1, ..., l_k\}$, $V(L)(m) = \sum_{\boldsymbol{s} \in S_k(m)} \bigcap_{j=1}^{k} X(l_j)(s_j)$.

Conditions (3.3) – (3.6) imply that $\mathcal{V}$ (i.e. $\{S(m) | m \in M\}$) has the following properties:

$$(i, j) \in S(m) \Leftrightarrow (j, i) \in S(m) \tag{3.9}$$

$$i \neq j \Rightarrow S(i) \cap S(j) = \emptyset \tag{3.10}$$

$$V(L + L')(0) = V(L)(0) \cap V(L')(0) \text{ i.e. } S(0) = \{(0, 0)\} \tag{3.11}$$

$$m > 0 \Rightarrow (m, m) \notin S(m) \tag{3.12}$$

$$(0, m), (m, 0) \in S(m),\ (0, i), (i, 0) \notin S(m)\ (i < m) \tag{3.13}$$

$$(m, i), (i, m) \notin S(m)\ (0 < i) \tag{3.14}$$

$$1 \leq i \leq m \Rightarrow \exists m' < m\ (T(m, i) = T(m', 0)) \tag{3.15}$$

$$T(m, 0) = \{0, 1, .., m\} \tag{3.16}$$

$$T(m, m) = \{0\} \tag{3.17}$$

$$V(L_1)(0) \cap V(L_2)(0) = V(L_1 \cup L_2)(0) \tag{3.18}$$

$$V(L_1)(m) \cap V(L_2)(0) = V(L_1 \setminus L_2)(m) \cap V(L_2)(0) \tag{3.19}$$

Finally, we assume two conditions of V related to occurrence probabilities:

$$0 < \Pr[V(L)(0)] \quad \text{for } L \subset \Delta^* \tag{3.20}$$

$$\begin{aligned} &V(L)(i) \text{ and } V(L')(j) \text{ are independent} \\ &\text{if} \quad L, L' \neq \emptyset,\ L \cap L' = \emptyset,\ i, j \in M \end{aligned} \tag{3.21}$$

3.2 Examples

In what follows, we provide some examples of characteristics which can be treated in our model and satisfy the above conditions.

(Packet loss)

Note that "No loss on a path" $\Leftrightarrow$ "No loss on every link in a path". Thus we interpret –

- X_l is an event that no loss occurs on a link l. Since $M = \{0\}$, we omit the parameter (0) in this case.
- $V(L)$ is defined by the relation that $V(L+L') = V(L) \cap V(L')$ (if $L \cap L' = \emptyset$) so that conditions (3.3) – (3.6) are held.
- Y_r is an event that a packet traversing a path r reaches the end.

If all packets get lost on a link, we cannot calculate the characteristics beyond that link. Thus we should remove paths r from ***Path*** on which $\Pr[Y_r] = 0$ so that condition (3.20) of V is held. Furthermore, we expect that condition (3.21) is loosely held in a large and diverse network where a number of independent traffic flows pass through internal-links (routers) and cause packet loss and queuing delay on them.

(Queuing delay)

Note that "A delay on a path" = "A sum of delays on every link in a path". In addition, we regard a delay as a discrete value, and define an unit time of delay. Thus we interpret –

- X_l is an event that a queuing delay on a link l is $m \in M = \{0, 1, .., max\}$.
- $V(L)(m)$ is defined by the relation that $V(L + L')(m) = \sum_{j=0}^{m} V(L)(j) \cap V(L')(m-j)$ (i.e. $S(m) = \{(s_1, s_2) | s_i \in M, s_1 + s_2 = m\}$) so that conditions (3.3) – (3.6) are held.
- $Y_r(m)$ is an event that a packet traversing a path r experiences a queuing delay m.

Since a total delay that a packet experiences on a link mainly consists of $(\textit{propagation time}) + (\textit{packet size})/(\textit{bandwidth}) + (\textit{queuing delay})$, we can extract a queuing delay of each path from a series of observed total packet delays on that path.

Since each router is expected not to be always in congestion, "queuing delay" is likely to sometimes 0. Thus condition (3.20) of V is held in such networks. Furthermore, we expect that condition (3.21) is loosely held because of the same reason as packet loss.

Note that we should choose the discrete unit of delay time by carefully considering rounding errors and computing costs in inference.

It should be noted that our model and method have potential possibilities to be applied to other kind of inference problems such as, for example, inferring an OD traffic matrix. Origin-destination (OD) traffic matrix inference problem is to infer "traffic counts" (the amount of traffic in a fixed time interval) of a directed flow between each OD pairs from traffic counts of directed links measured at routers' incoming and outgoing interfaces [15].

Assume, for each OD flow, we can know (identify) the set of router interfaces which the OD flow traverses. Then we regard each router interface as a "path" (in terms of our model), and each OD flow as a "link" (in terms of our model). Since "traffic counts on a router interface" = "A sum of traffic counts on every OD flow passing a router interface", we can treat it in our model, and thus, if some conditions of traffic to be inferred are satisfied, we can apply our inference method to this problem.

4 Method of inferring characteristics of links

In this section, we explain our inference method (calculations) on a general model we present in the previous sections.

Assume ***Path*** (all paths), Δ^* (all links), X_l (events on a link l) for each $l \in \Delta^*$, Y_r (events on a path r) for each $r \in$ ***Path*** and V (a relation between them) are given. Then, for each $m \in M$,

$$\begin{aligned}
\Psi &\stackrel{\text{def}}{=} \{R \subset \boldsymbol{Path} | \overline{L}(R) \neq \emptyset\} && (3.22)\\
\alpha_l(m) &\stackrel{\text{def}}{=} \Pr[X_l(m)] \\
\boldsymbol{\alpha}(m) &\stackrel{\text{def}}{=} \{\alpha_l(i) | 0 \leq i \leq m,\ l \in \Delta^*\} \\
\underline{\gamma}(R)(m) &\stackrel{\text{def}}{=} \Pr[\bigcup_{k=0}^{m} \underline{Z}(R)(k)] && (3.23)\\
\underline{\gamma}(m) &\stackrel{\text{def}}{=} \{\underline{\gamma}(R)(i) | 0 \leq i \leq m,\ R \in \Psi\} \\
\overline{\gamma}(R, R')(m) &\stackrel{\text{def}}{=} \Pr[\overline{Z}(R - R')(0) \cap \overline{Z}(R')(m)] && (3.24)\\
\overline{\gamma}(m) &\stackrel{\text{def}}{=} \{\overline{\gamma}(R, R')(i) | 0 \leq i \leq m, R' \subset R, R' \neq R, |R| > 1, R \in \Psi\} \\
\boldsymbol{\gamma}(m) &\stackrel{\text{def}}{=} \underline{\gamma}(m) \cup \overline{\gamma}(m) && (3.25)
\end{aligned}$$

Under conditions (3.3), (3.4), (3.5) and (3.21) of V, we can calculate $\boldsymbol{\gamma}(m)$ from $\boldsymbol{\alpha}(m)$ as polynomial expressions of $\boldsymbol{\alpha}(m)$. We denote this map from $\boldsymbol{\alpha}(m)$ to $\boldsymbol{\gamma}(m)$ by Γ_m. Note that $0 \leq \alpha_l(m)$, and $\sum_{i=0}^{m} \alpha_l(i) \leq 1$ for $l \in \Delta^*$ and $m \in M$. In addition, we assume condition (3.20) so that $0 < \alpha_l(0)$ for $l \in \Delta^*$.

Let $d_m \stackrel{\text{def}}{=} |\boldsymbol{\alpha}(m)|$ and $e_m \stackrel{\text{def}}{=} |\boldsymbol{\gamma}(m)|$. We can regard $\boldsymbol{\alpha}(m)$ and $\boldsymbol{\gamma}(m)$ as a point in $[0,1]^{d_m}$ and in $[0,1]^{e_m}$ by arranging elements of $\boldsymbol{\alpha}(m)$ and of $\boldsymbol{\gamma}(m)$ in a fixed order, respectively. We denote the domain in which $\boldsymbol{\alpha}(m)$ ($\boldsymbol{\gamma}(m)$) exists by $\mathcal{D}_m$ ($\mathcal{E}_m$, respectively).

$$\Gamma_m : \mathcal{D}_m \to \mathcal{E}_m \quad \text{such that } \boldsymbol{\gamma}(m) = \Gamma_m(\boldsymbol{\alpha}(m)) \tag{3.26}$$

$$\begin{aligned}
\mathcal{D}_m &\stackrel{\text{def}}{=} \{(x_{k,i})_{1 \leq k \leq |\Delta^*|, 0 \leq i \leq m} \in [0,1]^{d_m} | \ 0 < x_{k,0},\ 0 \leq x_{k,i},\ \sum_{i=0}^{m} x_{k,i} \leq 1\} \\
\mathcal{E}_m &\stackrel{\text{def}}{=} \Gamma_m(\mathcal{D}_m) \subset [0,1]^{e_m}
\end{aligned}$$

Our goal of this section is to find a subspace E'_m of $[0,1]^{e_m}$ and an inverse map $(\Gamma_m|_{E'_m})^{-1}$ of $\Gamma_m|_{E'_m}$ which is the projection of Γ_m on E'_m:

$$\alpha(m) = (\Gamma_m|_{E'_m})^{-1}(\gamma(m)|_{E'_m}) \tag{3.27}$$

We present our main results. Proofs and calculation algorithms are shown in Appendix. If conditions (3.3), (3.4), (3.5), (3.6), (3.20) and (3.21) of V are satisfied, then the following Lemmas, and thus Theorem 1 are held. We denote an occurrence probability of an event $V(\overline{L}(R))(m)$ by:

$$A_R(m) \stackrel{\text{def}}{=} \Pr[V(\overline{L}(R))(m)] \quad \text{for } R \in \Psi, m \in M.$$

[Lemma 1]

For each set R of paths satisfying that the MIP number of R is not equal to 1, $A_R(m)$ $(m \in M)$ is uniquely determined as follows: Let the MIP number of R be N,

1. If $N = 0$, then $\{r\} = R$, and $A_R(m)$ is calculated from an occurrence probability of an event on a path r: $A_R(m) = \Pr[Y_r(m)]$.
2. If $N \geq 2$, then there exists a family $\{R_0, R_1, .., R_N\}$ of subsets of R satisfying that $A_R(m)$ is calculated from occurrence probabilities of OR-events on them: $\{\Pr[\bigcup_{k=0}^{i} \underline{Z}(R_j)(k)] | 0 \leq i \leq m,\ 0 \leq j \leq N\}$

[Lemma 2]

For each set R of paths satisfying that the MIP number of R is not equal to 0, there exists R_0 satisfying that $R_0 \subset R$, $R_0 \neq R$, and $A_R(m)$ $(m \in M)$ is calculated from:

- Occurrence probabilities of AND-events on R, $R - R_0$: $\Pr[\overline{Z}(R-R_0)(0)]$, $\{\Pr[\overline{Z}(R-R_0)(0) \cap \overline{Z}(R_0)(i)] | 0 \leq i \leq m\}$ and,
- $\{A_{R'}(i) | 0 \leq i \leq m,\ R_0 \subset R' \subset R, R' \neq R\}$.

[Lemma 3]

For each link l_R $(R \in \Delta)$, an occurrence probability $\Pr[X_{l_R}(m)]$ $(m \in M)$ is calculated from: $\{A_{R'}(i) | 0 \leq i \leq m,\ R \subset R',\ R' \in \Delta\}$

We present Theorem 1. We use notations from (3.22) to (3.26). Combining Lemma 1, 2 and 3, for each $R \in \Delta$ and $m \in M$, $\alpha_{l_R}(m)$ can be determined from a subset of $\gamma(m)$, via $\{A_{R'}(i) | 0 \leq i \leq m,\ R \subset R',\ R' \in \Delta\}$.

Since, for each R satisfying that $\overline{L}(R) \neq \emptyset$, there may exist R_0 satisfying that $\overline{L}(R) = \overline{L}(R_0)$ and (the MIP number of R) $\neq$ (the MIP number of R_0), we have more than one way to calculate $A_R(i)$ by using Lemma 1 and 2 in general. Note that if $R \in \Delta$, then $\exists R_0 \subset R\ \forall l_{R'} \in \underline{L}(R) - \overline{L}(R)\ (R_0 \setminus R' \neq \emptyset) \Leftrightarrow \exists R_0 \neq R\ (\overline{L}(R) = \overline{L}(R_0))$.

[Theorem 1]

Γ_m is injection. There exist two subsets Ψ_1 and Ψ_2 of Ψ, a subspace $E_m^{\Psi_1,\Psi_2}$ of $[0,1]^{e_m}$ and a map $\Gamma^{-1}_{m,\Psi_1,\Psi_2}$ which is defined in a neighborhood of the domain $\mathcal{E}_m|_{E_m^{\Psi_1,\Psi_2}}$ satisfying that:

$$\begin{aligned}\boldsymbol{\alpha}(m) &= \Gamma^{-1}_{m,\Psi_1,\Psi_2}(\boldsymbol{\gamma}(m)|_{E_m^{\Psi_1,\Psi_2}}) \\ \underline{\boldsymbol{\gamma}}^{\Psi_1}(m) \cup \overline{\boldsymbol{\gamma}}^{\Psi_2}(m) &= \boldsymbol{\gamma}(m)|_{E_m^{\Psi_1,\Psi_2}}\end{aligned}$$

where we regard $\boldsymbol{\alpha}(m)$ and $\boldsymbol{\gamma}(m)$ as a point in $[0,1]^{d_m}$ and in $[0,1]^{e_m}$, respectively, and we choose a proper $R_0 \subset R$ for each $R \in \Psi_2$:

$$\begin{aligned}\underline{\boldsymbol{\gamma}}^{\Psi_1}(m) &\stackrel{\text{def}}{=} \{\underline{\gamma}(R)(i) \mid 0 \le i \le m,\ R \in \Psi_1\} \\ \overline{\boldsymbol{\gamma}}^{\Psi_2}(m) &\stackrel{\text{def}}{=} \{\overline{\gamma}(R, R_0)(i) \mid 0 \le i \le m,\ R \in \Psi_2\}\end{aligned}$$

Moreover, $\Gamma^{-1}_{m,\Psi_1,\Psi_2}$ is continuously differentiable.

In what follows, since our calculations include an extension of MINC's inference method, we present a brief review of MINC's method and the relation to our method. We consider a tree path-topology like (I) of Fig. 3.1, and end-to-end probe packets from a root sender to each leaf receiver.

Let 0 be a root node (sender), $r_{i,j}$ be a partial path from node i to j, e_j be a directed link from node j' to j (j' is the parent of j), and R_j be a set of leaf nodes (receivers) descended from j. $r_{0,j}$ is a (end-to-end) path if $j \in R_0$.

Let k be an internal node in a tree. A partial path $r_{0,k}$ can be regarded as a shared-link set of paths $\{r_{0,j} | j \in R_k\}$. Inference calculations are performed by two steps:

1. Calculating characteristic on each shared-link set $r_{0,k}$ from characteristics on paths in $\{r_{0,j} | j \in R_k\}$.
2. Calculating characteristic on each link e_k from characteristics on $r_{0,k}$ and $r_{0,k'}$ (k' is the parent of k).

For example, consider packet loss in (I) of Fig. 3.1. In the above notations, path a, b and c are $r_{0,1}$, $r_{0,2}$ and $r_{0,3}$ respectively. Let p and q be the first (upper) and the second (lower) internal node in a path from 0 respectively. Thus $R_p = \{1,2,3\}$, $R_q = \{1,2\}$, and e_q is a directed link from p to q.

We denote an event that a packet passes successfully (no loss occurs) through a path $r_{0,j}$ (a partial path $r_{i,j}$) by Y_j ($V_{i,j}$, respectively). Moreover we also denote an event that no loss occurs on a link e_k by X_k.

Let $Z_p = \bigcup_{j \in R_p} Y_j$: an event that a packet passes successfully through at least one of three paths $r_{0,1}$, $r_{0,2}$ and $r_{0,3}$. Similarly let $Z_q = \bigcup_{j \in R_q} Y_j$, $W_q = \bigcup_{j \in R_q} V_{p,j}$. Then, since we can assume $V_{0,p}$, W_q and $V_{p,3}$ are independent, there exists the following relation among occurrence probabilities of events.

$$\begin{aligned}
\Pr[Z_p] &= \Pr[V_{0,p} \cap (W_q \cup V_{p,3})] = \Pr[V_{0,p}](\Pr[W_q] + \Pr[V_{p,3}] - \Pr[W_q]\Pr[V_{p,3}]) \quad (3.28)\\
\Pr[Z_q] &= \Pr[V_{0,p}]\Pr[W_q]\\
\Pr[Y_3] &= \Pr[V_{0,p}]\Pr[V_{p,3}]
\end{aligned}$$

where the left hand sides of the equations are occurrence probabilities of states of paths which can be inferred from end-to-end measurements. Therefore we can obtain $\Pr[V_{0,p}]$ (an occurrence probability of states of a shared-link set $r_{0,p}$) by solving the equations. Similarly we can obtain $\Pr[V_{0,q}]$. Then we calculate $\Pr[X_q] = \Pr[V_{0,q}]/\Pr[V_{0,p}]$.

Next we extend it to general path-topologies. Note that no explicit information about orders of links in a path is used in calculating the above occurrence probabilities. This implies that the calculation is based upon a combination of paths and links, but is not on a path-topology. Therefore we formalize it to a lattice Δ.

Key equation (3.28) in the previous example uses an occurrence probability of an OR-event on a set of paths from 0 to descendants of p, and lies on the fact that every internal node k has more than one child node. In other words, a set of partial paths from k to descendants can be divided into more than one independent path group. Since this division is clear if Δ is a tree, MINC's calculation can be extended in a straightforward way where Δ is a tree such as (I) of Fig. 3.2, even if the path-topology is not a tree.

Furthermore, even if Δ is not a tree such as (II) or (III) of Fig. 3.2, by introducing the MIP number of R, we can extend the calculation of A_R when the MIP number is not 1 (Lemma 1).

Finally we need to develop a different principle of calculation for the case when the MIP number is 1 (Lemma 2). This calculation requires occurrence probabilities of AND-events on some sets of paths, and can extract occurrence probabilities of events on links from mutually overlapped (more than two) paths. Theoretically, since every A_R can be calculated by using only 1. of Lemma 1 and Lemma 2, 2. of Lemma 1 is not necessary. However, in actual inference problems where $\gamma(m)$ are given by end-to-end measurements, we may use 2. of Lemma 1 if it is applicable because these OR-event based method and AND-event based method may be different in properties of inference such as efficiency and accuracy.

5 Case studies

5.1 Round-trip measurements

We estimate network-internal characteristics by measuring traffic on round trip paths from clients to servers in (I) of Fig. 3.3.

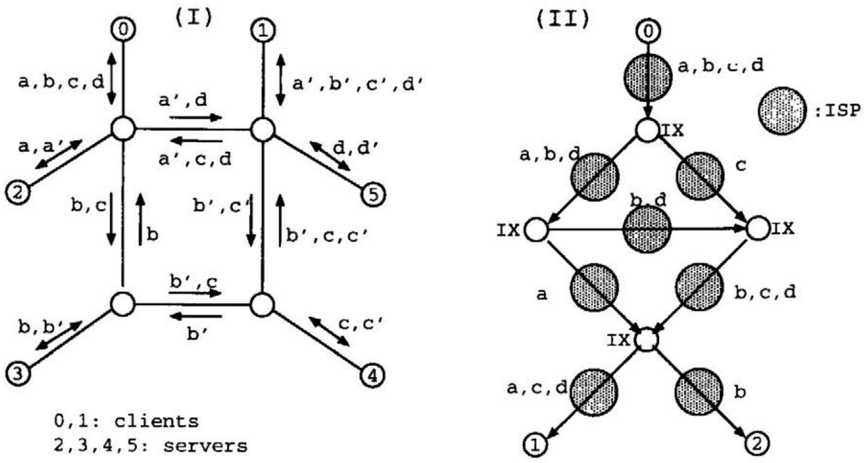

Figure 3.3: Case studies

We assume only clients are under our management, thus we cannot measure characteristics of servers and physical links directly. The purpose of this measurements is quantitative estimations of performance responsibilities for each servers (including physical leaf links) and physical backbone links.

- node $0, 1$ are clients, and node $2, 3, 4$ and 5 are servers. A client sends a packet and then receives its reply packet by a server. Other unnumbered nodes (circles) indicate routers.
- We regard a round-trip route as a path. Some packets are sent by node 0 to nodes $2, 3, 4, 5$ and, in reply to them, packets go back to node 0 through paths a, b, c, d, respectively. Others are sent by node 1 to nodes $2, 3, 4, 5$ and their reply packets go back to node 1 through paths a', b', c', d', respectively, as described in (I) of Fig. 3.3.

Note that each leaf (logical) link includes two physical leaf links (both directions) so that we cannot distinguish characteristic of each direction, although it is not fatal for the measurement purpose in this case.

5.2 One way measurements

We estimate each ISP's characteristics from one way measurements on paths across many ISPs in (II) of Fig. 3.3.

We assume only end-nodes (a sender and receivers) are under our management, thus we cannot measure characteristics of each ISP directly. The purpose of this measurements is quan-

Table 3.1: Shared link set and the MIP number

(I) Round trip measurements

link	shared-link set	MIP #	MIP div.
l_{abcd}	l_{abcd}	3	$\{a\}, \{b\}, \{d\}$
$l_{a'b'c'd'}$	$l_{a'b'c'd'}$	3	$\{a'\}, \{b'\}, \{d'\}$
$l_{a'cd}$	$l_{a'cd}$	2	$\{a'\}, \{c\}$
$l_{b'cc'}$	$l_{b'cc'}$	1	–
$l_{aa'}$	$l_{aa'}$	2	$\{a\}, \{a'\}$
$l_{a'd}$	$l_{a'd}, l_{a'cd}$	2	$\{a'\}, \{d\}$
l_{bc}	l_{bc}, l_{abcd}	2	$\{b\}, \{c\}$
$l_{bb'}$	$l_{bb'}$	2	$\{b\}, \{b'\}$
$l_{b'c'}$	$l_{b'c'}, l_{b'cc'}, l_{a'b'c'd'}$	2	$\{b'\}, \{c'\}$
$l_{b'c}$	$l_{b'c}, l_{b'cc'}$	2	$\{b'\}, \{c\}$
$l_{cc'}$	$l_{cc'}, l_{b'cc'}$	2	$\{c\}, \{c'\}$
$l_{dd'}$	$l_{dd'}$	2	$\{d\}, \{d'\}$
l_b	$l_b, l_{bb'}, l_{bc}, l_{abcd}$	0	–
$l_{b'}$	$l_{b'}, l_{bb'}, l_{b'c}, l_{b'c'}, l_{bcc'}, l_{a'b'c'd'}$	0	–

(II) One way measurements

link	shared-link set	MIP #	MIP div.
l_{abcd}	l_{abcd}	1	–
l_{abd}	l_{abd}, l_{abcd}	2	$\{a\}, \{b\}$
l_{acd}	l_{acd}, l_{abcd}	2	$\{a\}, \{c\}$
l_{bcd}	l_{bcd}, l_{abcd}	2	$\{b\}, \{c\}$
l_{bd}	$l_{bd}, l_{bcd}, l_{abcd}$	2	$\{b\}, \{d\}$
l_a	$l_a, l_{abd}, l_{acd}, l_{abcd}$	0	–
l_b	$l_b, l_{bd}, l_{abd}, l_{bcd}, l_{abcd}$	0	–
l_c	$l_c, l_{bcd}, l_{acd}, l_{abcd}$	0	–

titative estimations of performance responsibilities for each ISP on end-to-end paths. Therefore we regard a transit route through an ISP as a "link".

- node 0 is a source (sender) and nodes 1 and 2 are destination (receiver) nodes. Each IX (Internet eXchange) node is an exchange-point among ISPs.
- Some packets are sent from node 0 to 1 through paths a, c and d, respectively. Others are sent from node 0 to 2 through path b, as described in (II) of Fig. 3.3.

Finally we show concrete expressions to calculate no-loss probability of each link in this example.

In order to use our inference method, we estimate $\underline{\gamma}_R$ (a probability that no loss occurs along at least one of path in R) or, in some case, $\overline{\gamma}_R$ (a probability that no loss occurs along all path in R) for some Rs, from end-to-end measurements. Then we calculate A_R (a probability that no loss occurs on all links in the $\overline{\mathrm{L}}(R)$) from them, and finally calculate α_R (a probability that no loss occurs on the link l_R) for each $R \in \Delta$. In this example, since the MIP number of $\{a, b, c\}$ is equal to 1, we use $\overline{\gamma}_{abc}$ instead of $\underline{\gamma}_{abc}$ in order to calculate A_{abc} ($= A_{abcd}$).

For example, we have $A_a = \gamma_a$, $A_b = \gamma_b$, $A_c = \gamma_c$, $A_{bd} = (\gamma_b\gamma_d)/(\gamma_b + \gamma_d - \underline{\gamma}_{bd})$, $A_{abd} = A_{ab} = (\gamma_a\gamma_b)/(\gamma_a+\gamma_b-\underline{\gamma}_{ab})$, $A_{acd} = A_{ac} = (\gamma_a\gamma_c)/(\gamma_a+\gamma_c-\underline{\gamma}_{ac})$, $A_{bcd} = A_{bc} = (\gamma_b\gamma_c)/(\gamma_b+\gamma_c - \underline{\gamma}_{bc})$. And then,

$$A_{abcd} = A_{abc} = \frac{\overline{\gamma}_{abc}A_{ab}A_{ac}}{\overline{\gamma}_{bc}A_a} = \frac{\overline{\gamma}_{abc}\gamma_a\gamma_b\gamma_c}{\overline{\gamma}_{bc}(\gamma_a + \gamma_b - \underline{\gamma}_{ab})(\gamma_a + \gamma_c - \underline{\gamma}_{ac})}$$

Consequently,

$$\alpha_{abcd} = \frac{\overline{\gamma}_{abc}\gamma_a\gamma_b\gamma_c}{\overline{\gamma}_{bc}(\gamma_a + \gamma_b - \underline{\gamma}_{ab})(\gamma_a + \gamma_c - \underline{\gamma}_{ac})}, \quad \alpha_{abd} = \frac{\overline{\gamma}_{bc}(\gamma_a + \gamma_c - \underline{\gamma}_{ac})}{\gamma_c\overline{\gamma}_{abc}}$$
$$\alpha_{acd} = \frac{\overline{\gamma}_{bc}(\gamma_a + \gamma_b - \underline{\gamma}_{ab})}{\gamma_b\overline{\gamma}_{abc}}, \quad \alpha_{bcd} = \frac{\overline{\gamma}_{bc}(\gamma_a + \gamma_b - \underline{\gamma}_{ab})(\gamma_a + \gamma_c - \underline{\gamma}_{ac})}{\gamma_a\overline{\gamma}_{abc}(\gamma_b + \gamma_c - \underline{\gamma}_{bc})}$$
$$\alpha_{bd} = \frac{\gamma_d\overline{\gamma}_{abc}(\gamma_b + \gamma_c - \underline{\gamma}_{bc})}{\overline{\gamma}_{bc}(\gamma_b + \gamma_d - \underline{\gamma}_{bd})(\gamma_a + \gamma_c - \underline{\gamma}_{ac})}; \quad \alpha_a = \frac{\overline{\gamma}_{abc}}{\overline{\gamma}_{bc}}$$
$$\alpha_b = \frac{\gamma_b + \gamma_d - \underline{\gamma}_{bd}}{\gamma_d}, \quad \alpha_c = \frac{\overline{\gamma}_{abc}(\gamma_b + \gamma_c - \underline{\gamma}_{bc})}{\overline{\gamma}_{bc}(\gamma_a + \gamma_b - \underline{\gamma}_{ab})}$$

Note that in this example, due to a special relation that $\gamma_a + \gamma_b - \underline{\gamma}_{ab} = \overline{\gamma}_{ab}$, above expressions can be expressed more concisely.

6 Discussions and future work

6.1 Conditions of V

We discuss the conditions of V (a relation between path events and link events). Condition (3.21) is "assumption of independence of all links". In actual networks, links are not usually independent each other so that this fact can cause errors in inference. Nevertheless, this assumption simplifies inference calculations, and for this ease, we first employ it here.

Note that if the co-relation among links is very weak, the error will be negligible. Furthermore, if we know (estimate) the degree of co-relation in some way, we can correct inference calculations. Some analytical and simulation results regarding packet loss rates and queing delay distributions on tree path-topologies have been presented ([4], [16] and [17]).

Special properties of state-0, condition (3.20) and property (3.11) derived from (3.5), are essential for our inference. As an example of V which violates these properties, we consider:

$$V(L+L')(0) = V(L)(0) \cap V(L')(0) + V(L)(1) \cap V(L')(1)$$
$$V(L+L')(1) = V(L)(0) \cap V(L')(1) + V(L)(1) \cap V(L')(0)$$

where V can be regarded as a model for "toggling a flag (ON/OFF) in a packet".

- The flag in a packet is OFF at source nodes.
- The flag may be toggled when it passes a link.
- Can we infer the occurrence probabilities of toggle on each link from observing packets at destination nodes?

Unfortunately we can show some cases in which the solution is not unique.

6.2 Multi-routes

Our method can treat static multi-routes, by which we mean that a router decides the next physical link for a received packet by using not only a destination IP address but some static information like policy routing. Although there are more than one partial path from a router to a destination, and more than one path from a source to a destination in such a case, for each observed packet, we can identify the path which the packet traverses by its information such as its source IP address, port numbers, MAC address, and so on.

Note that, in Section III, we consider a general event form $V(L)(m)$ which is parameterized by a state m and a set L of links, thus this formalization implicitly assumes that we don't mention a relation between a statistical characteristic of each set of links and an information on the types of packet passes through the links. In other words, if some link behaves differently for packets of different types (for example, if a router gives a priority to packets coming from some special source nodes), and if we need to distinguish the difference, we consider there exist two (virtually different) links, and need to determine which link is passed by each traversing packet.

6.3 Measurements of end-to-end path characteristics

Finally we mention the relation between inference of characteristics of paths by monitoring end-to-end traffic on the paths and our method of determining characteristic of each link from the inferred characteristics of the paths related to the link.

In general, recalling Section IV, let $\Gamma : \boldsymbol{\gamma} = \Gamma(\boldsymbol{\alpha})$ be the map from "occurrence probabilities of events on each link" ($\boldsymbol{\alpha} \in \mathcal{D} \subset [0,1]^{|\boldsymbol{\alpha}|}$) to "occurrence probabilities of events on each set of paths" ($\gamma \in \mathcal{E} \subset [0,1]^{|\gamma|}$).

We have shown that there exists a subspace E' of $[0,1]^{|\gamma|}$ and $\Gamma'^{-1} : \mathcal{E}|_{E'} \to \mathcal{D}$ satisfying $\boldsymbol{\alpha} = \Gamma'^{-1}(\boldsymbol{\gamma'})$ where Γ' is the projection of Γ onto E' such that $\boldsymbol{\gamma'} = \Gamma'(\boldsymbol{\alpha})$.

Then, in order to infer $\boldsymbol{\alpha}$, what we have to do is to obtain $\boldsymbol{\gamma'}$. We consider the case in which we infer $\boldsymbol{\gamma'}$ from n independent end-to-end path measurements by "Strong law of large numbers", and denote the inferred value by $\hat{\gamma}'_n$. That law tells us $\hat{\gamma}'_n$ converges to γ' almost surely as $n \to \infty$. Then, let $\hat{\boldsymbol{\alpha}}_n \stackrel{\text{def}}{=} \Gamma'^{-1}(\hat{\gamma}'_n)$, we expect that $\hat{\boldsymbol{\alpha}}_n$ converges to $\boldsymbol{\alpha}$ almost surely. However, since Γ'^{-1} is well defined only when a input value is in $\mathcal{E}'$, we cannot define $\hat{\boldsymbol{\alpha}}_n$ when $\hat{\gamma}'_n$ approximates γ' from outside of $\mathcal{E}'$.

Therefore γ' should be restricted in an open set $\mathcal{E}'_0 \subset \mathcal{E}'$. We should do it by removing some paths from ***Path*** if necessary. Under this condition, since $\hat{\gamma}'_n \to \gamma'$ a.s. $(n \to \infty)$, for sufficiently large n, $\hat{\gamma}'_n$ stays in $\mathcal{E}'_0$. Hence $\hat{\boldsymbol{\alpha}}_n \to \boldsymbol{\alpha}$ a.s. if $n \to \infty$ because of continuity of Γ'^{-1}. Moreover, some analysis of convergence rate can be done by means of continuous differentiability.

In inferring characteristics from actual measurements, there exist the following three fundamental problems related to three different time-scales. They lead to errors or decreasing convergence rate in inference, and strongly depend on concrete characteristics and path-topologies rather than abstract models. Therefore we need to study practical improvements to overcome these errors in actual inference problems, by using end-to-end measurement techniques and statistical error (bias) corrections.

- Concurrent events on paths

  In our inference method, we use occurrence probabilities of concurrent events (OR-events, AND-events) on paths. It is the most important and the most difficult part to estimate them accurately from observations of end-to-end traffic.

  Especially in using "unicast" traffic, we need more than one packet which traverse the set of links shared by these paths within a trial in which we assume the links' states do not change. However, the numbers of paths on a concurrent event increasing, a series of packets in a trial may meet changes of state of links.

- Independence among each observations

  The fact that $\hat{\gamma}'_n$ converges to γ' almost surely as $n \to \infty$ relies on the assumption of independence among each observations. However, actual networks have temporal dependencies like heavy-tailed behaviors of the Internet traffic.

- Stationarity of links

  When we infer occurrence probabilities of events on each link, we assume there exists a stable probabilistic model on links. Since long-term observations may violate this assumption, we should complete a set of observations within a proper time-scale for consistent inferring.

Finally, for efficient and accurate inference, it is of practical importance to find a set of paths to be observed that is suitable for a given set of links to be inferred. For a given physical link e, we can automatically enumerate all sets of paths satisfying that e is "distinguishable" in the set of paths in our model. However, from the above candidates, we need to select a suitable set of paths based not only on the size of the set (the number of paths be observed) but on various practical factors such as the topology and configuration of a network or occurrence probabilities of observed events on the network dependent on end-to-end measurement techniques.

7 Conclusion

In this work, we have proposed a general model and method of determining characteristic (i.e. occurrence probabilities of states) of each internal-link from given characteristics of end-to-end paths. They can be applied to various characteristics of links on an arbitrary path-topology. Some examples of such characteristics, and case studies of network (path) topologies have been shown.

Our model and method are so general that they can give a unified viewpoint to various practical methods of inferring network-internal characteristics, and thus can give useful insights into advantages and disadvantages of such inference methods.

8 Appendix (proofs of Lemmas and Theorems)

[Proof of Lemma 1]

Let the MIP number of R be N. If $N = 0$, then $\{r\} = R$ and $\overline{L}(R) = L_r$. Therefore $A_R(m) = \Pr[V(\overline{L}(R))(m)] = \Pr[V(L_r)(m)] = \Pr[Y_r(m)]$.

If $N \geq 2$, then $A_R(m)$ can be calculated as follows: Let $K_r \stackrel{\text{def}}{=} L_r - \overline{L}(R)$, R_0 be $\{r \in R | K_r \neq \emptyset\}$, and $\{R_1, .., R_N\}$ be the MIP division of R_0.

Furthermore we define $W_j(i,s) \stackrel{\text{def}}{=} \bigcup_{r\in R_j} \sum_{s' \in T(i,s)} V(K_r)(s')$ for $0 \leq i \leq m$ where $T(i,s)$ is defined by (3.6). We have:

$$\begin{aligned}
\bigcup_{i=0}^{m} \underline{Z}(R_0)(i) &= \bigcup_{i=0}^{m} \bigcup_{r\in R_0} V(\overline{L}(R) + K_r)(i) \\
&= \bigcup_{j=1}^{N} \bigcup_{r\in R_j} \sum_{i=0}^{m} \sum_{s\in S(i)} \left(V(\overline{L}(R))(s_1) \cap V(K_r)(s_2)\right) \\
&= \sum_{s=0}^{m} \left(V(\overline{L}(R))(s) \cap \bigcup_{j=1}^{N} W_j(m,s)\right) \\
\bigcup_{i=0}^{m} \underline{Z}(R_j)(i) &= \bigcup_{i=0}^{m} \bigcup_{r\in R_0} V(\overline{L}(R) + K_r)(i) \\
&= \sum_{s=0}^{m} \left(V(\overline{L}(R))(s) \cap W_j(m,s)\right) \qquad 1 \leq j \leq N
\end{aligned}$$

Note that we assume condition (3.21). By the definition of the MIP number, $(\bigcup_{r\in R_j} K_r) \cap (\bigcup_{r\in R_{j'}} K_r) = \emptyset$ $(j \neq j')$, and thus $W_j(m,s)$ and $W_{j'}(m,s)$ are independent. Similarly, because of $\overline{L}(R) \cap K_r = \emptyset$, $V(\overline{L}(R))(s)$ and $W_j(m,s)$ are independent. Thus we have:

$$\Pr[\bigcup_{i=0}^{m} \underline{Z}(R_0)(i)] = \sum_{s=0}^{m} A_R(s)(1 - \prod_{j=1}^{N}(1 - \Pr[W_j(m,s)])) \tag{3.29}$$

$$\Pr[\bigcup_{i=0}^{m} \underline{Z}(R_j)(i)] = \sum_{s=0}^{m} A_R(s) \Pr[W_j(m,s)] \quad 1 \leq j \leq N \tag{3.30}$$

Let $A(i) \stackrel{\text{def}}{=} A_R(i)$, $B_j(i,s) \stackrel{\text{def}}{=} \Pr[W_j(i,s)]$, $\underline{\gamma}_0(i) \stackrel{\text{def}}{=} \Pr[\bigcup_{k=0}^{i} \underline{Z}(R_0)(k)]$, and $\underline{\gamma}_j(i) \stackrel{\text{def}}{=} \Pr[\bigcup_{k=0}^{i} \underline{Z}(R_j)(k)]$.

In what follows, we show that $\{A(i), B_j(i,s) | 0 \leq i \leq m, 0 \leq s \leq m, 1 \leq j \leq N\}$ are uniquely determined from

$$\mathcal{C}_{R_0,m} \stackrel{\text{def}}{=} \{\underline{\gamma}_0(i), \underline{\gamma}_j(i) | 0 \leq i \leq m,\ 1 \leq j \leq N\}$$

by solving equations (3.29) and (3.30) inductively with regard to m.

Note that, by (3.15), (3.16) and (3.17),

$$1 \leq s \leq m \Rightarrow \exists m' < m \ (B_j(m,s) = B_j(m',0)) \tag{3.31}$$

$$B_j(m,m) = \Pr[W_j(m,m)] = \Pr[\bigcup_{r \in R_j} V(K_r)(0)] > 0 \tag{3.32}$$

$$B_j(m,m) \leq B_j(m,0) \tag{3.33}$$

(case: $m = 0$**)**

From condition (3.20), $A(0), B_j(0,0)$ $(1 \leq j \leq N)$ are positive, thus we have

$$\begin{aligned} 0 &< \underline{\gamma}_0(0), \underline{\gamma}_j(0) \quad 1 \leq j \leq N & (3.34)\\ \underline{\gamma}_0(0) &< \sum_{j=1}^{N} \underline{\gamma}_j(0) & (3.35) \end{aligned}$$

We rewrite equations (3.29) and (3.30): $\underline{\gamma}_0(0) = A(0)(1 - \prod_{j=1}^{N}(1 - B_j(0,0)))$ and $\underline{\gamma}_j(0) = A(0)B_j(0,0)$, hence

$$\underline{\gamma}_0(0) = A(0)(1 - \prod_{j=1}^{N}(1 - \frac{\underline{\gamma}_j(0)}{A(0)})) \tag{3.36}$$

Then we consider an equation with variable x:

$$\begin{aligned} f(x) &\stackrel{\text{def}}{=} \underline{\gamma}_0(0) - x(1 - \prod_{j=1}^{N}(1 - \frac{\underline{\gamma}_j(0)}{x})) & (3.37)\\ f(x) &= 0 & (3.38) \end{aligned}$$

From the above conditions (3.34) and (3.35) of coefficients, *Lemma 1* in [4] tells us that equation (3.38) with the range of $0 < \underline{\gamma}_j(0) \leq \underline{\gamma}_0(0) \leq x \leq 1$ has a unique solution. Therefore, let such the solution of (3.38) be x^*, we obtain $A(0) = x^*$ and $B_j(0,0) = \underline{\gamma}_j(0)/x^*$ by solving (3.38). Thus $A(0), B_j(0,0)$ are determined from $\underline{\mathcal{C}}_{R_0,0}$.

Note that $\exists j(\underline{\gamma}_j(0) = \underline{\gamma}_0(0)) \Leftrightarrow x^* = \underline{\gamma}_0(0)$, and $\underline{\gamma}_0(0) = 1 - \prod_{j=1}^{N}(1 - \underline{\gamma}_j(0)) \Leftrightarrow x^* = 1$.

(case: $m \geq 1$**)**

Suppose that $0 < A(0), B_j(0,0) \leq 1$ and $0 \leq A(i), B_j(i,0) \leq 1$ $\quad (1 \leq i \leq m-1)$ are determined from $\underline{\mathcal{C}}_{R_0,m-1}$. Then we show $A(m), B_j(m,0)$ are determined from $\underline{\mathcal{C}}_{R_0,m}$. We rewrite equations (3.29) and (3.30):

$$\begin{aligned} \underline{\gamma}_0(m) &= \sum_{s=0}^{m} A(s)(1 - \prod_{j=1}^{N}(1 - B_j(m,s))) & (3.39)\\ &= A(m)C_m + D_m + A(0)(1 - \prod_{j=1}^{N}(1 - B_j(m,0)))\\ &= A(m)C_m + D_m + A(0)(1 - \prod_{j=1}^{N}(1 - \frac{\underline{\gamma}_j(m) - E_{m,j} - A(m)B_j(m,m)}{A(0)})) \end{aligned}$$

$$\begin{aligned}\underline{\gamma}_j(m) &= \sum_{s=0}^{m} A(s)B_j(m,s) \\ &= A(m)B_j(m,m) + E_{m,j} + A(0)B_j(m,0)\end{aligned} \tag{3.40}$$

where

$$\begin{aligned}C_m &\stackrel{\text{def}}{=} 1 - \prod_{j=1}^{N}(1 - B_j(m,m)) \\ D_m &\stackrel{\text{def}}{=} \sum_{s=1}^{m-1} A(s)(1 - \prod_{j=1}^{N}(1 - B_j(m,s))) \\ E_{m,j} &\stackrel{\text{def}}{=} \sum_{s=1}^{m-1} A(s)B_j(m,s)\end{aligned}$$

Since each $B_j(m,s)$ $(1 \le s \le m)$ equals one of $B_j(i,0)$ $(0 \le i \le m-1)$ by (3.31), C_m, D_m and $E_{m,j}$ are calculated by $A(i), B_j(i,0)$ for $0 \le i \le m-1$. Then we consider an equation with variable x:

$$f(x) \stackrel{\text{def}}{=} \underline{\gamma}_0(m) - xC_m - D_m - A(0)(1 - \prod_{j=1}^{N}(1 - \frac{\underline{\gamma}_j(m) - E_{m,j} - xB_j(m,m)}{A(0)})) \tag{3.41}$$

$$f(x) = 0 \tag{3.42}$$

In what follows, we will show that equation (3.42) with the range of $0 < x$ always has two solutions, and the second largest solution equals $A(m)$. Therefore, let the second largest solution be x^*, we obtain $A(m) = x^*$ and $B_j(m,0) = (\underline{\gamma}_j(m) - E_{m,j} - x^*B_j(m,m))/A(0)$ by solving (3.42). Thus $A(m), B_j(m,0)$ are determined from $\underline{\mathcal{C}}_{R_0,m}$.

To show this, we analyze polynomial (3.41) in the same way of *Lemma 1* in [17]. Let

$$g(y) \stackrel{\text{def}}{=} f(A(m) + yA(0)) \tag{3.43}$$

We shall show that $y = 0$ is the second largest solution of the equation $g(y) = 0$. In other words, we shall show the equation $g(y) = 0$ with the range of $0 < y$ has a unique solution. Let $b_j(m) \stackrel{\text{def}}{=} B_j(m,0)$ and $b_j(0) \stackrel{\text{def}}{=} B_j(m,m) = B_j(0,0)$.

$$\begin{aligned}g(y) &= \underline{\gamma}_0(m) - A(m)C_m - yA(0)C_m - D_m - A(0)(1 - \prod_{j=1}^{N}(1 - \frac{A(0)b_j(m) - yA(0)b_j(0)}{A(0)}))\ )) \\ &= \underline{\gamma}_0(m) - A(m)C_m - yA(0)C_m - D_m - A(0)(1 - \prod_{j=1}^{N}(1 - b_j(m) + yb_j(0)))\end{aligned}$$

Let $H \stackrel{\text{def}}{=} \{0,1\}^N - \{1\}^N - \{0\}^N$ and $H(k) \stackrel{\text{def}}{=} \{\boldsymbol{h} \in H \mid \sum_{h_j \in \boldsymbol{h}} h_j = N - k\}$. It is clear that:

$$\prod_{j=1}^{N}(1 - b_j(m) + yb_j(0)) = \prod_{j=1}^{N}(1 - b_j(m)) + \sum_{\boldsymbol{h} \in H}\prod_{j=1}^{N}(1 - b_j(m))^{h_j}(yb_j(0))^{1-h_j} + \prod_{j=1}^{N} yb_j(0)$$

Thus, by (3.39), we have:

$$
\begin{aligned}
g(y) &= \underline{\gamma}_0(m) - A(m)C_m - yA(0)C_m - D_m - A(0)(1 - \prod_{j=1}^{N}(1 - b_j(m))) \\
&\quad + A(0)\sum_{\boldsymbol{h}\in H}\prod_{j=1}^{N}(1 - b_j(m))^{h_j}(yb_j(0))^{1-h_j} + A(0)\prod_{j=1}^{N} yb_j(0) \\
&= -yA(0)C_m + A(0)\sum_{\boldsymbol{h}\in H}\prod_{j=1}^{N}(1 - b_j(m))^{h_j}(yb_j(0))^{1-h_j} + A(0)\prod_{j=1}^{N} yb_j(0) \\
&= y^N G_0 A(0) + \sum_{k=2}^{N-1} y^k G_{m,k} A(0) + yF_m A(0)
\end{aligned}
$$

where

$$
\begin{aligned}
F_m &\overset{\text{def}}{=} \sum_{j=1}^{N}\Big(\prod_{i\neq j}^{1\leq i\leq N}(1 - b_i(m))\Big)b_j(0) - C_m \qquad (3.44) \\
&= \sum_{j=1}^{N}\Big(\prod_{i\neq j}^{1\leq i\leq N}(1 - b_i(m))\Big)b_j(0) + \prod_{j=1}^{N}(1 - b_j(0)) - 1 \\
G_0 &\overset{\text{def}}{=} \prod_{j=1}^{N} b_j(0) \\
G_{m,k} &\overset{\text{def}}{=} \sum_{\boldsymbol{h}\in H(k)}\prod_{j=1}^{N}(1 - b_j(m))^{h_j} b_j(0)^{1-h_j}
\end{aligned}
$$

Therefore

$$
\begin{aligned}
\frac{dg}{dy}(y) &= y^{N-1}NG_0A(0) + \sum_{k=1}^{N-2} y^k(k+1)G_{m,k+1}A(0) + F_mA(0) \\
\frac{d^2g}{dy^2}(y) &= y^{N-2}(N-1)NG_0A(0) + \sum_{k=1}^{N-3} y^k(k+1)(k+2)G_{m,k+2}A(0) + 2G_{m,2}A(0)
\end{aligned}
$$

We show $F_m < 0$. It is clear that, for any set $\{W_1, ..., W_N\}$ of events,

$$
\begin{aligned}
\Omega &= \bigcap_{j=1}^{N} W_j + \sum_{k=1}^{N-1}\sum_{\boldsymbol{h}\in H(k)}\Big(\big(\bigcap_{h_i=1}^{1\leq i\leq N} W_i\big)\cap\big(\bigcap_{h_i=0}^{1\leq i\leq N} W_i^c\big)\Big) + \bigcap_{j=1}^{N} W_j^c \\
\therefore \Omega &- \sum_{k=1}^{N-2}\sum_{\boldsymbol{h}\in H(k)}\Big(\big(\bigcap_{h_i=1}^{1\leq i\leq N} W_i\big)\cap\big(\bigcap_{h_i=0}^{1\leq i\leq N} W_i^c\big)\Big) = \bigcap_{j=1}^{N} W_j + \sum_{j=1}^{N}\Big(\big(\bigcap_{i\neq j}^{1\leq i\leq N} W_i^c\big)\cap W_j\Big) + \bigcap_{j=1}^{N} W_j^c
\end{aligned}
$$

Consider $W_j(m, m)$ as W_j. Since $W_j(m, m)$ and $W_{j'}(m, m)$ are independent, and $b_j(m) \leq b_j(0)$ by (3.33),

$$
\begin{aligned}
1 &\geq \prod_{j=1}^{N}\Pr[W_j(m,m)] + \sum_{j=1}^{N}\Big(\prod_{i\neq j}^{1\leq i\leq N}(1 - \Pr[W_i(m,m)])\Big)\Pr[W_j(m,m)] + \prod_{j=1}^{N}(1 - \Pr[W_j(m,m)]) \\
&= \prod_{j=1}^{N} b_j(0) + \sum_{j=1}^{N}\Big(\prod_{i\neq j}^{1\leq i\leq N}(1 - b_j(0))\Big)b_j(0) + \prod_{j=1}^{N}(1 - b_j(0))
\end{aligned}
$$

$$\geq \prod_{j=1}^{N} b_j(0) + \sum_{j=1}^{N}\Big(\prod_{\substack{1\leq i\leq N\\ i\neq j}} (1 - b_j(m))\Big)b_j(0) + \prod_{j=1}^{N}(1 - b_j(0)) \quad = \quad \prod_{j=1}^{N} b_j(0) + F_m + 1$$

Since $b_j(0) = B_j(m, m) > 0$ by (3.32), $F_m \leq -\prod_{j=1}^{N} b_j(0) < 0$.

Therefore we have: $g(0) = 0$, $dg/dy(0) = F_m A(0) < 0$, $d^2g/dy^2(y) > 0$ $(0 < y)$. Furthermore, since $G_0A(0)$, the coefficient of the principal term, is positive, $\lim_{y\to\infty} g(y) > 0$. Thus, the equation $g(y) = 0$ with the range of $0 < y$ has a unique solution.

[Proof of Lemma 2]

For convenience, we define $V(\emptyset)(0)$ as the whole event Ω, and, for $m \geq 1$, $V(\emptyset)(m)$ as the empty event $\emptyset$.

For each set R of paths satisfying that $\overline{L}(R) \neq \emptyset$ and $\overline{L}(R) \neq \underline{L}(R)$, there exists at least one $R_1 \in \Delta$ such that $l_{R_1} \in \underline{L}(R)$ and $l_{R_1} \notin \overline{L}(R)$. Let $R_2 \overset{\text{def}}{=} R \cap R_1$, thus we have $R_2 \neq R$ and $l_{R_1} \in \boldsymbol{L}^R_{R_2}$. Therefore $\mathcal{R}_R \overset{\text{def}}{=} \{R' \subset R | R' \neq R, \boldsymbol{L}^R_{R'} \neq \emptyset\}$ is not empty.

Then we choose R_0 as one of the smallest (in size) element in $\mathcal{R}_R$ (i.e. $R_0 \in \mathcal{R}_R$, and $|R_0| \leq |R'|$ for every $R' \in \mathcal{R}_R$).

We shall show that $A_R(m)$ is calculated from $\bigcup_{R'\subset R}^{R'\neq R} \mathcal{A}^{R_0}_{R',m}$ and $\overline{\gamma}^{R_0}_{R,m}$ where

$$\begin{aligned}
\overline{\gamma}^{R_0}_{R,m} &\overset{\text{def}}{=} \{\Pr[\overline{Z}(R - R_0)(0)]\} \cup \{\Pr[\overline{Z}(R - R_0)(0) \cap \overline{Z}(R_0)(i)] | 0 \leq i \leq m\} \\
\mathcal{A}^{R_0}_{R',m} &\overset{\text{def}}{=} \{A_{R''}(i) | 0 \leq i \leq m,\ R_0 \subset R'' \subset R'\}
\end{aligned}$$

We define $U^R_{R'}(m)$ as follows:

$$U^R_{R'}(m) \overset{\text{def}}{=} \Pr[V(\boldsymbol{L}^R_{R'})(m)]$$

Note that $U^R_R(m) = A_R(m)$. Furthermore, by using (3.2), we have:

$$U^R_{R'}(m) = \sum_{s\in S(m)} U^{R+\{r\}}_{R'+\{r\}}(s_1) U^{R+\{r\}}_{R'}(s_2) \tag{3.45}$$

(step 1):

First we show that $U^{R'}_{R''}(m)$ for each R', R'' (such that $R_0 \subset R'' \subset R' \subset R$) is calculated from $\mathcal{A}^{R_0}_{R',m}$ by induction with regard to $|R' - R''|$ and m.

(case: $|R' - R''| = 0$)

Since $R' = R''$, $U^{R'}_{R''}(m) = A_{R'}(m) \in \mathcal{A}^{R_0}_{R',m}$.

(case: $|R' - R''| > 0$**)**

We choose $a \in R' \wedge a \notin R''$. Note that $|R' - (R'' + \{a\})| = |(R' - \{a\}) - R''| < |R' - R''|$.

For $m = 0$, then, by (3.45), (3.20) and (3.11), we have:

$$U_{R''}^{R'}(0) = \frac{U_{R''}^{R'-\{a\}}(0)}{U_{R''+\{a\}}^{R'}(0)}$$

where $U_{R''}^{R'-\{a\}}(0)$ and $U_{R''+\{a\}}^{R'}(0)$ are calculated from $\mathcal{A}_{R',0}^{R_0}$ by the induction hypothesis, since $\mathcal{A}_{R'-\{a\},0}^{R_0} \subset \mathcal{A}_{R',0}^{R_0}$. Thus, so is $U_{R''}^{R'}(0)$.

For $m > 0$, then, by (3.45), (3.20), (3.13) and (3.14), we have:

$$U_{R''}^{R'}(m) = \frac{U_{R''}^{R'-\{a\}}(m) - \sum_{s \in S(m)}^{s_2 < m} U_{R''+\{a\}}^{R'}(s_1) U_{R''}^{R'}(s_2)}{U_{R''+\{a\}}^{R'}(0)}$$

where $U_{R''}^{R'-\{a\}}(m)$, $U_{R''+\{a\}}^{R'}(s_1)$ (for $0 \leq s_1 \leq m$), $U_{R''}^{R'}(s_2)$ (for $0 \leq s_2 \leq m-1$) and $U_{R''+\{a\}}^{R'}(0)$ are calculated from $\mathcal{A}_{R',m}^{R_0}$ by the induction hypothesis, thus, so is $U_{R''}^{R'}(m)$.

(step 2):

Next, we show that $U_{R'}^{R}(m)$ for each R' (such that $R_0 \subset R' \subset R$) is calculated from $\overline{\gamma}_{R,m}^{R_0}$ and $\bigcup_{R' \subset R}^{R' \neq R} \mathcal{A}_{R',m}^{R_0}$ by induction with regard to $|R' - R_0|$ and m.

(case: $|R' - R_0| = 0$**)**

Note that

$$\begin{aligned}
\overline{Z}(R - R_0)(0) \cap \overline{Z}(R_0)(m) &= V(\underline{\boldsymbol{L}}(R - R_0))(0) \cap \bigcap_{r \in R_0} V(\boldsymbol{L}_r)(m) \\
&= V(\underline{\boldsymbol{L}}(R - R_0))(0) \cap \bigcap_{r \in R_0} V(\boldsymbol{L}_r \setminus \underline{\boldsymbol{L}}(R - R_0))(m) \\
&= V(\underline{\boldsymbol{L}}(R - R_0))(0) \cap V(\overline{\boldsymbol{L}}(R_0) \setminus \underline{\boldsymbol{L}}(R - R_0))(m) \quad (3.46) \\
&= \overline{Z}(R - R_0)(0) \cap V(\boldsymbol{L}_{R_0}^{R})(m)
\end{aligned}$$

Then we have:

$$U_{R_0}^{R}(m) = \frac{\Pr[\overline{Z}(R - R_0)(0) \cap \overline{Z}(R_0)(m)]}{\Pr[\overline{Z}(R - R_0)(0)]} \tag{3.47}$$

which is calculated from $\overline{\gamma}_{R,0}^{R_0}$.

To justify (3.46), we shall show, for every $r \in R_0$,

$$\boldsymbol{L}_r \setminus \underline{\boldsymbol{L}}(R - R_0) \subset \overline{\boldsymbol{L}}(R_0) \setminus \underline{\boldsymbol{L}}(R - R_0) \tag{3.48}$$

Suppose $\exists r_0 \in R_0$ and $\exists R_1 \in \Delta$ such that $l_{R_1} \in \boldsymbol{L}_{r_0} \setminus (\underline{\boldsymbol{L}}(R - R_0) \cup \overline{\boldsymbol{L}}(R_0))$.

Then we have $R_1 \cap (R - R_0) = \emptyset$, $R_0 \setminus R_1 \neq \emptyset$, and $R_1 \cap R \neq \emptyset$. Hence, $R_1 \cap R \subset R_0$, $R_0 - R_1 \cap R \neq \emptyset$, and thus $|R_0| > |R_1 \cap R|$.

On the other hand, since $l_{R_1} \in \overline{L}(R_1 \cap R)$ and $l_{R_1} \notin \underline{L}(R - R_1 \cap R)$, we have

$$L^R_{R_1 \cap R} = \overline{L}(R_1 \cap R) \setminus \underline{L}(R - R_1 \cap R) \neq \emptyset$$

Since we choose R_0 as one of the smallest (in size) $R' \subset R$ satisfying that $L^R_{R'} \neq \emptyset$, the fact that $L^R_{R_1 \cap R} \neq \emptyset$ implies $|R_0| \leq |R_1 \cap R|$.

This contradiction negatives the hypothesis, thus, (3.48) is held.

(case: $|R' - R_0| > 0$)

We choose $b \in R' \wedge b \notin R_0$. Note that $R - \{b\} \subset R$, $|R' - R_0| > |R' - \{b\} - R_0|$.

For $m = 0$, then we have:

$$U^R_{R'}(0) = \frac{U^{R-\{b\}}_{R'-\{b\}}(0)}{U^R_{R'-\{b\}}(0)}$$

where $U^{R-\{b\}}_{R'-\{b\}}(0)$ is calculated from $\mathcal{A}^{R_0}_{R-\{b\},0}$, and $U^R_{R'-\{b\}}(0)$ is calculated from $\overline{\gamma}^{R_0}_{R,0}$ and $\bigcup^{R'\neq R}_{R'\subset R} \mathcal{A}^{R_0}_{R',0}$ by the induction hypothesis. Since $\mathcal{A}^{R_0}_{R-\{b\},0} \subset \bigcup^{R'\neq R}_{R'\subset R} \mathcal{A}^{R_0}_{R',0}$, so is $U^R_{R'}(0)$.

For $m > 1$, then we have:

$$U^R_{R'}(m) = \frac{U^{R-\{b\}}_{R'-\{b\}}(m) - \sum^{s_1<m}_{s\in S(m)} U^R_{R'}(s_1) U^R_{R'-\{b\}}(s_2)}{U^R_{R'-\{b\}}(0)}$$

where $U^{R-\{b\}}_{R'-\{b\}}(m)$ is calculated from $\mathcal{A}^{R_0}_{R-\{b\},0}$, , and $U^R_{R'}(s_1)$ (for $0 \leq s_1 \leq m-1$), $U^R_{R'-\{b\}}(s_2)$ (for $0 \leq s_2 \leq m$), and $U^R_{R'-\{b\}}(0)$ are calculated from $\overline{\gamma}^{R_0}_{R,m}$, and $\bigcup^{R'\neq R}_{R'\subset R} \mathcal{A}^{R_0}_{R',m}$ by the induction hypothesis. Thus so is $U^R_{R'}(m)$.

Note that, we can calculate $A_R(m)$ for $R \notin \Delta$ from $\{A_{R'}(i) | R \subset R', R' \in \Delta, 0 \leq i \leq m\}$, because $\alpha_R(m)$ is calculated from $\{A_{R'}(i) | R \subset R', R' \in \Delta, 0 \leq i \leq m\}$ by Lemma 3, and further, $A_R(m)$ is calculated from $\{\alpha_{R'}(i) | l_{R'} \in \overline{L}(R), 0 \leq i \leq m\}$ by its definition.

[Proof of Lemma 3]

For each $R \in \Delta$ and $m \in M$, we show that $\alpha_R(m)$ ($= \Pr[X_{l_R}(m)]$) is calculated from:

$$A_{R,m} \stackrel{\text{def}}{=} \{A_{R'}(i) | R \subset R', R' \in \Delta, 0 \leq i \leq m\}$$

by induction with regard to m and R.

(case: $|R| = \max_{R\in\Delta}|R|$)

If R is one of the largest element in Δ (i.e. $|R| = \max_{R\in\Delta}|R|$), then $\overline{L}(R) = \{l_R\}$. Because if $\overline{L}(R)$ includes another link $l_{R'}$ ($\neq l_R$), then $R \subset R'$ and $R \neq R'$, thus $|R| < |R'|$. This conflicts the hypothesis of R.

Then we have $A_R(m) = \Pr[V(l_R)(m)] = \Pr[X_{l_R}(m)]$, and thus $\alpha_R(m) = A_R(m)$.

(case: $|R| < \max_{R\in\Delta}|R|$)

If $\overline{L}(R) = \{l_R\}$, then, in the same way as in the previous case, we have $\alpha_R(m) = A_R(m)$.

Otherwise, $\overline{L}(R)$ includes a link $l_{R'}$ other than l_R.

For $m = 0$, suppose that, for each $R' \in \Delta$ such that $|R| < |R'|$, $\alpha_{R'}(0)$ is already obtained from $\boldsymbol{A}_{R',0}$.

Let $\overline{L}(R) = \{l_R, l_{R'_1}, l_{R'_2}, .., l_{R'_n}\}$.

$$A_R(0) \;=\; \Pr[V(\sum_{j=1}^{n} l_{R'_j})(0) \cap V(\{l_R\})(0)] \;=\; (\prod_{j=1}^{n} \alpha_{R'_j}(0))\alpha_R(0)$$

where, since $R \subset R'_j$ and $R \neq R'_j$, we have $|R| < |R'_j|$, and thus, $\alpha_{R'_j}(0)$ are already obtained.

Therefore $\alpha_R(0) = A_R(0)/(\prod_{j=1}^{n} \alpha_{R'_j}(0))$ and thus, $\alpha_R(0)$ is obtained from $\boldsymbol{A}_{R,0}$.

For $m \geq 1$, suppose that, for each $R' \in \Delta$ such that $|R| < |R'|$, $\alpha_{R'}(i) \quad (0 \leq i \leq m)$ are already obtained from $\boldsymbol{A}_{R',m}$. Moreover, for R itself, suppose $\alpha_R(i) \quad (0 \leq i \leq m-1)$ are already obtained from $\boldsymbol{A}_{R,m-1}$.

Let $\overline{L}(R) = \{l_R, l_{R'_1}, l_{R'_2}, .., l_{R'_n}\}$. $S_n(m)$ is defined by (3.7).

$$\begin{aligned} A_R(m) &= \Pr[\sum_{\boldsymbol{s}\in S(m)} V(\sum_{j=1}^{n} l_{R'_j})(s_1) \cap V(\{l_R\})(s_2)] \\ &= \sum_{\boldsymbol{s}\in S(m)} \Pr[V(\sum_{j=1}^{n} l_{R'_j})(s_1)]\alpha_R(s_2) \;=\; \alpha_R(m)C_m + D_m \end{aligned}$$

where

$$C_m \stackrel{\text{def}}{=} \sum_{(s_1,m)\in S(m)} \Big(\sum_{\boldsymbol{s}'\in S_n(s_1)} \prod_{j=1}^{n} \alpha_{R'_j}(s'_j)\Big) \;=\; \prod_{j=1}^{n} \alpha_{R'_j}(0)$$

$$D_m \stackrel{\text{def}}{=} \sum_{(s_1,s_2)\in S(m)}^{s_2<m} \Big(\sum_{\boldsymbol{s}'\in S_n(s_1)} \prod_{j=1}^{n} \alpha_{R'_j}(s'_j)\Big)\alpha_R(s_2)$$

Since $R \subset R'_j$ and $R \neq R'_j$, we have $|R| < |R'_j|$, and thus $\alpha_{R'_j}(i)$ are already obtained. Thus C_m, D_m are calculated by $\alpha_{R'_j}(i) \quad (0 \leq i \leq m, 1 \leq j \leq n)$ and $\alpha_R(i) \quad (0 \leq i \leq m-1)$. Since $C_m > 0$,

$$\alpha_R(m) = (A_R(m) - D_m)/C_m$$

and thus, $\alpha_R(m)$ is obtained from $\boldsymbol{A}_{R,m}$.

[Proof of Theorem 1]

First we number all (logical) links. Let $K = |\Delta^*|$, $\Delta^* = \{l^{(k)} | 1 \leq k \leq K\}$. And, for each $l^{(k)}$, let $R^{(k)}$ be the influence-path set for $l^{(k)}$ (i.e. $l^{(k)} = l_{R^{(k)}}$). We can take the numbering of $l^{(k)}$ satisfying that $R^{(k)} \subset R^{(k')} \Rightarrow k < k'$.

We also number all sets of paths related to our calculation. Let $\Psi \stackrel{\text{def}}{=} \{R \subset \boldsymbol{Path} | \overline{L}(R) \neq \emptyset\}$, $J \stackrel{\text{def}}{=} |\Psi|$, $\Psi_0 \stackrel{\text{def}}{=} \{R \in \Psi |\; |R| = 1\}$, and $J_0 \stackrel{\text{def}}{=} |\Psi_0|$. We can take the numbering of R_j satisfying that $\Psi_0 = \{R_j | 1 \le j \le J_0\}$, $\Psi - \Psi_0 = \{R_j | J_0 < j \le J\}$ and $R_j \subset R_{j'} \Rightarrow j < j'$. Moreover, for each R_j $(J_0 < j \le J)$, we choose $R_{z_j} \subset R_j$ as R_0 in Lemma 2.

In what follows, we fix some $m \in M$. For $0 \le i \le m, 1 \le k \le K, 1 \le j \le J$, we define:

$$\begin{aligned}
\alpha_{k,i} &\stackrel{\text{def}}{=} \Pr[X_{l(k)}(i)] \\
A_{k,i} &\stackrel{\text{def}}{=} \Pr[V(\overline{L}(R^{(k)}))(i)] \\
\gamma_{j,i} &\stackrel{\text{def}}{=} \Pr[\bigcup_{h=0}^{i} \underline{Z}(R_j)(h)] \\
\overline{\gamma}_{j,i} &\stackrel{\text{def}}{=} \Pr[\overline{Z}(R_j - R_{z_j})(0) \cap \overline{Z}(R_{z_j})(i)] \quad \text{if} \quad J_0 < j \\
\boldsymbol{\alpha} &= \{\alpha_{k,i} | 0 \le i \le m, 1 \le k \le K\} \\
\boldsymbol{A} &= \{A_{k,i} | 0 \le i \le m, 1 \le k \le K\} \\
\boldsymbol{\gamma} &= \{\gamma_{j,i}, \overline{\gamma}_{j',i} | 0 \le i \le m, 1 \le j \le J, J_0 < j' \le J\}
\end{aligned}$$

Then we consider a map:

$$\Gamma : \mathcal{D} \to \mathcal{E} \quad \text{such that } \boldsymbol{\gamma}(m) = \Gamma(\boldsymbol{\alpha})$$

$$\begin{aligned}
\mathcal{D} &\stackrel{\text{def}}{=} \{(x_{k,i})_{1\le k\le K, 0\le i\le m} \in [0,1]^{(m+1)K} \,|\, 0 < x_{k,0},\ 0 \le x_{k,i},\ \sum_{i=0}^{m} x_{k,i} \le 1\} \\
\mathcal{E} &\stackrel{\text{def}}{=} \Gamma(\mathcal{D}) \subset [0,1]^{(m+1)(2J-J_0)}
\end{aligned}$$

According to the proofs of Lemma 1, 2 and 3, we have $\Gamma = \chi \circ \psi$ where $\psi(\boldsymbol{\alpha}) = \boldsymbol{A}$ and $\chi(\boldsymbol{A}) = \boldsymbol{\gamma}$. Thus $\Gamma^{-1} \stackrel{\text{def}}{=} \psi^{-1} \circ \chi^{-1}$.

Since ψ has an explicit expression, it follows directly that ψ is injection and ψ^{-1} is continuously differentiable. Therefore we shall show so is χ^{-1}.

We have the implicit function for χ:

$$\boldsymbol{F}(\boldsymbol{x}, \boldsymbol{y}) = (F_{k,i}(\boldsymbol{x}, \boldsymbol{y}))_{1\le k\le K, 0\le i\le m} \text{ such that } \boldsymbol{F}(\boldsymbol{A}, \boldsymbol{\gamma}) = 0$$

According to Lemma 1 and 2, each $A_{k,i}$ can be determined by one or two types of calculation, if it is applicable. Therefore $F_{k,i}$ can be constructed as follows.

(case: 0)

For k such that $A_{k,i}$ is determined by 1. of Lemma 1, there exists $R_j = \{r\}$ (such that $R_j \subset R^{(k)}$, $\overline{L}(R^{(k)}) = \overline{L}(R_j)$ and the MIP number of $R_j = 0$) where $\gamma_{j,0} = A_{k,0}$ and $\gamma_{j,i} - \gamma_{j,i-1} = A_{k,i}$ for $1 \le i \le m$. Therefore $F_{k,0}(\boldsymbol{x}, \boldsymbol{y}) = y_{j,0} - x_{k,0}$ and $F_{k,i}(\boldsymbol{x}, \boldsymbol{y}) = y_{j,i} - y_{j,i-1} - x_{k,i}$.

(case: 1)

For k such that $A_{k,i}$ is determined by Lemma 2, there exists R_j (such that $R_j \subset R^{(k)}$, $\overline{L}(R^{(k)}) = \overline{L}(R_j)$ and the MIP number of $R_j \geq 1$) and $R_{j'} = R_j - R_{z_j}$ where $A_{k,i}$ is calculated by $\overline{\gamma}_{j,i'}$, $\overline{\gamma}_{j',0}$ or $\gamma_{j',0}$, and $A_{k',i'}$ for $0 \leq i' \leq i$ and k' satisfying that $R_{z_j} \subset R^{(k')} \subset R^{(k)}$.

Thus we can construct an explicit function $f_{k,i}$ satisfying that $A_{k,i} = f_{k,i}(\boldsymbol{A} \setminus \{A_{k,i}\}, \boldsymbol{\gamma})$. Therefore $F_{k,i}(\boldsymbol{x}, \boldsymbol{y}) = f_{k,i}(\boldsymbol{x} \setminus \{x_{k,i}\}, \boldsymbol{y}) - x_{k,i}$ and is continuously differentiable in a neighborhood of the point $(\boldsymbol{A}, \boldsymbol{\gamma})$.

(case: 2)

For k such that $A_{k,i}$ is determined by 2. of Lemma 1, there exists R_{j_0} (such that $R_{j_0} \subset R^{(k)}$, $\overline{L}(R^{(k)}) = \overline{L}(R_{j_0})$ and the MIP number of $R_{j_0} \geq 2$) where the MIP number of R_{j_0} is N_k, and its division is $\{R_{j_1}, ..., R_{j_{N_k}}\}$ satisfying that

$$\gamma_{j_0,0} - A_{k,0}\Big(1 - \prod_{h=1}^{N_k}\Big(1 - \frac{\gamma_{j_h,0}}{A_{k,0}}\Big)\Big) = 0 \qquad (3.49)$$

$$\gamma_{j_0,i'} - A_{k,i'}C_{k,i'} - D_{k,i'} - A_{k,0}\Big(1 - \prod_{h=1}^{N_k}\Big(1 - \frac{\gamma_{j_h,i'} - E_{j_h,i'} - A_{k,i'}B_{j_h,i'}}{A_{k,0}}\Big)\Big) = 0 \qquad (3.50)$$

$$1 \leq i' \leq i$$

Combining (3.49) and (3.50), we have:

$$F_{k,i}(\boldsymbol{x}, \boldsymbol{y}) = f_{k,i}(x_{k,0}, x_{k,1}, .., x_{k,i}, y_{j_0^{(k)},0}, y_{j_0^{(k)},1}, .., y_{j_0^{(k)},i}, y_{j_1^{(k)},0}, ..., y_{j_{N_k}^{(k)},i})$$

where $F_{k,i}(\boldsymbol{A}, \boldsymbol{\gamma}) = 0$ for $1 \leq k \leq K, 0 \leq i \leq m$.

By the Implicit Function Theorem, continuously differentiability of χ^{-1} (where $\boldsymbol{A} = \chi^{-1}(\boldsymbol{\gamma})$) follows by:

1. $F_{k,i}(\boldsymbol{x}, \boldsymbol{y})$ is continuously differentiable in a neighborhood of the point $(\boldsymbol{A}, \boldsymbol{\gamma})$. This follows from $A_{k,0} > 0$ in (3.49) and (3.50).
2. The linear map (i.e. a matrix)

$$\frac{\partial \boldsymbol{F}}{\partial \boldsymbol{x}}(\boldsymbol{A}, \boldsymbol{\gamma}) = \Big(\frac{\partial F_{k,i}}{\partial x_{k',i'}}(\boldsymbol{A}, \boldsymbol{\gamma})\Big)_{1 \leq k,k' \leq K, 0 \leq i,i' \leq m}$$

has the inverse map.

We shall show 2. It is clear if $k < k'$ or $i < i'$, then $\frac{\partial F_{k,i}}{\partial x_{k',i'}}(\boldsymbol{x}, \boldsymbol{y}) = 0$.

We consider the diagonal elements (i.e. $k = k'$ and $i = i'$). For k in the (case: 0) and (case: 1), $A_{k,i}$ has an explicit expression so that $F_{k,i}(\boldsymbol{x}, \boldsymbol{y}) = f_{k,i}(\boldsymbol{x} \setminus \{x_{k,i}\}, \boldsymbol{y}) - x_{k,i}$ as mentioned above, thus $\frac{\partial F_{k,i}}{\partial x_{k,i}}(\boldsymbol{x}, \boldsymbol{y}) = -1$.

Otherwise, if $i = 0$, then,

$$\frac{\partial F_k}{\partial x_k}(\boldsymbol{x}, \boldsymbol{y}) = -x_k(-\sum_{j=1}^{N_k} \frac{y_j}{x_k^2} \prod_{h \neq j}^{1 \leq h \leq N_k} (1 - \frac{y_h}{x_k})) - (1 - \prod_{j=1}^{N_k}(1 - \frac{y_j}{x_k}))$$

Thus we have:

$$\frac{\partial F_k}{\partial x_k}(\boldsymbol{A}, \gamma) = \sum_{j=1}^{N_k} b_j(0) \prod_{h \neq j}^{1 \leq h \leq N_k} (1 - b_h(0)) + \prod_{j=1}^{N_k}(1 - b_j(0)) - 1 = F_0 < 0$$

where F_0 is defined in (3.44).

Finally if $i > 0$, then, by using the function g in (3.43), $F_{k,i}|_{x_{k,i}=A_{k,i}+zA_{k,0}} = g(z)$, and thus,

$$\begin{aligned} \frac{dg}{dz}(0) &= A_{k,0}\frac{\partial F_{k,i}}{\partial x_{k,i}}(\boldsymbol{A}, \gamma) \\ \frac{\partial F_{k,i}}{\partial x_{k,i}}(\boldsymbol{A}, \gamma) &= \frac{dg}{dz}(0)/A_{k,0} \leq -\frac{\prod_{h=1,N_k} b_{k,j_h}(0)}{A_{k,0}} < 0 \end{aligned}$$

Therefore $\frac{\partial \boldsymbol{F}}{\partial \boldsymbol{y}}(\boldsymbol{A}, \gamma)$ has the inverse.

Chapter 4

Inferring link loss rates from unicast-based end-to-end measurement

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 [7], 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 observable paths).

In this work, based on our general framework, we present a feasible method of inferring packet loss rates on 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, as mentioned later (Section 4), for most of path-topologies in actual networks, loss rates on 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 the MINC (Multicast-based Inference of Network-internal Characteristics) project. In [4], 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, which is still major in the current Internet. 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 correlation in concurrent events on paths. As mentioned later (Section 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 work 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 [5], [25], [24]. In particular, the method in [25] 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 [30]. On the other hand, our method infers more quantitative characteristics, i.e., packet loss rates on individual links.

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

2 Tree path-topology

2.1 Basic model and problem

Let us consider a single-level binary tree ((I) of Fig. 4.1). 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 on l_R. Our goal is to determine x_R from end-to-end

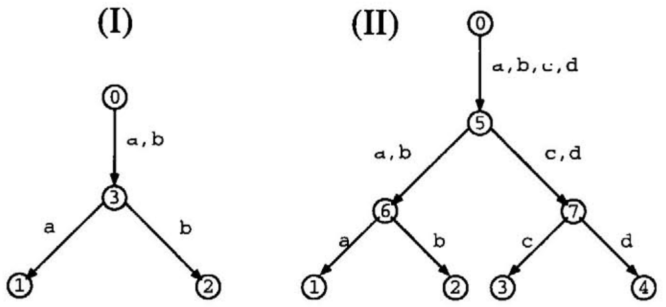

Figure 4.1: Tree path-topologies

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:

$$y_a = x_{ab}x_a, \; y_b = x_{ab}x_b, \; y_{ab} = x_{ab}x_a x_b \tag{4.1}$$

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

$$x_a = y_{ab}/y_b, \; x_b = y_{ab}/y_a, \; x_{ab} = y_a y_b / y_{ab} \tag{4.2}$$

However, the following problems arise in this naive inference:

($*$) Concurrent events on paths — In (4.1), 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 (4.1), 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 [4] and [16]. 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., [4]). 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 $M_{a.b}$ 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 $M_{a.b}$ 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 $M_{a.b}$ and M_b, let $X^{(D)}$ denote the event in measurement D ($D \in \{a.b, b\}$). 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_i denote the event that the i-th probe entering the shared link l_{ab} in a trial passes through the link ($i \in \{1, 2\}$).

For each trial in $M_{a.b}$, 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) $X_{ab,a} \cap X^c_{ab,b} = V_1 \cap (V_2)^c$, (e3) $X^c_{ab,a} \cap X_{ab,b} = (V_1)^c \cap V_2$, (e4) $X^c_{ab,a} \cap X^c_{ab,b} = (V_1)^c \cap (V_2)^c$, 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) is negligible, and is essential to our inference method. On the other hand, the occurrence of (e2) is not harmful to the inference.

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

$$
\begin{aligned}
x^*_{ab} &\stackrel{\text{def}}{=} \Pr[V_1^{(a.b)} \cap V_2^{(a.b)}], \quad \varepsilon \stackrel{\text{def}}{=} \Pr[(V_1^{(a.b)})^c | V_2^{(a.b)}] \\
x_{ab} &\stackrel{\text{def}}{=} \Pr[V_1^{(b)}] = \Pr[V_1^{(a.b)}] \\
x_a &\stackrel{\text{def}}{=} \Pr[X^{(a.b)}_{a,a} | V_1^{(a.b)}], \quad x_b \stackrel{\text{def}}{=} \Pr[X^{(b)}_{b,b} | V_1^{(b)}] \\
x'_b &\stackrel{\text{def}}{=} \Pr[X^{(a.b)}_{b,b} | V_1^{(a.b)} \cap V_2^{(a.b)}] = \Pr[X^{(a.b)}_{b,b} | V_2^{(a.b)}]
\end{aligned}
$$

where we introduce an approximation $\Pr[V_1^{(b)}] = \Pr[V_1^{(a.b)}]$, 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 preceded probes, where we introduce an approximation $\Pr[X_{b,b}^{(a.b)}|V_1^{(a.b)} \cap V_2^{(a.b)}] = \Pr[X_{b,b}^{(a.b)}|V_2^{(a.b)}]$.

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

$$\begin{aligned}
y_b^{(b)} &\stackrel{\text{def}}{=} \Pr[Y_b^{(b)}] = \Pr[X_{b,b}^{(b)} \cap V_1^{(b)}] \\
y_a &\stackrel{\text{def}}{=} \Pr[Y_a^{(a.b)}] = \Pr[X_{a,a}^{(a.b)} \cap V_1^{(a.b)}] \\
y_b &\stackrel{\text{def}}{=} \Pr[Y_b^{(a.b)}] = \Pr[X_{b,b}^{(a.b)} \cap V_2^{(a.b)}] \\
y_{ab} &\stackrel{\text{def}}{=} \Pr[Y_a^{(a.b)} \cap Y_b^{(a.b)}] \\
&= \Pr[X_{a,a}^{(a.b)} \cap X_{b,b}^{(a.b)} \cap V_1^{(a.b)} \cap V_2^{(a.b)}]
\end{aligned}$$

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

$$\Pr[X_{a,a}^{(a.b)} \cap X_{b,b}^{(a.b)}|V_1^{(a.b)} \cap V_2^{(a.b)}] = x_a x'_b(1+\delta)$$

Then we have the relation between $\boldsymbol{x}$ and $\boldsymbol{y}$:

$$\begin{aligned}
y_b^{(b)} &= x_{ab}x_b, \quad y_a = x_{ab}x_a, \quad y_b = x^*_{ab}x'_b/(1-\varepsilon), \\
y_{ab} &= x^*_{ab}x_a x'_b(1+\delta)
\end{aligned} \tag{4.3}$$

with six unknown variable x_a, x_b, x_{ab}, x'_b, x^*_{ab} and ε. Note that, if we regard ε 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.

$$\begin{aligned}
x_a &= \frac{y_{ab}}{y_b(1-\varepsilon)(1+\delta)}, \quad x_b = \frac{y_b^{(b)}y_{ab}}{y_a y_b(1-\varepsilon)(1+\delta)}, \\
x_{ab} &= \frac{y_a y_b(1-\varepsilon)(1+\delta)}{y_{ab}}
\end{aligned}$$

Since we assume both ε and δ 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.

$$\hat{x}_a \stackrel{\text{def}}{=} \frac{y_{ab}}{y_b}, \quad \hat{x}_b \stackrel{\text{def}}{=} \frac{y_b^{(b)}y_{ab}}{y_a y_b}, \quad \hat{x}_{ab} \stackrel{\text{def}}{=} \frac{y_a y_b}{y_{ab}} \tag{4.4}$$

We estimate errors between x in (4.3) and $\hat{x}$ in (4.4) as follows ($r \in \{a, b\}$):

$$\frac{\hat{x}_{ab}}{x_{ab}} = \frac{1}{(1-\varepsilon)(1+\delta)}, \quad \frac{\hat{x}_r}{x_r} = (1-\varepsilon)(1+\delta)$$

where ε and δ 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. 4.1). 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, $M_{a.b}$, $M_{b.c}$, $M_{c.d}$ and M_d, where $M_{r.r'}$ ($(r, r') \in \{(a, b), (b, c), (c, d)\}$) 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 $y_a^{(a.b)}$, $y_{ab}^{(a.b)}$, $y_b^{(b.c)}$, $y_{bc}^{(b.c)}$, $y_c^{(c.d)}$, $y_{cd}^{(c.d)}$, and $y_d^{(d)}$.

According to the above method for single-level binary trees, we can obtain $x_{abcd}x_{ab}$ and x_a from $M_{a.b}$, x_{abcd} and $x_{ab}x_b$ from $M_{b.c}$, $x_{abcd}x_{cd}$ and x_c from $M_{c.d}$, and $x_{abcd}x_{cd}x_d$ from M_d, respectively. Combining them, we can infer x_a, x_b, x_c, x_d, x_{ab}, x_{cd}, 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:

$$\begin{aligned}
\hat{x}_{ab} &\stackrel{\text{def}}{=} (y_a^{(a.b)} y_b^{(a.b)} y_{bc}^{(b.c)}) / (y_{ab}^{(a.b)} y_b^{(b.c)} y_c^{(b.c)}) \\
\frac{\hat{x}_{ab}}{x_{ab}} &= \left(\frac{1-\varepsilon_{abcd}}{1-\varepsilon_{ab}} + \frac{\varepsilon_{abcd}}{\rho}\right) \frac{(1+\delta_{ab,cd})(1+\delta_{b,c})}{1+\delta_{a,b}}
\end{aligned}$$

where $\delta_{R,R'}$ is a degree of correlation between links l_R and $l_{R'}$, ε_R is $\Pr[(V_1)^c | V_2]$ on a shared link l_R, and ρ is $\Pr[V_2 | V_1]$ on l_{ab} ($= x_{ab}^*/x_{ab}$). Note that $\rho \leq 1$ and ρ 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 a same link along paths), then we have:

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

Unlike δ, the effect of ε_R 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. 4.2). Unicast probe packets can be sent from nodes 1 and 2 to 0 along paths a and b, respectively.

Let M_{ab} be a series of trials where each trial sends an unordered pair (P_a, P_b) of probes. In M_{ab}, 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_{ab}, M_a

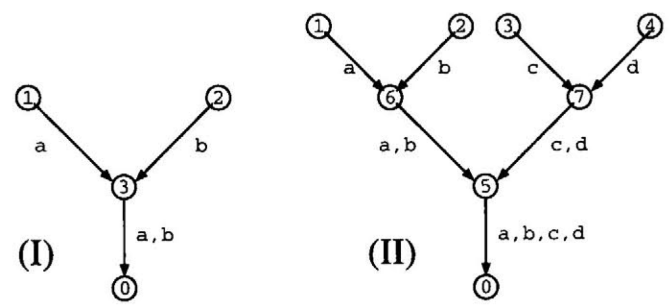

Figure 4.2: Inverse tree path-topologies

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_{ab}. Let S_{ab} denote the event that probes P_a and P_b enter l_{ab} so closely in a trial in M_{ab} 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 $W_{ab} \stackrel{\text{def}}{=} X_{a,a}^{(ab)} \cap X_{b,b}^{(ab)}$.

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

$$
\begin{aligned}
h_a &\stackrel{\text{def}}{=} \Pr[H_a \cap S_{ab} | W_{ab}] \\
h_b &\stackrel{\text{def}}{=} \Pr[H_b \cap S_{ab} | W_{ab}] \\
x^*_{ab,a} &\stackrel{\text{def}}{=} \Pr[V_1^{(ab)} \cap V_2^{(ab)} | H_a \cap S_{ab} \cap W_{ab}] \\
\varepsilon_b &\stackrel{\text{def}}{=} \Pr[(V_1^{(ab)})^c | V_2^{(ab)} \cap H_b \cap S_{ab} \cap W_{ab}] \\
x^*_{ab,b} &\stackrel{\text{def}}{=} \Pr[V_1^{(ab)} \cap V_2^{(ab)} | H_b \cap S_{ab} \cap W_{ab}] \\
\varepsilon_a &\stackrel{\text{def}}{=} \Pr[(V_1^{(ab)})^c | V_2^{(ab)} \cap H_a \cap S_{ab} \cap W_{ab}] \\
x_{ab} &\stackrel{\text{def}}{=} \Pr[V_1^{(ab)} | W_{ab}] = \Pr[V_1^{(ab)} | X_{a,a}^{(ab)}] = \Pr[V_1^{(ab)} | X_{b,b}^{(ab)}] \\
&= \Pr[V_1^{(a)} | X_{a,a}^{(a)}] = \Pr[V_1^{(b)} | X_{b,b}^{(b)}] \\
x_a &\stackrel{\text{def}}{=} \Pr[X_{a,a}^{(ab)}] = \Pr[X_{a,a}^{(a)}] \\
x_b &\stackrel{\text{def}}{=} \Pr[X_{b,b}^{(ab)}] = \Pr[X_{b,b}^{(b)}]
\end{aligned}
$$

where we introduce approximations $\Pr[V_1^{(ab)} | W_{ab}] = \Pr[V_1^{(ab)} | X_{a,a}^{(ab)}] = \Pr[V_1^{(ab)} | X_{b,b}^{(ab)}] = \Pr[V_1^{(a)} | X_{a,a}^{(a)}] = \Pr[V_1^{(b)} | X_{b,b}^{(b)}]$, $\Pr[X_{a,a}^{(ab)}] = \Pr[X_{a,a}^{(a)}]$ and $\Pr[X_{b,b}^{(ab)}] = \Pr[X_{b,b}^{(b)}]$. These approximations allow us to regard x_{ab} as the "general" no loss rate of probes entering link l_{ab} without interference from other probes given that the 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} (resp. y'_{ab}) 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".)

$$\begin{aligned}
y^{(a)}_a &\stackrel{\text{def}}{=} \Pr[Y^{(a)}_a] = \Pr[V^{(a)}_1 \cap X^{(a)}_{a,a}] \\
y^{(b)}_b &\stackrel{\text{def}}{=} \Pr[Y^{(b)}_b] = \Pr[V^{(b)}_1 \cap X^{(b)}_{b,b}] \\
y_a &\stackrel{\text{def}}{=} \Pr[Y^{(ab)}_a] = \Pr[X^{(ab)}_{ab,a} \cap X^{(ab)}_{a,a}] \\
y_b &\stackrel{\text{def}}{=} \Pr[Y^{(ab)}_b] = \Pr[X^{(ab)}_{ab,b} \cap X^{(ab)}_{b,b}] \\
y^*_{ab} &\stackrel{\text{def}}{=} \Pr[V^{(ab)}_1 \cap V^{(ab)}_2 \cap S_{ab} \cap W_{ab}] \\
y'_{ab} &\stackrel{\text{def}}{=} \Pr[V^{(ab)}_1 \cap V^{(ab)}_2 \cap (S_{ab})^c \cap W_{ab}]
\end{aligned}$$

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 δ 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 $\boldsymbol{x}$ and $\boldsymbol{y}$:

$$\begin{aligned}
y^{(a)}_a &= x_a x_{ab}, \quad y^{(b)}_b = x_b x_{ab} \\
y_a &= x_a(1 - x_b(1+\delta))x_{ab} + x_a x_b (1+\delta)(h_a x_{ab} + h_b x^*_{ab,b}/(1-\varepsilon_b)) \\
&\quad + x_a x_b (1+\delta)(1 - h_a - h_b) x_{ab} \\
&= x_a x_{ab} - x_a x_b (1+\delta) x_{ab} h_b + x_a x_b (1+\delta) x^*_{ab,b} h_b/(1-\varepsilon_b) \\
y_b &= x_b x_{ab} - x_a x_b (1+\delta) x_{ab} h_a + x_a x_b (1+\delta) x^*_{ab,a} h_a/(1-\varepsilon_a) \\
y^*_{ab} &= x_a x_b (1+\delta)(h_a x^*_{ab,a} + h_b x^*_{ab,b}) \\
y'_{ab} &= x_a x_b (1+\delta)(1 - h_a - h_b) x^2_{ab}
\end{aligned}$$

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

$$\begin{aligned}
h &\stackrel{\text{def}}{=} h_a + h_b, \quad \rho_r \stackrel{\text{def}}{=} x^*_{ab,r}/x_{ab} \quad (r \in \{a, b\}), \\
\varepsilon &\stackrel{\text{def}}{=} \frac{1}{h}\left(\frac{\rho_a \varepsilon_a h_a}{1-\varepsilon_a} + \frac{\rho_b \varepsilon_b h_b}{1-\varepsilon_b}\right), \\
z_1 &\stackrel{\text{def}}{=} (y^{(a)}_a - y_a) + (y^{(b)}_b - y_b) + y^*_{ab}, \\
z_2 &\stackrel{\text{def}}{=} y^{(a)}_a y^{(b)}_b - y'_{ab}
\end{aligned}$$

Consequently we have the following equations.

$$\begin{aligned}
y_a^{(a)} &= x_a x_{ab}, \quad y_b^{(b)} = x_b x_{ab}, \\
z_1 &= x_a x_b (1+\delta) x_{ab} (1-\varepsilon) h, \\
z_2 &= x_a x_b (1+\delta) x_{ab}^2 h - x_a x_b \delta x_{ab}^2
\end{aligned} \tag{4.5}$$

Since $\rho_a, \rho_b \le 1$, ε can be estimated as $\varepsilon \le \max(\varepsilon_a/(1-\varepsilon_a), \varepsilon_b/(1-\varepsilon_b))$. Similarly to the previous case of tree path-topologies, if we regard ε as a parameter, then x_a, x_b and x_{ab} are uniquely solved: $x_{ab} = (z_2 + \delta y_a^{(a)} y_b^{(b)})(1-\varepsilon)/z_1$, $x_a = y_a^{(a)} z_1/((z_2 + \delta y_a^{(a)} y_b^{(b)})(1-\varepsilon))$, and $x_b = y_b^{(b)} z_1/((z_2 + \delta y_a^{(a)} y_b^{(b)})(1-\varepsilon))$. Since we assume ε_a, ε_b and δ 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.

$$\begin{aligned}
\hat{x}_a &\stackrel{\text{def}}{=} \frac{y_a^{(a)} z_1}{z_2} = \frac{(y_a^{(a)} - y_a) + (y_b^{(b)} - y_b) + y_{ab}^*}{y_b^{(b)} - y'_{ab}/y_a^{(a)}}, \\
\hat{x}_b &\stackrel{\text{def}}{=} \frac{y_b^{(b)} z_1}{z_2} = \frac{(y_a^{(a)} - y_a) + (y_b^{(b)} - y_b) + y_{ab}^*}{y_a^{(a)} - y'_{ab}/y_b^{(b)}}, \\
\hat{x}_{ab} &\stackrel{\text{def}}{=} \frac{z_2}{z_1} = \frac{y_a^{(a)} y_b^{(b)} - y'_{ab}}{(y_a^{(a)} - y_a) + (y_b^{(b)} - y_b) + y_{ab}^*}
\end{aligned} \tag{4.6}$$

We estimate errors between $\boldsymbol{x}$ in (4.5) and $\hat{\boldsymbol{x}}$ in (4.6) from $z_2 \le y_a^{(a)} y_b^{(b)} \le 1$:

$$\begin{aligned}
\frac{1}{(1+\delta/z_2)(1-\varepsilon)} &\le \frac{\hat{x}_{ab}}{x_{ab}} \le \frac{1}{(1+\delta)(1-\varepsilon)} \\
(1+\delta)(1-\varepsilon) &\le \frac{\hat{x}_r}{x_r} \le (1+\delta/z_2)(1-\varepsilon)
\end{aligned}$$

which indicate that errors coming from δ may increase in inverse proportion to z_2 (i.e., $y_a^{(a)} y_b^{(b)} - y'_{ab}$). However, if we can assume that losses are moderated (i.e., $y_a^{(a)} y_b^{(b)} \gg 0$) and S_{ab} is likely to occur (i.e., $y'_{ab} \ll y_a^{(a)} y_b^{(b)}$), then the 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. 4.2). Unicast probe packets can be sent from node 1, 2, 3 and 4 to 0 along paths a, b, c and d, respectively. We consider seven independent measurements, M_a, M_{ab}, M_b, M_{bc}, M_c, M_{cd} and M_d.

According to the above method for single-level inverse trees, we can obtain $x_{abcd} x_{ab}$, x_a and x_b from triple (M_a, M_{ab}, M_b) of measurements, x_{abcd}, $x_{ab} x_b$ and $x_{cd} x_c$ from (M_b, M_{bc}, M_c), and $x_{abcd} x_{cd}$, x_c and x_d from (M_c, M_{cd}, M_d), respectively. Combining them, we can infer x_a, x_b, x_c, x_d, x_{ab}, x_{cd}, and x_{abcd}. For example, we show $\hat{x}_{ab}$ (an inferred value of x_{ab}) and its error estimation:

$$\begin{aligned}
\hat{x}_{ab} &\stackrel{\text{def}}{=} (z_2^{(ab)} z_1^{(bc)})/(z_1^{(ab)} z_2^{(bc)}) \\
\frac{\hat{x}_{ab}}{x_{ab}} &= \frac{z_2^{(ab)} (z_2^{(bc)} + \delta^{(bc)} y_b^{(b)} y_c^{(c)})(1-\varepsilon^{(bc)})}{(z_2^{(ab)} + \delta_{a,b}^{(ab)} y_a^{(a)} y_b^{(b)})(1-\varepsilon^{(ab)}) z_2^{(bc)}}
\end{aligned}$$

where

$$
\begin{aligned}
z_1^{(ab)} &\stackrel{\text{def}}{=} (y_a^{(a)} - y_a^{(ab)}) + (y_b^{(b)} - y_b^{(ab)}) + y_{ab}^{*(ab)} \\
&= x_a x_b (1 + \delta_{a,b}^{(ab)}) x_{ab} x_{abcd} h^{(ab)} (1 - \varepsilon^{(ab)}) \\
z_2^{(ab)} &\stackrel{\text{def}}{=} y_a^{(a)} y_b^{(b)} - y_{ab}'^{(ab)} \\
&= x_a x_b (1 + \delta_{a,b}^{(ab)}) x_{ab}^2 x_{abcd}^2 h^{(ab)} - x_a x_b \delta_{a,b}^{(ab)} x_{ab}^2 x_{abcd}^2 \\
z_1^{(bc)} &\stackrel{\text{def}}{=} (y_b^{(b)} - y_b^{(bc)}) + (y_c^{(c)} - y_c^{(bc)}) + y_{bc}^{*(bc)} \\
&= x_b x_{ab} x_c x_{cd} (1 + \delta^{(bc)}) x_{abcd} h^{(bc)} (1 - \varepsilon^{(bc)}) \\
z_2^{(bc)} &\stackrel{\text{def}}{=} y_b^{(b)} y_c^{(c)} - y_{bc}'^{(bc)} \\
&= x_b x_{ab} x_c x_{cd} (1 + \delta^{(bc)}) x_{abcd}^2 h^{(bc)} - x_b x_{ab} x_c x_{cd} \delta^{(bc)} x_{abcd}^2 \\
\varepsilon^{(ab)} &\stackrel{\text{def}}{=} \frac{1}{h^{(ab)}} \Big(h_a^{(ab)} \big(\frac{\rho_a \rho_a' \varepsilon_{abcd,a}^{(ab)}}{1 - \varepsilon_{abcd,a}^{(ab)}} + \frac{\rho_a' \varepsilon_{ab,a}^{(ab)}}{1 - \varepsilon_{ab,a}^{(ab)}} \big) + h_b^{(ab)} \big(\frac{\rho_b \rho_b' \varepsilon_{abcd,b}^{(ab)}}{1 - \varepsilon_{abcd,b}^{(ab)}} + \frac{\rho_b' \varepsilon_{ab,b}^{(ab)}}{1 - \varepsilon_{ab,b}^{(ab)}} \big) \Big) \\
\varepsilon^{(bc)} &\stackrel{\text{def}}{=} \frac{1}{h^{(bc)}} \Big(\frac{h_b^{(bc)} \rho_b \varepsilon_{abcd,b}^{(bc)}}{1 - \varepsilon_{abcd,b}^{(bc)}} + \frac{h_c^{(bc)} \rho_c \varepsilon_{abcd,c}^{(bc)}}{1 - \varepsilon_{abcd,c}^{(bc)}} \Big)
\end{aligned}
$$

and, $h^{(ab)} \stackrel{\text{def}}{=} h_a^{(ab)} + h_b^{(ab)}$, $\rho_r \stackrel{\text{def}}{=} x_{abcd,r}^* / x_{abcd}$, $\rho_r' \stackrel{\text{def}}{=} x_{ab,r}^* / x_{ab}$ in M_{ab} ($r \in \{a, b\}$); $h^{(bc)} \stackrel{\text{def}}{=} h_b^{(bc)} + h_c^{(bc)}$, $\delta^{(bc)} \stackrel{\text{def}}{=} (1 + \delta_{b,c}^{(bc)})(1 + \delta_{ab,cd}^{(bc)}) - 1$, $\rho_r^{(bc)} \stackrel{\text{def}}{=} x_{abcd,r}^* / x_{abcd}$ in M_{bc} ($r \in \{b, c\}$).

If we assume that $\delta \stackrel{\text{def}}{=} \delta_{a,b}^{(ab)} = \delta_{ab,cd}^{(bc)}$, and $\delta_{b,c}^{(bc)} = 0$, then we have:

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

Since each of ρ_r, ρ_r' and $\rho_r^{(bc)}$ is close to (and is not more than) 1, if we assume that $\varepsilon_{abcd} \stackrel{\text{def}}{=} \varepsilon_{abcd,a}^{(ab)} = \varepsilon_{abcd,b}^{(ab)} = \varepsilon_{abcd,b}^{(bc)} = \varepsilon_{abcd,c}^{(bc)}$ and $\varepsilon_{ab} \stackrel{\text{def}}{=} \varepsilon_{ab,a}^{(ab)} = \varepsilon_{ab,b}^{(ab)}$, then we also have:

$$
\begin{aligned}
\frac{1 - \varepsilon^{(bc)}}{1 - \varepsilon^{(ab)}} &\leq 1 + \frac{\varepsilon_{ab}}{1 - 2\varepsilon_{ab} - (1 - \varepsilon_{ab}) \varepsilon_{abcd} / (1 - \varepsilon_{abcd})} \\
&\approx 1 + \varepsilon_{ab}
\end{aligned}
$$

4 General path-topology

In the previous sections, we show how to infer loss rates on 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. 4.3, we assume unicast probe packets can be sent from node 0 to 2, 1 to 2 and 1 to 3 along paths a, b and c, respectively. We consider a tree traversed by P_b and P_c, and an inverse tree traversed by P_a

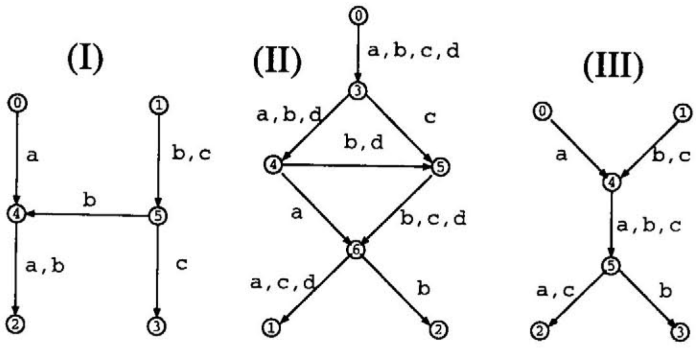

Figure 4.3: General path-topologies

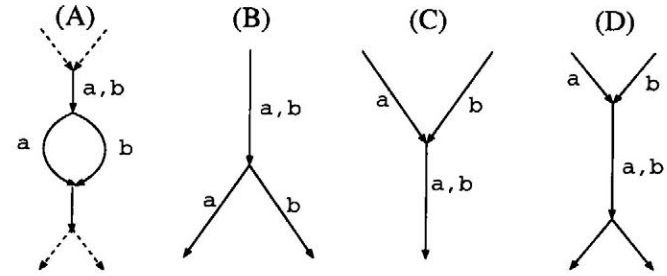

Figure 4.4: Relations between two paths sharing a common part

and P_b, and four independent measurements on them, i.e., M_a, M_{ab}, $M_{b.c}$ and M_c. Combining both inference for a tree and an inverse tree, we can infer x_a, x_b, x_c, x_{ab} and x_{bc}.

$$\hat{x}_a \stackrel{\text{def}}{=} \frac{(y_a^{(a)} - y_a^{(ab)}) + (y_b^{(b.c)} - y_b^{(ab)}) + y_{ab}^{*(ab)}}{y_b^{(b.c)} - y_{ab}^{'(ab)}/y_a^{(a)}}$$

$$\hat{x}_b \stackrel{\text{def}}{=} \frac{((y_a^{(a)} - y_a^{(ab)}) + (y_b^{(b.c)} - y_b^{(ab)}) + y_{ab}^{*(ab)})y_{bc}^{(b.c)}}{(y_a^{(a)}y_b^{(b.c)} - y_{ab}^{'(ab)})y_c^{(b.c)}}$$

$$\hat{x}_c \stackrel{\text{def}}{=} \frac{y_c^{(c)}y_{bc}^{(b.c)}}{y_b^{(b.c)}y_c^{(b.c)}}$$

$$\hat{x}_{ab} \stackrel{\text{def}}{=} \frac{y_a^{(a)}y_b^{(b.c)} - y_{ab}^{'(ab)}}{(y_a^{(a)} - y_a^{(ab)}) + (y_b^{(b.c)} - y_b^{(ab)}) + y_{ab}^{*(ab)}}$$

$$\hat{x}_{bc} \stackrel{\text{def}}{=} \frac{y_b^{(b.c)}y_c^{(b.c)}}{y_{bc}^{(b.c)}}$$

In what follows, we explain how this inference works in general, based on the principle and notations in [7]. First we consider a set ***Path*** of observable 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 $\Delta \subset 2^{\boldsymbol{Path}}$, thus the set of all links are denoted by $\{l_R | R \in \Delta\}$. For conciseness, we use abc instead of $\{a, b, c\}$ as an expression of set R ($\subset$ $\boldsymbol{Path}$) consisting of paths a, b and c, for example.

Let $\boldsymbol{L}_r$ be a set of links included by path $r \in \boldsymbol{Path}$, $\overline{\boldsymbol{L}}(R)$ be a set of links included by all paths in $R \subset \boldsymbol{Path}$ (i.e., $\overline{\boldsymbol{L}}(R) = \bigcap_{r \in R} \boldsymbol{L}_r$), and $\underline{\boldsymbol{L}}(R)$ be a set of links included by at least one path in R (i.e., $\underline{\boldsymbol{L}}(R) = \bigcup_{r \in R} \boldsymbol{L}_r$). Furthermore, let $A(L)$ denote an occurrence probability of "no loss" on set L of links. We assume losses on individual links in L are independent so that $A(L) = \prod_{l_R \in L} x_R$, where x_R is the "no loss" rate on link l_R ($R \in \Delta$), and also assume each x_R is positive (non-zero).

Then it can be shown that $\{x_R | R \in \Delta\}$ is uniquely determined by $\{A(\overline{\boldsymbol{L}}(R)) | R \in \Delta\}$ under the above assumptions. Therefore, our goal is to infer $A(\overline{\boldsymbol{L}}(R))$ for each link l_R from end-to-end path measurement on a general path-topology. If $\overline{\boldsymbol{L}}(R) = \boldsymbol{L}_a$ for path $\exists a \in R$, then $A(\overline{\boldsymbol{L}}(R)) = A(\boldsymbol{L}_a) = y_a$ is simply obtained from measurement of path a.

Otherwise, we consider whether every pair of paths in R has a shared link besides (outside) $\overline{\boldsymbol{L}}(R)$ or not. The cases in which all paths in R are mutually overlapped outside $\overline{\boldsymbol{L}}(R)$, like a, b, c and d for l_{abcd} in (II) of Fig. 4.3, 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 $\overline{\boldsymbol{L}}(R)$. For such a pair (a, b), it can be shown that $\overline{\boldsymbol{L}}(R) = \overline{\boldsymbol{L}}(ab)$ and $(\boldsymbol{L}_a - \overline{\boldsymbol{L}}(ab)) \cap (\boldsymbol{L}_b - \overline{\boldsymbol{L}}(ab)) = \emptyset$. For example, for l_{abd} in (II) of Fig. 4.3, we see $\overline{\boldsymbol{L}}(abd) = \{l_{abd}, l_{abcd}\} = \overline{\boldsymbol{L}}(ab)$, $\boldsymbol{L}_a - \overline{\boldsymbol{L}}(ab) = \{l_a, l_{acd}\}$ and $\boldsymbol{L}_b - \overline{\boldsymbol{L}}(ab) = \{l_b, l_{bd}, l_{bcd}\}$. Therefore, what we should do is to obtain $A(\overline{\boldsymbol{L}}(ab))$ for an appropriate pair (a, b) of paths in R.

Let us consider topological relations among $\boldsymbol{L}_a$, $\boldsymbol{L}_b$ and $\overline{\boldsymbol{L}}(ab)$. In general, there exist four cases: (A), (B), (C), (D) of Fig. 4.4. (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 $A(\overline{\boldsymbol{L}}(ab))$ can be inferred as mentioned in the previous sections.

Finally, in case (D), it is difficult to detect whether the two probes enter the shared part $\overline{\boldsymbol{L}}(ab)$ "closely" or not in end-to-end measurement of paths a and b. This also makes it difficult to infer $A(\overline{\boldsymbol{L}}(ab))$ directly in general. However, if there exists a set L of links including $\overline{\boldsymbol{L}}(ab)$ satisfying that both $A(L)$ and $A(L - \overline{\boldsymbol{L}}(ab))$ are inferred, then $A(\overline{\boldsymbol{L}}(ab))$ can be obtained as $A(L)/A(L - \overline{\boldsymbol{L}}(ab))$. For example, for l_{abc} in (III) of Fig. 4.3, we can infer x_{ac} from measurements $M_{c.b}$ on a tree, and $x_{ac}x_{abc}$ from (M_a, M_c, M_{ac}) on an inverse tree, and consequently x_{abc} can be obtained.

5 Simulation

We examine four path-topologies shown in Fig. 4.1 and Fig. 4.2 by using the network simulator ns [32]. 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^i 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^i and T_{j+1}^i, or T_n^i and T_1^{i+1}) change randomly within some range. The range of time intervals we choose are [8,24] msec for Fig. 4.1, [50,80] msec for (I) of Fig. 4.2, and [70,100] msec for (II) of Fig. 4.2, 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 $\times$ 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 $P_{r'}$ 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. Although this (too simple) strategy may not be optimal, our simulations show acceptable results.

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. 4.2; 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. 4.2, respectively. Time intervals between executing adjacent trials change randomly within [0.6,0.9] msec for (I) of Fig. 4.2, and [0.7,1] msec for (II) of Fig. 4.2, respectively

In Fig. 4.5, Fig. 4.6 and Fig. 4.7, 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. 4.5, 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. 4.6, we see inaccuracy and slow convergence with instability. In Fig. 4.7, however, we can see better accuracy and stability in a high bandwidth network (with many background TCP flows).

In general, the above errors are mainly due to 1) correlation between links in a trial, i.e., spatial dependence, 2) correlation between trials in different measurements, i.e., temporal dependence, 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 by observing the inter-arrival time of these probes at a receiver.

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 work, we have presented a method of inferring packet loss rates on 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.

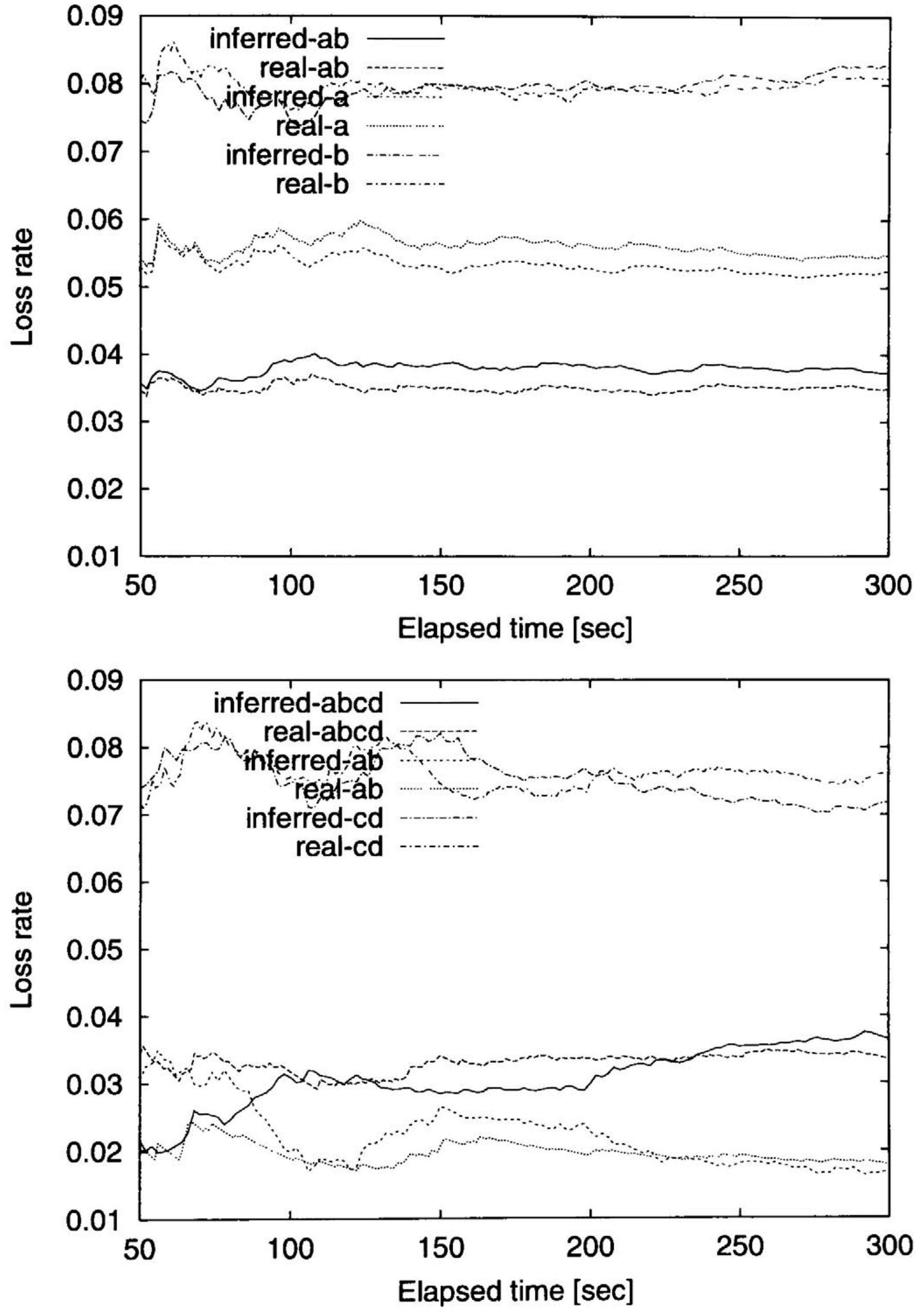

Figure 4.5: Loss rates inference on tree path-topologies

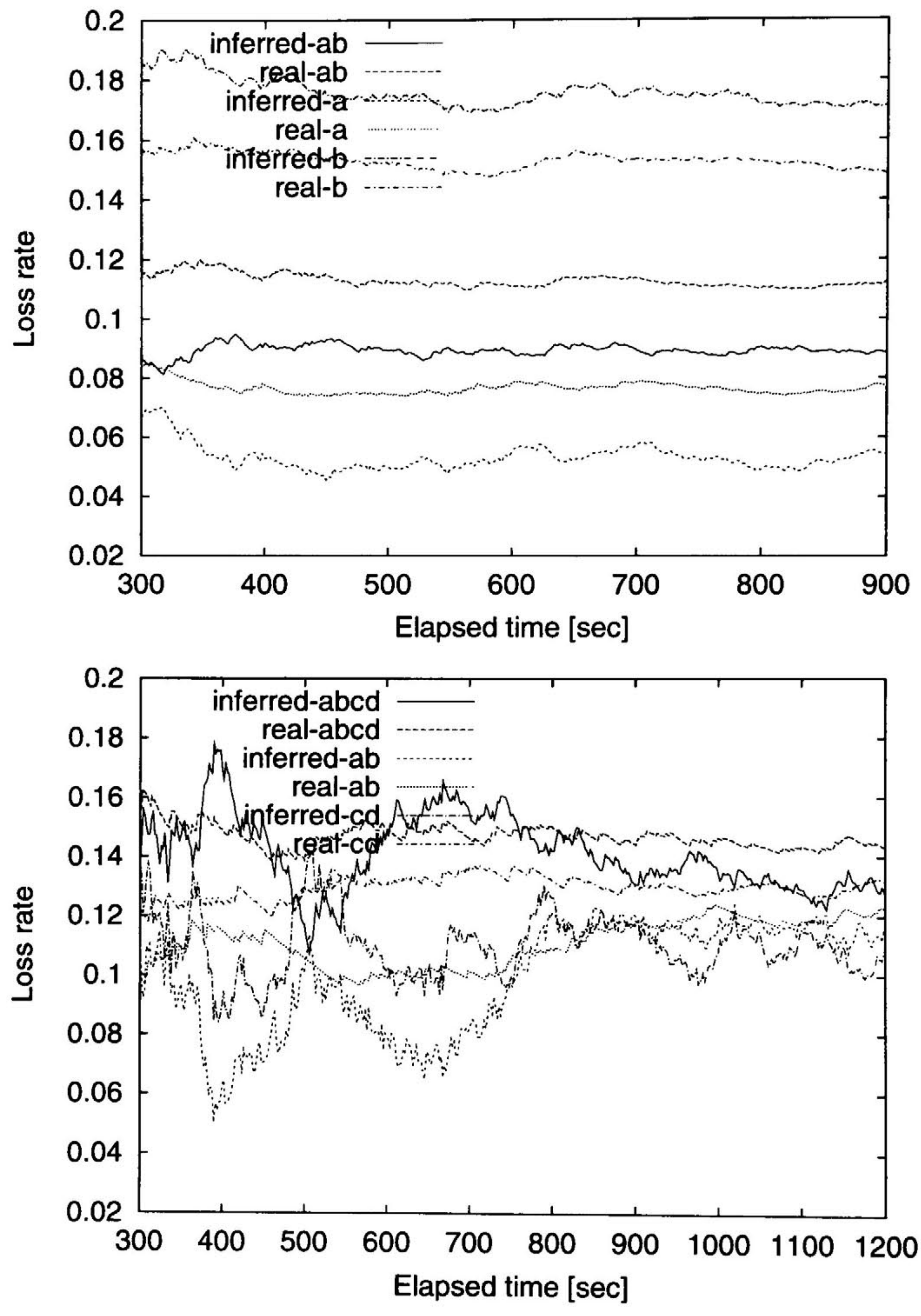

Figure 4.6: Loss rates inference on inverse tree path-topologies

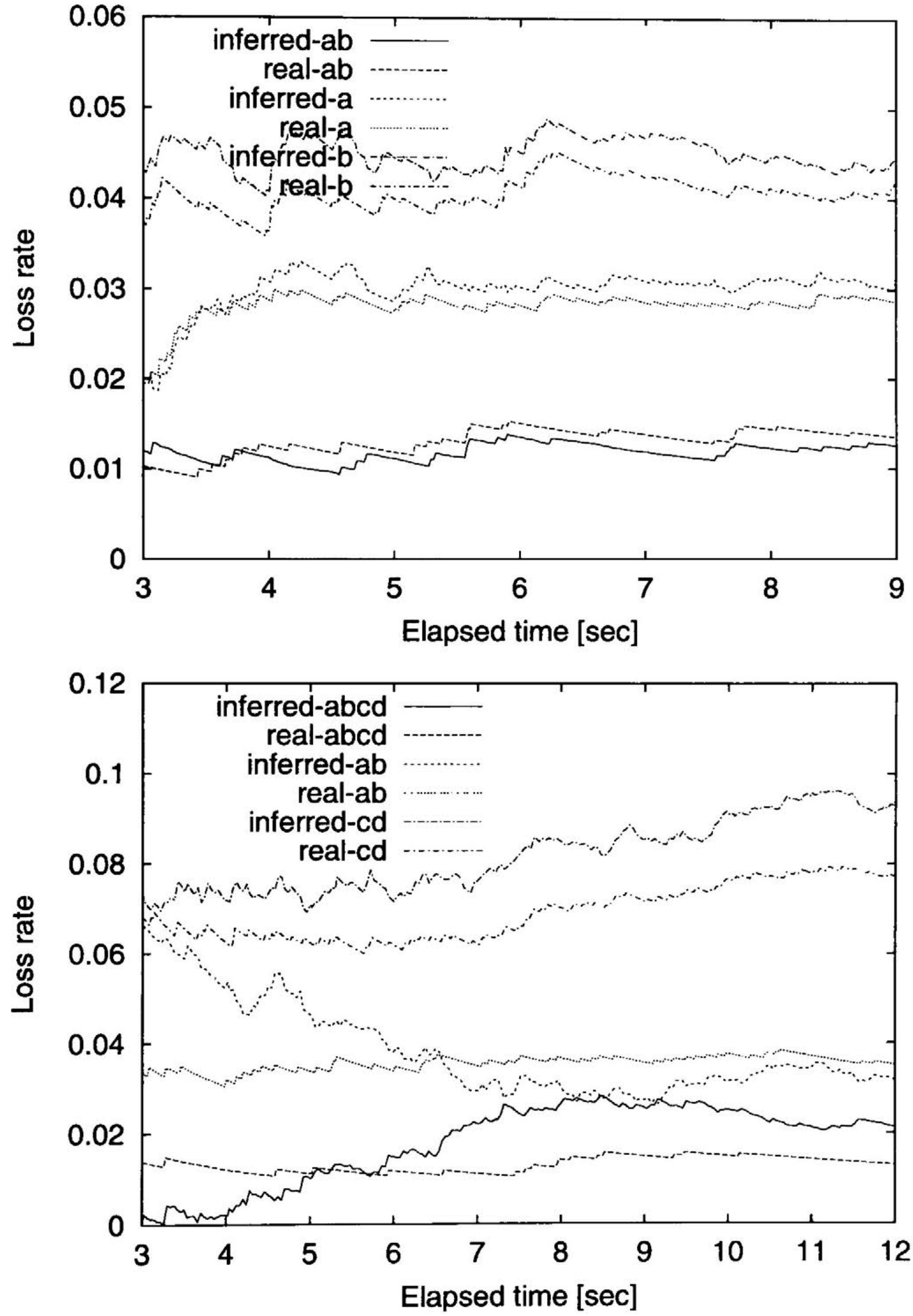

Figure 4.7: Loss rates inference on inverse trees on high speed links

Chapter 5

Inferring traffic flow characteristics from aggregated-flow measurement

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 whose characteristics should be measurable. For example, a statistical perspective of global traffic flows has been considered as an important key to network management, e.g., configuration, provisioning and traffic engineering. Nonetheless, it is expensive or sometime difficult to measure statistics of each flow directly (although some researches try it by capturing and analyzing raw traffic data at some routers in a network, e.g., [33]). Therefore, it is of practical importance to infer unobservable statistical characteristics of individual flows from characteristics of the aggregated-flows, which are easily observed at some points in the network.

In this work, we propose a new approach to infer characteristics of each flow from given characteristics of the aggregated-flows. We regard a "flow" as a series of (some kind of) packets from an origin node to a destination node. Here we intend that a "node" does not correspond to a single host but to a large set of hosts (i.e., a network or a set of networks), and thus a "flow" is not related to source-destination IP addresses directly but to a partial topological structure of routing paths in a network under a fixed routing scheme. Let us consider flows $f_1, f_2, .., f_p$, and directed-links $l_1, l_2, .., l_q$. Each link l_i is associated with a set F_i of flows where all (and only) flows in the set F_i pass through the link l_i. The problem is to infer characteristics of each flow f_j ($1 \leq j \leq p$) from observation of the aggregated-flow F_i at the link l_i ($1 \leq i \leq q$).

The arrival rate, i.e., the number of arriving traffic bytes or packets in a unit time-interval, is a typical example of flow characteristics. The inference for arrival rates is known as the origin-destination (OD) traffic matrix problem. Originally, the OD traffic matrix problem is to infer unobservable OD flow traffic volume (byte counts) from the link traffic volume (byte counts) measured at some routers' interfaces, and several researches studied this problem (e.g., [2], [3], [14] and [15]). They assumed that all OD byte counts were modeled by independent normal (or Poisson) distributions, and were *iid* over successive measurement time-intervals or something like that. Then they employed the EM method to calculate the maximum likelihood estimators for parameters of the models, which were expected to perform well for flows having relative large traffic volume with a relative long measurement period.

Our approach is different from the above approach. It is based on the same principle as a general framework we previously proposed for inferring network-internal (i.e., link) characteristics from given end-to-end path characteristics ([7]), which can be regarded as a generalization and extension of [17]. We infer some characteristics (occurrence probabilities of some discrete states) of each flow from correlation among characteristics of aggregated-flows at different links with an arbitrary network (routing) topology.

As concerns the case of inferring the arrival rates, we infer a discrete distribution of the number of arriving packets in a measurement time-interval. Our method is applicable to models with general distributions that cannot be captured by normal-based parametric models. Moreover the computational cost is less than the above existing approach using the EM method. On the other hand, our method requires that the number of arrivals in a measurement time-interval should sometime take 0. Therefore, in this work, instead of the inference for the arrival rate of the whole traffic, we focus especially on the inference for the arrival rate of some kind of special packets, where by "special" we mean that such packets do not always arise in each measurement interval. Of course, this condition is relative to the scale of measurement periods, and thus a very short measurement period allows us to infer the arrival rates of the whole traffic. However, our intention is to infer some irregular events with a distribution that is not covered by the existing normal-based methods for the OD traffic matrix inference. For example, we intend to infer the arrival rates of (some kind of) ICMP packets, or packets with (some kind of) IP options. Other end-to-end events related to TCP or application layers can be regarded as targets of our method if routers count such events. It is expected that the dynamics of such special events on each flow often indicate useful information, e.g., detection of anomalous congestion, malicious activities, or deployment of some optional functions.

Note that to count such events (passages of such packets) is quite easier than to record or classify the source-destination IP addresses of the arriving packets. Furthermore, since a "flow" we mention here is related to a partial topological structure of routing paths, it is not always easy to map the source-destination IP addresses of the arriving packet to the flow to which the packet belongs. In addition, we need to treat cases that source IP addresses are not reliable (e.g., watching malicious packets).

The remainder of this chapter is organized as follows. Section 2 describes a general model consisting of flows, links, aggregated-flows, and characteristics of flows to be inferred. Section 3 explains how to apply our inference method to inferring arrival rates of packets with some examples . Section 4 shows simulation results. Finally Section 5 concludes this work.

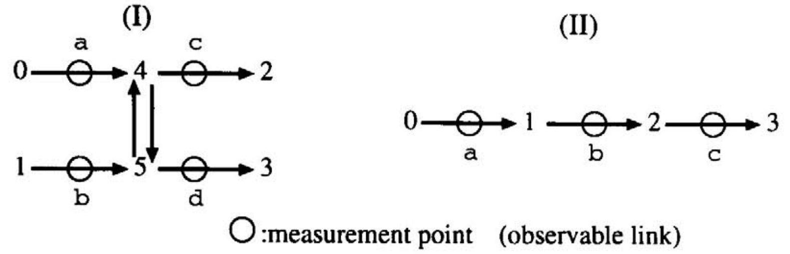

Figure 5.1: Examples of the network model (nodes, links and flows)

2 General model

According to the framework in [7], we define a general model for our inference method.

2.1 Links and flows

Let us consider a directed-graph consisting nodes (vertexes) and directed-links (edges), and flows on the graph. Each flow is a series of some kind of packets from an origin node to a destination node along a fixed sequence of directed-links without a loop. We call a set of flows passing through the same link by an aggregated-flow passing through that link.

Link is defined as a set consisting of all observable links at which the characteristic of the aggregated-flow passing through the link can be obtained from observations. Typically, ***Link*** corresponds to a set of (incoming and/or outgoing) interfaces of one or more routers in a network. ***Flow*** is defined as a set consisting of all flows passing through at least one of links in ***Link***.

For each flow $e \in \boldsymbol{Flow}$, we define passing-link set $\boldsymbol{R}(e)$ as a set of links in ***Link*** that are passed through by the flow e. Note that $\boldsymbol{R}(e) \neq \emptyset$. We denote a set consisting of all passing-link sets by Δ: $\Delta \stackrel{\text{def}}{=} \{\boldsymbol{R}(e) | e \in \boldsymbol{Flow}\}$. Then, for $R \in \Delta$, we define "distinguishable flow" f_R as a set $\{e | \boldsymbol{R}(e) = R\}$ of flows. In other words, we label each (distinguishable) flow by its passing-link set R. We also denote a set consisting all "distinguishable flow"s by Δ^*. In what follows, we use a term "flow" as a "distinguishable flow".

For each link $l \in \boldsymbol{Link}$, we define $\boldsymbol{F}_l$ as a set of flows (an aggregated-flow) passing through the link l: $\boldsymbol{F}_l \stackrel{\text{def}}{=} \{f_R | l \in R,\ R \in \Delta\}$, for $l \in \boldsymbol{Link}$. We assume that $\boldsymbol{F}_{l_1} \neq \boldsymbol{F}_{l_2}$ if $l_1 \neq l_2$.

Fig. 5.1 shows some examples. Both (I) and (II) have four end nodes 0, 1, 2 and 3. In (I), $\boldsymbol{Link} = \{a, b, c, d\}$, and $\Delta = \{ac, ad, bc, bd\}$, where f_{ac} is a flow from node 0 to 2, f_{ad} is from 0 to 3, f_{bc} is from 1 to 2, and f_{bd} is from 1 to 3, respectively. Aggregated-flows are $\boldsymbol{F}_a = \{f_{ac}, f_{ad}\}$, $\boldsymbol{F}_b = \{f_{bc}, f_{bd}\}$, $\boldsymbol{F}_c = \{f_{ac}, f_{bc}\}$, and $\boldsymbol{F}_d = \{f_{ad}, f_{bd}\}$.

In (II), $\boldsymbol{Link} = \{a, b, c\}$, and $\Delta = \{a, b, c, ab, bc, abc\}$, where f_a is a flow from node 0 to 1, f_b is from 1 to 2, f_c is from 2 to 3, f_{ab} is from 0 to 2, f_{bc} is from 1 to 3, and f_{abc} is from

0 to 3, respectively. Aggregated-flows are $\boldsymbol{F}_a = \{f_a, f_{ab}, f_{abc}\}$, $\boldsymbol{F}_b = \{f_b, f_{ab}, f_{bc}, f_{abc}\}$, and $\boldsymbol{F}_c = \{f_c, f_{bc}, f_{abc}\}$.

2.2 Characteristics of flows to be inferred

We define some notations to describe characteristics of a flow or a set of flows as follows.

- $\boldsymbol{M} = \{0, 1, .., M\}$: A set of integers that represents states related to characteristic of a flow or a set of flows.
- $X_R(m)$: An event that the state of a flow f_R ($R \in \Delta$) is $m \in \boldsymbol{M}$.
- $V(F)(m)$: An event that the state of a set F of flows is $m \in \boldsymbol{M}$. If $F = \{f_R\}$ then $V(\{f_R\})(m) = X_R(m)$.
- $Y_l(m) \stackrel{\text{def}}{=} V(\boldsymbol{F}_l)(m)$: An event that the state of an aggregated-flow $\boldsymbol{F}_l$ is $m \in \boldsymbol{M}$.

We also define the occurrence probabilities related to the above events.

$$\begin{aligned} x_R(m) &\stackrel{\text{def}}{=} \Pr[X_R(m)] \\ \underline{y}(R)(m) &\stackrel{\text{def}}{=} \Pr[\bigcup_{k=0}^{m} \bigcup_{l \in R} Y_l(k)] \\ \overline{y}(R, R')(m) &\stackrel{\text{def}}{=} \Pr[\bigcap_{l \in R - R'} Y_l(0) \cap \bigcap_{l \in R'} Y_l(m)] \end{aligned}$$

Roughly speaking, if $\{V(F)(m) | F \subset \Delta^*, F \neq \emptyset, m \in \boldsymbol{M}\}$ satisfies the following conditions, then we can uniquely determine unobservable probabilities $\{x_R(m) | 0 \leq m \leq M,\ R \in \Delta\}$ from observable probabilities $\{\underline{y}(R)(m),\ \overline{y}(R', R'')(m) | 0 \leq m \leq M,\ R \in \Psi, R' \in \Psi', R''$ is a subset of $R'\}$, where Ψ and Ψ' are appropriate subsets of $2^{\boldsymbol{Link}}$. Therefore, by using inference of $\{\underline{y}(R)(m)\}$ and $\{\overline{y}(R', R'')(m)\}$, we can infer $\{x_R(m)\}$, which is our goal.

The conditions that $\{V(F)(m)\}$ should satisfy are:

1. $\Pr[V(F)(i) \cap V(F)(j)] = 0$ if $i \neq j$
2. $0 < \Pr[V(F)(0)]$
3. $V(F)(i)$ and $V(F')(j)$ are independent if $F, F' \neq \emptyset,\ F \cap F' = \emptyset,\ i, j \in M$
4. A certain technical condition on the relation between $V(F + F')(m)$ and $V(F)(s_1) \cap V(F')(s_2)$ for $\forall s_1, s_2 \leq m$. Note that if $V(F)(m)$ satisfies

$$V(F + F')(m) = \sum_{j=0}^{m} V(F)(j) \cap V(F')(m - j)$$

for $0 \leq m \leq M$, then it also satisfies that condition.

3 Inference method for arrival rates

3.1 General description

We explain how to infer the distribution of the arrival rates of packets on each flow. Let T be a measurement period, and $[(i-1)T, iT)$ be the i-th measurement interval. For each $i \in \{1, 2, .., n\}$, let w_i^l and v_i^R be the number of target packets arriving to aggregated-flow $\boldsymbol{F}_l$ and flow f_R in the i-th measurement interval, respectively. The number of arrivals ranges from 0 to M. We assume $\{v_i^R | 1 \leq i \leq n\}$ is *iid* for each R, and let v^R be the number of target packets arriving to flow f_R in a measurement interval. We ignore the problem of transmission time delay and clock synchronization between different measurement points (links), and thus assume that w_i^l can be observed at link l. Our goal is to infer the distribution of v^R (i.e., $x_R(m) \stackrel{\text{def}}{=} \Pr[v^R = m]$ for $m = 0, 1, .., M$) for each flow f_R, which leads to the average arrival rate $\sum_{m=0}^{M} m x_R(m)/T$ of f_R.

We have the relation (linear equations) among observable w_i^l and unobservable v_i^R for each $i \in \{1, 2, .., n\}$.

$$w_i^l = \sum_{f_R \in \boldsymbol{F}_l} v_i^R \quad \text{for } l \in \boldsymbol{Link} \tag{5.1}$$

If the above relation (5.1) is uniquely solvable, then we have the map $\boldsymbol{H}^R$ such that $v_i^R = \boldsymbol{H}^R(w_i^l;\ l \in \boldsymbol{Link})$. In this case, for a sufficient large n, we can directly estimate $x_R(m)$ as:

$$\begin{aligned}\hat{x}_R(m) &\stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(v_i^R = m) \\ &= \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(\boldsymbol{H}^R(w_i^l;\ l \in \boldsymbol{Link}) = m)\end{aligned}$$

where $\mathbf{1}(.)$ denotes the indication function.

Hereafter, we consider cases in which relation (5.1) is not uniquely solvable. Even if each v_i^R cannot be uniquely solved in a deterministic way, we show that the distribution of v^R can be determined in a statistical way.

In accordance with the general model mentioned in the previous section, let $V(F)(m)$ be an event that $\sum_{f_R \in F} v^R = m$. Then $X_R(m)$ (resp. $Y_l(m)$) is an event that $v^R = m$ (resp. $\sum_{f_R \in \boldsymbol{F}_l} v^R = m$) for $F \subset \Delta^*$, $R \in \Delta$, $l \in \boldsymbol{Link}$, and $0 \leq m \leq M$.

Let us check the conditions on V in the previous section. Condition 1. is clear. Moreover we can show $V(F + F')(m) = \sum_{j=0}^{m} V(F)(j) \cap V(F')(m-j)$ for $0 \leq m \leq M$, so that condition 4. is satisfied. Whether condition 2. and 3. are satisfied or not depends on both the nature of the target traffic and measurement period T, and thus these conditions are requirements (restrictions) for our approach. Note that condition 3. is expected to be satisfied approximately because of diversity of traffic in actual networks.

Several estimators can be derived by a framework in [7]. For the examples in the next section, since the MIP number of each flow is 2 or 0 (in term of [7]), we can employ a basic estimator based on the relation among the shared-part and two independent-parts. For concise descriptions, we use the following definitions of $\tilde{x}_R(m)$, $y_l(m)$ and $y_{l_1 l_2}(m)$ for $R \in \Delta$ and $l, l_1, l_2 \in \boldsymbol{Link}$.

$$\begin{aligned}
\tilde{x}_R(m) &\stackrel{\text{def}}{=} \Pr[v^R \leq m] = \Pr[\bigcup_{k=0}^{m} X_R(k)] = \sum_{k=0}^{m} x_R(k) \\
y_l(m) &\stackrel{\text{def}}{=} \underline{y}(l)(m) = \Pr[\bigcup_{k=0}^{m} Y_l(k)] \\
y_{l_1 l_2}(m) &\stackrel{\text{def}}{=} \underline{y}(l_1)(m) + \underline{y}(l_2)(m) - \underline{y}(l_1 l_2)(m) \\
&= \Pr[\bigcup_{k=0}^{m} Y_{l_1}(k) \cap \bigcup_{k=0}^{m} Y_{l_2}(k)]
\end{aligned}$$

where $\underline{y}(R)(m)$ is defined in the previous section.

For a sufficient large n, we can estimate $y_l(m)$ (resp. $y_{l_1 l_2}(m)$) as $\hat{y}_l(m)$ (resp. $\hat{y}_{l_1 l_2}(m)$):

$$\begin{aligned}
\hat{y}_l(m) &\stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(w_i^l \leq m) \\
\hat{y}_{l_1 l_2}(m) &\stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(w_i^{l_1} \leq m \wedge w_i^{l_2} \leq m)
\end{aligned}$$

Then we can show that $\{x_R(m) | R \in \Delta, 0 \leq m \leq M\}$ are determined from $\{y_l(m), y_{l_1 l_2}(m) | l, l_1, l$ $\boldsymbol{Link}, 0 \leq m \leq M\}$ by solving the relation among x_R, y_l and $y_{l_1 l_2}$. To be more precise, we have the map $\boldsymbol{G}_{R,m}$ such that $x_R(m) = \boldsymbol{G}_{R,m}(y_l(i), y_{l_1 l_2}(i);\ l, l_1, l_2 \in \boldsymbol{Link}, 0 \leq i \leq m)$. Consequently we infer $x_R(m)$ as follows.

$$\hat{x}_R(m) \stackrel{\text{def}}{=} \boldsymbol{G}_{R,m}(y_l(i), y_{l_1 l_2}(i);\ l, l_1, l_2 \in \boldsymbol{Link}, 0 \leq i \leq m)$$

3.2 Examples

We provide some examples in Fig. 5.2. For (I) in Fig. 5.2 (modeled by (I) in Fig. 5.1), we observe aggregated-flows at four interfaces a, b, c and d of a router, and can obtain, $\hat{y}_a(m)$, $\hat{y}_b(m)$, $\hat{y}_c(m)$, $\hat{y}_d(m)$, $\hat{y}_{ac}(m)$, $\hat{y}_{ad}(m)$, $\hat{y}_{bc}(m)$, and $\hat{y}_{bd}(m)$, for $0 \leq m \leq M$. Although the number of links is equal to the number of flows, relation (5.1) is not uniquely solvable because of the implicit restriction $w_i^a + w_i^b = w_i^c + w_i^d$.

Let us infer x_{ac} from $\hat{y}_a$, $\hat{y}_c$ and $\hat{y}_{ac}$, for example. We have the following relation.

$$\begin{aligned}
y_a(m) &= \sum_{i=0}^{m} x_{ac}(i) \tilde{x}_{ad}(m-i) \\
y_c(m) &= \sum_{i=0}^{m} x_{ac}(i) \tilde{x}_{bc}(m-i) \\
y_{ac}(m) &= \sum_{i=0}^{m} x_{ac}(i) \tilde{x}_{ad}(m-i) \tilde{x}_{bc}(m-i)
\end{aligned}$$

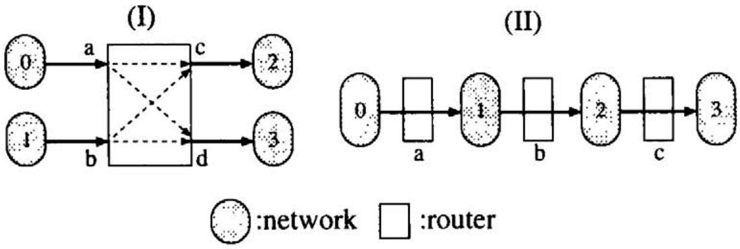

Figure 5.2: Examples of actual networks

Then we can solve it inductively as follows.

$$\begin{aligned}
x_{ac}(m) &= \boldsymbol{G}_{ac,m}(y_a(i), y_c(i), y_{ac}(i);\ 0 \le i \le m) \\
&\stackrel{\text{def}}{=} \frac{x_{ac}(0)}{2}\Big(E_{ac}(m) - \\
&\qquad \sqrt{E_{ac}^2(m) - \frac{4C_a(m)C_c(m)}{y_a(0)y_c(0)} + \frac{4D_{ac}(m)}{y_{ac}(0)}}\Big) \\
\tilde{x}_{ad}(m) &= \frac{C_a(m) - x_{ac}(m)\tilde{x}_{ad}(0)}{x_{ac}(0)} \\
\tilde{x}_{bc}(m) &= \frac{C_b(m) - x_{ac}(m)\tilde{x}_{bc}(0)}{x_{ac}(0)}
\end{aligned}$$

where

$$\begin{aligned}
C_a(m) &\stackrel{\text{def}}{=} y_a(m) - \sum_{i=1}^{m-1} x_{ac}(i)\tilde{x}_{ad}(m-i) \\
C_c(m) &\stackrel{\text{def}}{=} y_c(m) - \sum_{i=1}^{m-1} x_{ac}(i)\tilde{x}_{bc}(m-i) \\
D_{ac}(m) &\stackrel{\text{def}}{=} y_{ac}(m) - \sum_{i=1}^{m-1} x_{ac}(i)\tilde{x}_{ad}(m-i)\tilde{x}_{bc}(m-i) \\
E_{ac}(m) &\stackrel{\text{def}}{=} \frac{C_a(m)}{y_a(0)} + \frac{C_c(m)}{y_c(0)} - 1
\end{aligned}$$

Using this $\boldsymbol{G}_{ac,m}$, we can infer $x_{ac}(m)$ as $\hat{x}_{ac}(m) \stackrel{\text{def}}{=} \boldsymbol{G}_{ac,m}(\hat{\boldsymbol{y}}(i);\ 0 \le i \le m)$. Similarly, we infer x_{ad} (from $\hat{y}_a$, $\hat{y}_d$ and $\hat{y}_{ad}$), x_{bc} (from $\hat{y}_b$, $\hat{y}_c$ and $\hat{y}_{bc}$), and x_{bd} (from $\hat{y}_b$, $\hat{y}_d$ and $\hat{y}_{bd}$). Figure 5.3 (I) shows the above relation between the flows and the aggregated-flows.

As mentioned previously, there exist similar estimators. For example, in the above derivation of an estimation for x_{ac}, we also have an estimator for $\tilde{x}_{ad}$, and so for x_{ad} ($x_{ad}(m) = \tilde{x}_{ad}(m) - \tilde{x}_{ad}(m-1)$). Another estimator for x_{ad} can be derived from a simple relation: $\tilde{x}_{ad}(m) = \overline{y}(ac, a)(m)/y_c(0)$.

For (II) in Fig. 5.2 (modeled by (II) in Fig. 5.1), we measure aggregated-flows at three routers a, b and c, and can obtain, $\hat{y}_a(m), \hat{y}_b(m), \hat{y}_c(m), \hat{y}_{ab}(m), \hat{y}_{ac}(m)$ and $\hat{y}_{bc}(m)$ for $0 \le m \le M$.

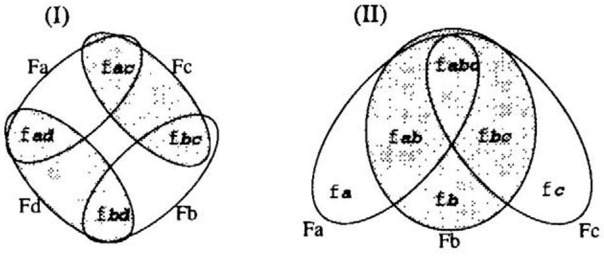

Figure 5.3: Relation between flows and aggregated-flows

Since the number of flows is greater than that of links, (5.1) is not uniquely solvable. In the same manner as the previous example, we infer x_{abc} (from $\hat{y}_a$, $\hat{y}_c$ and $\hat{y}_{ac}$) , $x_{ab}x_{abc}$ (from $\hat{y}_a$, $\hat{y}_b$, and $\hat{y}_{ab}$) , $x_{bc}x_{abc}$ (from $\hat{y}_b$, $\hat{y}_c$ and $\hat{y}_{bc}$). Figure 5.3 (II) shows the above relation between the flows and the aggregated-flows. Consequently, we can obtain x_{abc} , x_{ab} , x_{bc} , x_a , x_b , and x_c , separately.

Before ending this section, we note that to infer arrival rates as the number of arriving bytes (instead of arriving packets), we need to round the number of bytes by an adequate bin size that depends on acceptable cost and required accuracy.

4 Simulation

We examine two examples shown in Fig. 5.2 by using network simulator *ns* [32]. We dispatch a series of *pings* (i.e., ICMP echo request packets) as a target (i.e., to be inferred) flow, and a series of TCP packets as background traffic. We employ three types of distributions of interval time between adjacent pings: (U) uniform distribution in a range $[0.2 \times m, 1.8 \times m]$ with average m, (P) Pareto distribution with average m and shape ρ, (E) exponential distribution with average m. The bandwidth of each link is 1.5 Mbps with 10 ms of propagation delay. We count the number of pings arriving to each aggregated-flow in each measurement interval $[(i-1)T, iT)$ for $i = 1, 2, .., n$, where we choose 1 or 0.5 sec as measurement period T. Then we infer the average arrival rate of ping (the average number of pings arriving in 1 sec, i.e., pps) on each flow.

For (I) of Fig. 5.2, we generate three independent streams of type (U) ping with average interval time $m = 1.5$ sec on flow f_{ac}, a stream of type (P) ping with $m = 0.7$ and $\rho = 1.5$ on f_{ad}, a stream of type (E) ping with $m = 1.3$ on f_{bc}, and a stream of type (E) ping with $m = 0.3$ on f_{bd}, respectively. The theoretical average arrival rates (pps) are 2 (0.66×3) for f_{ac}, 1.4 for f_{ad}, 0.8 for f_{bc}, 3.3 for f_{bd}, respectively.

For (II) of Fig. 5.2, we generate three independent streams of type (U) ping with average interval time $m = 2.5$ sec on flow f_a, two independent streams of type (U) ping with $m = 2.5$

on flow f_b, a stream of type (P) ping with $m = 0.8$ and $\rho = 1.5$ on f_c, a stream of type (P) ping with $m = 1.5$ and $\rho = 1.3$ on f_{ab}, a stream of type (E) ping with $m = 1.3$ on f_{bc}, and a stream of type (E) ping with $m = 1.5$ on f_{abc}, respectively. The theoretical average arrival rates (pps) are 1.2 (0.4×3) for f_a, 0.8 (0.4×2) for f_b, 1.3 for f_c, 0.7 for f_{ab}, 0.8 for f_{bc}, 0.7 for f_{abc}, respectively.

Fig. 5.4, Fig. 5.5 and Fig. 5.6 show comparison between the real average arrival rates and the inferred average arrival rates in $[0, t)$ from $t = 0$ to 1500 (sec). They correspond to case (I) with measurement period $T = 1$ and 0.5 sec, case (II) with $T = 1$ sec, and case (II) with $T = 0.5$ sec, respectively.

In all figures, we do not see particular differences in inference accuracy among three types of distributions of ping intervals. In case (I), inference seems quite stable and accurate (Fig. 5.4). On the other hand, in case (II) with $T = 1$, although the inferred values roughly track the real values, we see inaccuracy and slow convergence with instability (Fig. 5.5), where we try to infer arrival rates on six individual flows from observation of only three aggregated-flows. Case (II) with $T = 0.5$ verifies that a shorter measurement period makes better stability and accuracy (Fig. 5.6).

5 Concluding remarks

In this work, we have presented a new approach to inferring statistical characteristics (occurrence probabilities of discrete states) of each flow from given characteristics of the aggregated-flows. By this approach, the distribution of the number of packets arriving to each flow (in a measurement interval) can be inferred from observation of the aggregated-flows at some links (interfaces of routers). Although our method requires some condition on dynamics of arrivals, it is applicable to general (irregular) distributions that cannot be captured by existing methods based on normal-based parametric models. For smaller average arrival rates and shorter measurement time-intervals, our discrete model is expected to be more suitable. Furthermore, our method is computationally light-weight, which makes real-time estimations feasible. We have provided some examples and shown simulation results, which indicate potential of our approach.

For development and deployment of practical methods based on the approach proposed in this work, we have many issues to examine and solve in actual networks, such as, reliability (limitation), distributed measurements, and scalability. Our method may also require additional functions to current routers in order to count some events in a short interval. However, this work has provided a starting point to establish a novel method for efficient inference of flow characteristics, which is useful and valuable for network management.

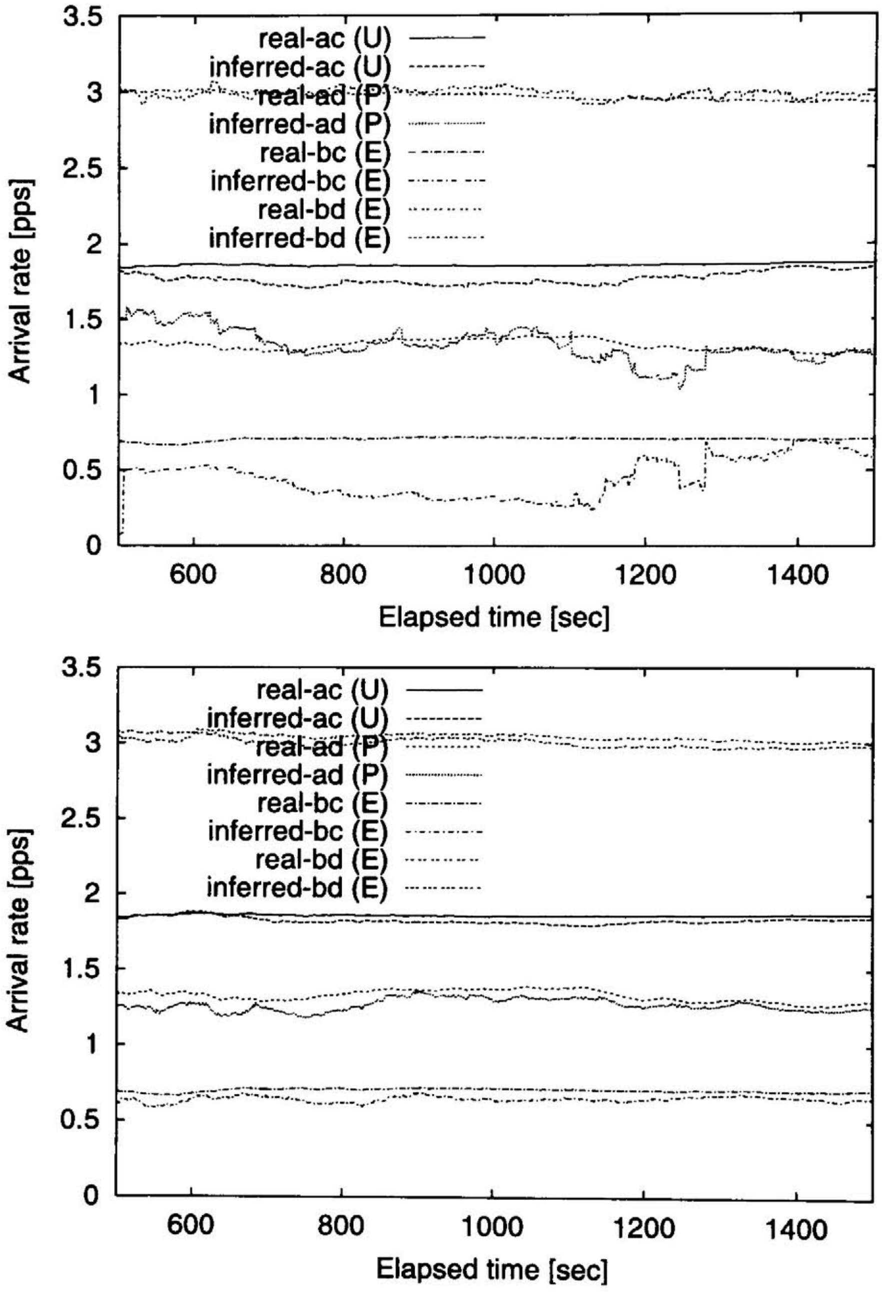

Figure 5.4: Arrival rates inference on each flow in (I) with $T = 1$ and $T = 0.5$

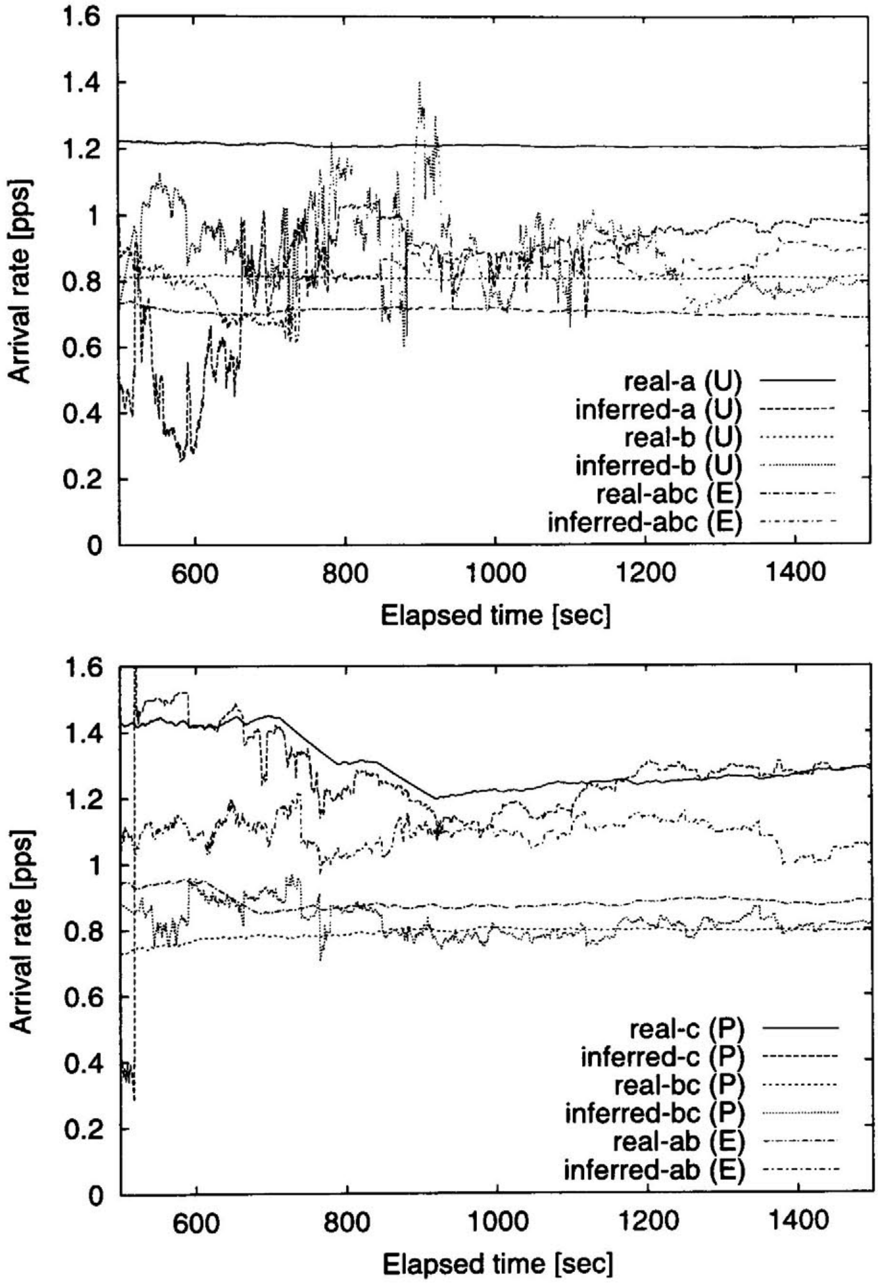

Figure 5.5: Arrival rates inference on each flow in (II) with $T = 1$

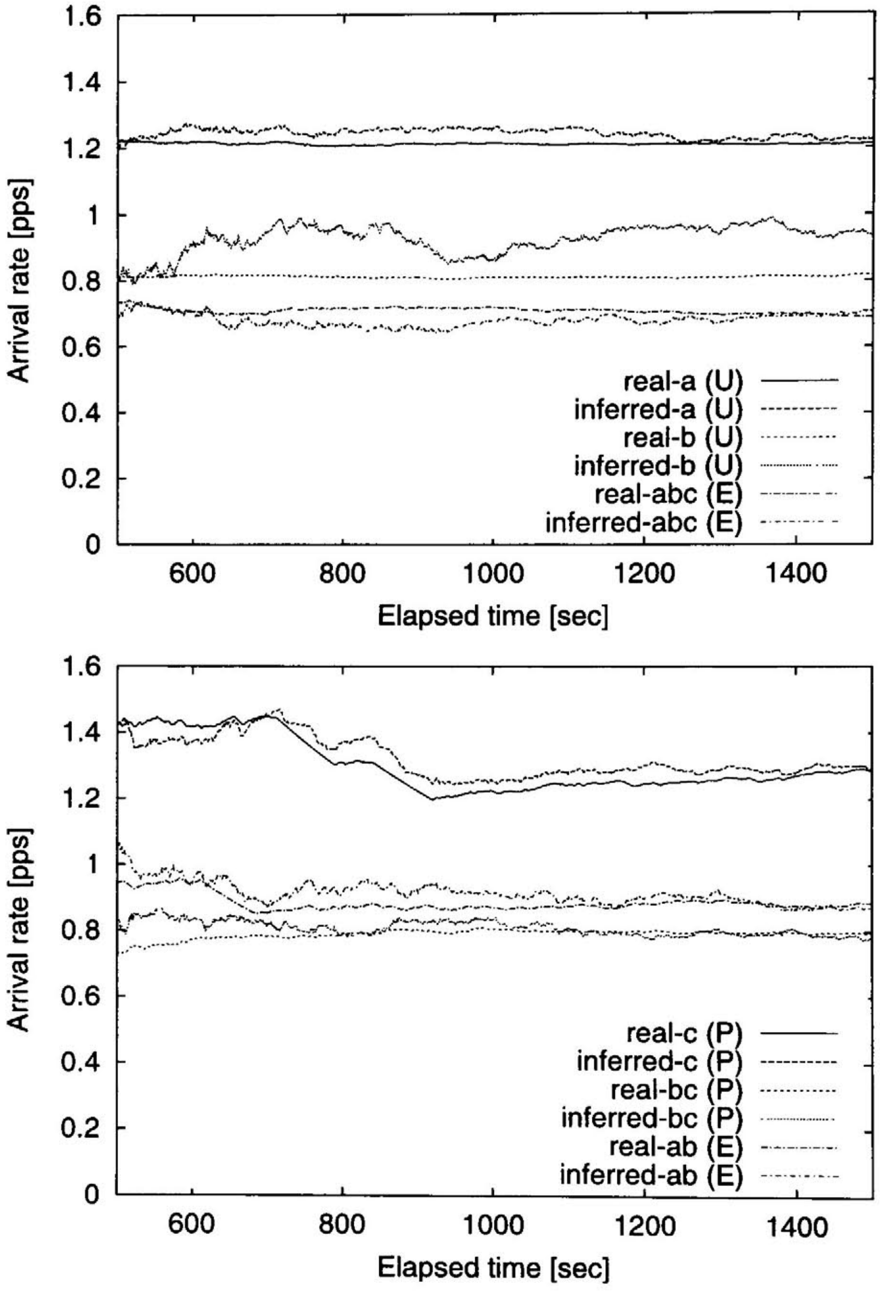

Figure 5.6: Arrival rates inference on each flow in (II) with $T = 0.5$

Chapter 6

Conclusion

In this thesis, I have studied on characteristics measurement and inference in statistical and indirect ways categorized as the "network tomography", which includes two typical forms; "inference of flow characteristics based on aggregated-flow measurements" and "inference of network-internal characteristics based on end-to-end path measurements".

My contributions are as follows. In Chapter 2, I have shown that a common framework of the network tomography has been roughly established, which can give a unified viewpoint to various practical methods in the network tomography, and thus can give useful insights into advantages and disadvantages of individual inference methods.

In Chapter 3, I have presented a principle of determining characteristics (i.e., occurrence probabilities of some states) of network-internal links from given characteristics of end-to-end paths with an arbitrary path topology, as a general framework for the "inverse function" approach. This generalization indicates that the "inverse function" approach is also applicable to the "inference of flow characteristics" based on aggregated-flow measurements, which is shown in Chapter 5.

In Chapter 4, I have presented a method, based on the "inverse function" approach with some extension, inferring packet loss rates on individual links from end-to-end measurement of unicast probe packets among several senders' and receivers' nodes. This method enables us to deal with inverse tree path topologies (and thus, almost general path topologies), while preceding methods have dealt with only tree path topologies.

In Chapter 5, I have presented a method, based on the "inverse function" approach, inferring arrival rates of (some kind of) packets on individual flows from measurement of aggregated-flows at several links (e.g., routers' interfaces). Although my method requires some condition on dynamics of arrivals due to the "inverse function" approach, it is applicable to general distributions that cannot be captured by preceding methods based on an MLE of normal-based parametric models.

Before ending this thesis, I remark research directions in this area. First, more deployment: although a number of simulation results and experimental results have been presented, there exists little practical use of the network tomography in the real Internet. We have many issues to examine and solve in actual networks, such as, reliability (limitation), distributed simultaneous measurements, and scalability.

There exist several methods (algorithms) to approximately solve an MLE equation in the "MLE solver" approach as well as several methods (estimators) derived from the "inverse function" approach, in which we should consider a trade-off between computational and/or operational costs and accuracy of the inference. We should also choose a number of detailed choices (parameters) in measurement and inference phases of a method. Analysis and verification of accuracy and reliability of the inference by each concrete method in actual environments are of practical importance. Note that the accuracy in experiments (on simulations or real networks) reported in the above researches (Table. 2.1) ranges from under 1% to over 20% (in some sense of "relative errors"), which depends not only on the type of the inference method but on the situation for which the method is applied. Moreover, acceptable accuracy also depends on the situation.

Second, more applications: the framework and principle (behind two typical forms) of the network tomography may be applicable to other inference problems not only in the IP layer but also in various applications. We also expect that the third form of the network tomography will be found.

Third, beyond independence assumptions: we often make assumptions of spatial and/or temporal independence to solve a problem. While several studies have dealt with the time-varying nature (non-stationarity), spatial independence is often essential to identifiability. Several studies have employed Bayesian inference to deal with some dependence, but prior distributions were often uncertain. It is a big challenge to find novel methodologies that can deal with temporal and spatial dependence.

Bibliography

[1] M. Tsuru and Y. Oie, "Internet characteristics measurement: Overview," *IPSJ Mag. (in Japanese)*, vol. 42, no. 2, pp. 192–197, Feb. 2001.

[2] R. J. Vanderbei and J. Iannone, "An EM approach to OD matrix estimation," SOR 94-04, Prinston Univ., 1994.

[3] Y. Vardi, "Network tomography: Estimating source-destination traffic intensities from link data," *J. Am. Stat. Assoc.*, vol. 91, pp. 365–377, 1996.

[4] R. Caceres, N. Duffield, J. Horowitz, and D. Towsley, "Multicast-based inference of network-internal loss characteristics," *IEEE Trans. Info. Theory*, vol. 45, no. 7, pp. 2462–2480, Nov. 1999.

[5] M. Coates and R. Nowak, "Network loss inference using unicast end-to-end measurement," in *Proc. ITC Conf. IP Traffic, Modeling and Management*, Monterey, Sept. 2000, pp. 28–1–28–9.

[6] N. Duffield, J. Horowitz, F. Lo Presti, and D.Towsley, "Network delay tomography from end-to-end unicast measurements," in *Proc. the 2001 Int. Workshop on Digital Communications*, Taormina, Italy, Sept. 2001.

[7] M. Tsuru, T. Takine, and Y. Oie, "Inferring link characteristics from end-to-end path measurements," in *Proc. IEEE ICC*, Helsinki, June 2001, pp. 1534–1538.

[8] M. Tsuru, T. Takine, and Y. Oie, "Inferring link loss rates from unicast-based end-to-end measurement," *IEICE Trans. Commun.*, vol. E85-B, no. 1, pp. 70–78, Jan. 2002.

[9] M. Tsuru, T. Takine, and Y. Oie, "Inferring traffic flow characteristics from aggregated-flow measurement," in *Proc. IEEE/IPSJ SAINT 2002*, Nara, Japan, Feb. 2002, pp. 256–261.

[10] A. Adams, T. Bu, T. Friedman, J. Horowitz, D. Towsley, R. Caceres, N. Duffield, F. Lo Presti, S. B. Moon, and V. Paxson, "The use of end-to-end multicast measurement for characterizing internal network behavior," *IEEE Commun. Mag.*, pp. 152–158, May 2000.

[11] M. Coates, A. Hero, R. Nowak, and B. Yu, "Large scale inference and tomography for network monitoring and diagnosis," Tech. Rep. 604, Stat. Dept, UCBerkeley, Aug 2001.

[12] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Royal Statistical Society Series B*, vol. 39, pp. 1–38, 1977.

[13] G. J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*, John Wiley and Sons, 1997.

[14] C. Tebaldi and M. West, "Bayesian inference on network traffic using link count data," *J. Am. Stat. Assoc.*, June 1998.

[15] J. Cao, D. Davis, S. V Wiel, and B. Yu, "Time-varying network tomography: Router link data," *J. Am. Stat. Assoc.*, vol. 95, pp. 1063–1075, 2000.

[16] R. Caceres, N. Duffield, J. Horowitz, D. Towsley, and T. Bu, "Multicast-based inference of network-internal characteristics: Accuracy of packet loss estimation," in *Proc. IEEE infocom*, New York, Mar. 1999.

[17] F. Lo Presti, N. Duffield, J. Horowitz, and D. Towsley, "Multicast-based inference of network-internal delay distributions," TR 99-55, UMass Computer Science, Nov 1999.

[18] N. Duffield and F. Lo Presti, "Multicast inference of packet delay variance at interior network links," in *Proc. IEEE infocom*, Tel Aviv, Israel, Mar. 2000.

[19] S. Ratnasamy and S. McCanne, "Inference of multicast routing tree topologies and bottleneck bandwidths using end-to-end measurements," in *Proc. IEEE infocom*, New York, Mar. 1999.

[20] R. Caceres, N. Duffield, J. Horowitz, F. Lo Presti, and D. Towsley, "Loss-based inference of multicast network topology," in *Proc. IEEE Conf. Decision and Control*, Dec. 1999.

[21] N. Duffield, J. Horowitz, and F. Lo Presti, "Adaptive multicast topology inference," in *Proc. IEEE infocom*, Anchorage, Apr. 2001.

[22] A. G. Ziotopolous, A. O. Hero, and K. Wasserman, "Estimation of network link loss rates via chaining in multicast trees," in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Proc.*, Salt Lake City, May 2001.

[23] R. Caceres, N. Duffield, and T. Friedman, "Impromptu measurement infrastructures using RTP," in *Proc. IEEE infocom, to appear*, New York, June 2002.

[24] Y. Tsang, M. Coates, and R. Nowak, "Passive unicast network tomography using EM algorithms," in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Proc.*, Salt Lake City, May 2001.

[25] N. Duffield, F. Lo Presti, V. Paxson, and D. Towsley, "Inferring link loss using striped unicast probes," in *Proc. IEEE infocom*, Anchorage, Apr. 2001.

[26] M. Coates and R. Nowak, "Network delay distribution inference from end-to-end unicast measurement," in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Proc.*, Salt Lake City, May 2001.

[27] M. Coates and R. Nowak, "Sequential monte carlo inference of internal delays in non-stationary communication networks," *IEEE Trans. Signal Proc., to appear*, 2001.

[28] M. F. Shih and A. O. Hero, "Unicast inference of network link delay distribution from edge measurement," in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Proc.*, Salt Lake City, May 2001.

[29] K. Lai and M. Baker, "Measuring link bandwidths using a deterministic model of packet delay," in *Proc. ACM SIGCOMM*, Stockholm, Aug 2000.

[30] D. Rubenstein, J. Kurose, and D. Towsley, "Detecting shared congestion of flows via end-to-end measurement," in *Proc. ACM SIGMETRICS*, San Jose, June 2000.

[31] T. M Chen (Ed.), "Network traffic measurements and experiments," *IEEE Commun. Mag.*, May 2000.

[32] "The network simulator - ns-2," `http://www.isi.edu/nsnam/ns/`.

[33] A. Feldmann, A. G. Greenberg, C. Lund, N. Reingold, J. Rexford, and F. True, "Deriving traffic demands for operational IP networks: methodology and experience," in *Proc. ACM SIGCOMM*, Stockholm, Aug 2000, pp. 257–270.