

Model-based End-to-End Available Bandwidth Inference using Queueing Analysis

Xiaojun Hei[†], Brahim Bensaou[‡] and Danny H.K. Tsang[†]

[†]Department of Electrical & Electronic Engineering

[‡]Department of Computer Science

Hong Kong University of Science and Technology

Clear Water Bay, Kowloon, Hong Kong

heixj@ee.ust.hk, brahim@cs.ust.hk and eetsang@ee.ust.hk

Abstract

End-to-end available bandwidth estimation between Internet hosts is important to understand network congestion and enhance the performance of Quality-of-Service (QoS) demanding applications. In this paper, we investigate model-based available bandwidth measurement via the use of an active probing stream. A general end-to-end measurement framework, which unifies the current research approaches and highlights insights for measurement practice, is proposed. Within this framework, the end-to-end available bandwidth is inferred based on the measurement of the performance metrics of an active probing stream. We study two probing streams: Poisson and periodic probing. Of particular interest to our investigations is the Squared Coefficient of Variation (SCV) of the inter-probing packet arrival time at the receiver. The performance comparison of the available bandwidth measurements based on loss models and delay models indicates that the delay-based measurement exhibits many advantages over the loss-based measurements, such as accuracy, overhead and robustness. We conducted a comparison study between the proposed SCV-based probing scheme, namely, SCVProbe, and Pathload using ns-2 simulation in terms of probing accuracy, convergence time and overhead. Our evaluation results indicate that SCVProbe achieves similar or even better measurement accuracy than Pathload with much less probing time and smaller overhead.

Keywords: available bandwidth measurement, active measurement, packet train, queueing analysis, delay process

Model-based End-to-End Available Bandwidth Inference using Queueing Analysis

I. INTRODUCTION

Available bandwidth estimation has become an active research topic in the past several years due to a wide range of applications. End-to-end available bandwidth is related to but different from end-to-end capacity. The end-to-end capacity is the maximum throughput that a path can provide when there is no cross traffic, and the end-to-end available bandwidth is the maximum throughput that a path can provide at a certain time period for a given cross traffic. The available bandwidth is largely influenced by cross traffic along the path and the congestion control protocols used by cross traffic connections; therefore, it can be highly dynamic because the cross traffic has widely differing characteristics.

Various end-to-end available bandwidth estimation techniques have been proposed in the past few years. They are classified into two categories: self-congestion [1]–[4], and model-based [5]–[10]. Despite the good characteristics, such as accuracy, simplicity, speed, robustness, etc., claimed by proponents of the self-congestion approach, the congestion introduced by the probing stream inevitably changes the load along the path in such a way that the TCP flows along the path are sometimes interfered. Therefore, non-intrusive measurement can hardly be guaranteed. Conversely, in the model-based approach, the bandwidth measurement tool is operated when the path is not congested. This brings forth the possibility of a truly non-intrusive available bandwidth estimation technique.

In the end-to-end context, the available bandwidth along a path cannot be measured directly because the measurement entity is normally not able to access the network internals. One possible measurement approach is actively sending out packets into the network, and the performance metrics of the probing stream are measured; thereafter, the available bandwidth is inferred accordingly. The essence of active measurement is to identify an “effect” in order to establish the relationship between the *target measurement metric*, i.e., available bandwidth, and the *measurable performance metrics* of the active probing stream. With a well established relationship, available bandwidth is inferred from the performance measurement of this active probing stream. Both self-congestion and model-based measurement approaches fall into this big picture. We present an end-to-end measurement framework to unify the current research approaches and highlight some insights for the measurement practice. This framework helps to design more efficient probing methodologies and to enhance the understanding of the network performance.

In the model-based available bandwidth measurement, the performance metrics of the active probing stream fall into two categories: loss-related metrics [6], [7] and delay-related metrics [8]–[10]. We compare the measurement performance of the loss-based models and the delay-based models for available bandwidth measurement using queueing analysis and extensive simulations. The comparison indicates that the delay-based measurement exhibits more advantages over the loss-based measurement, such as accuracy, overhead and robustness. The impact of various parameters in both loss-based and delay-based models are investigated. The buffer size along the path shows a significant impact on the loss-based measurement. When the buffer size is small, the loss measurements appear to fluctuate significantly; when the buffer size is large, the loss is hardly observable and, hence, no valid available bandwidth estimation can be obtained. In contrast, the delay-based measurement, (in particular, the SCV of the inter-probing packet arrival time at the receiver), is sensitive to the load of the cross traffic. Therefore, it provides a more advantageous approach in available bandwidth estimation. We also conducted a comparison study between SCVProbe and Pathload using ns-2 simulation. Our evaluation results indicate that SCVProbe achieves similar or even better measurement accuracy than Pathload with much less probing time and smaller overhead.

The rest of the paper is organized as follows. First, in Section II some related work is outlined. Next, in Section III a general end-to-end measurement framework is presented. The available bandwidth measurement is considered as one application of this framework. In Section IV, we evaluate and compare the available bandwidth measurement using the analysis of the loss and delay processes of an active probing stream. In particular, the impact of the buffer size on the loss and delay processes of the probing stream are investigated using queueing models and extensive simulations. In Section V, the available bandwidth inference using delay-models are evaluated along multi-hop paths and SCVProbe is evaluated against Pathload using ns-2. Finally, concluding remarks are made in Section VI.

II. RELATED WORK

Pathload [1] and Initial Increasing Gap (IGI)/Packet Transmission Rate (PTR) [2] are two available bandwidth measurement methodologies in a self-congestion approach. By trying different probing rates using a binary search, a reasonable estimate of the available bandwidth can be found in Pathload. In [2] Hu and Steenkiste reported the construction of a delay-gap model to understand how cross traffic changes the probing packet gap for a single-hop path. A turning point exists at which the average input gap equals the average output gap. At this turning point, IGI utilizes a fluid-flow based formula to estimate the available

bandwidth and PTR estimates the available bandwidth on the bottleneck link as the average rate of the packet train. Essentially, Hu and Steenkiste studied the first-order statistics of the delay process of the probing stream, and their methodology works when the path is self-congested by the additional probing load. In [4], TOPP shared the same self-congestion nature as Pathload and IGI/PTR. The difference is that TOPP utilized the throughput measurement to identify a throughput transition point when the path load ρ is 1.

In [11], Strauss et al. improved the probe gap model of [2] by setting the inter-gap time between two probe pairs to be exponentially distributed. Their measurement tool, Spruce, shows better accuracy than IGI; however, Spruce essentially retains the self-congestion principle. In [3], Ribeiro et al. proposed to create a self-induced congestion using an active probing stream with an exponential flight pattern, called “pathChirp”. The available bandwidth estimation uses the information on whether the delay of pathChirp packets is increasing or decreasing, and a turning point is also identified when the probe sending rates and receiving rates start to match.

In [5] Sharma and Mazumdar proposed the estimation of the traffic parameters of the Poisson-type cross traffic by measuring the delay experienced by an active probing stream. However, their inference schemes require that the network nodes are cooperative and provide the local queueing information, such as the mean queue length and the waiting time of a probing stream. In the end-to-end scenario, the network cooperation cannot be guaranteed.

The model-based measurement method most closely related to our work is that of [6], in which Alouf et al. proposed that the network characteristics are inferred via moment-based estimators. The inference (queueing) models were developed for a single node case based on the single bottleneck assumption. Two models were studied, $M_1 + M_2/M/1/K$ and $M_1 + M_2/D/1/K$, to simultaneously estimate the cross traffic intensity and the buffer size of the bottleneck link given the knowledge of the bottleneck capacity. Multiple inference schemes were proposed based on the moment estimators of various performance metrics of the probing stream. Following an approach similar to that in [6], cross traffic characteristics were inferred based on the loss process analysis of an $MMPP/M/1/N$ queueing model in [7].

We believe that the proposed loss-based metrics used in [6] and [7] do not necessarily perform well in the measurement practice. For example, considering the end-to-end measurement situation, packet loss might not be easily observable during the measurement period. In addition, packets may be dropped due

to Active Buffer Management (AQM) instead of buffer overflow. A more observable metric, the inter-arrival time between consecutive packets at the receiver, has not yet been investigated in the queuing framework to study the available bandwidth measurement to the best of our knowledge. More importantly the variation of such inter-arrival time is of particular interest. This fact motivated us to conduct research on network inference via an analysis of the inter-arrival time of two consecutive probing packets at the receiver.

III. A GENERAL END-TO-END MEASUREMENT FRAMEWORK

A. Overview

In the end-to-end measurement context, we present a general measurement framework depicted in Figure 1. This framework consists of two sub-problems, the problem of modelling an end-to-end path and the problem of inferring the unknown parameters of this path model. In the modelling problem, the end-to-end path is abstractly modelled as a series of tandem queues with a superposition of two arrival processes, the probing stream and the cross traffic stream. This network model is determined by a set of determinable or controllable parameters, Θ , of the probing actions and a set of unknown parameters, X , which include the target measurement metrics, i.e., the available bandwidth.

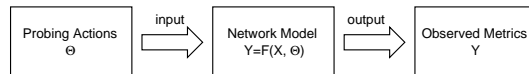


Fig. 1. A general end-to-end model-based measurement framework.

The output of the measurement process is the observed metrics Y . With an appropriate hypothetical model, the relationship between Y and (X, Θ) can be established, namely, $Y = F(X, \Theta)$. Therefore, the performance of an active probing stream from the sender to the receiver can be predicted given all the parameters of this hypothetical model. As an inverse process, in the inference problem, the parameters of this hypothetical model, that are of particular interests, including path capacity and available bandwidth, can be estimated by measuring the performance of this probing stream. In a mathematical form, $X = F^{-1}(Y, \Theta)$. With a different probing strategy, controlling the parameter set Θ of the input probing process, a different F can be identified. Relatively accurate tools for capacity measurement have been proposed in [12], [13]. Our focus in this paper is on the estimation of the cross traffic intensity under the assumption that the path capacity is known, which is provided for instance by the packet-pair technique. Given

knowledge of the capacity information, the available bandwidth is determined by the difference between the capacity and the cross traffic intensity.

B. The Modelling Problem

The accuracy of the model-based measurement largely depends on whether the hypothetical model is able to capture the relationship F correctly. The construction of the measurement model involves three components: the design of the probing sequence (determining an appropriate parameter set Θ), the characteristics of the cross traffic stream, and how packets are transmitted in the network.

A commonly used assumption in the active measurement literature is the existence of a single bottleneck along an end-to-end path, in that, the performance of the probing stream is dominated by the most congested link along the path [2], [6]–[8]. Despite the simplicity, the analysis of this single bottleneck leads to valuable insights into understanding the behavior of the complex network. As a simple starting point, an end-to-end Internet path is modelled as a single server queue with two concurrent streams, the probing stream and the cross traffic, as shown in Figure 2.

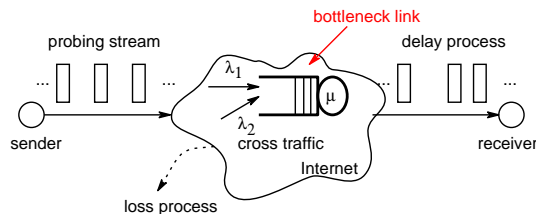


Fig. 2. The available bandwidth inference model.

1) *Probe Sequence Design*: A common measurement consists of multiple probe sequences. Probe sequence design sets the characteristics of both the distribution of packets sequences and the distribution of packets within a sequence [14]. A probe sequence consists of a number of sequential probing packets. The goal of this probe sequence design is to control the parameter set Θ for establishing a most inferrable $X = F^{-1}(Y, \Theta)$. Depicted in Figure 3, these parameters include the number of sequences (N), the number of probing packets within a sequence (M), the probing packet size (L), the inter-arrival time between probing packets ($1/\lambda_1$, λ_1 is the packet arrival rate), the inter-arrival time between probing sequences ($1/\gamma$, γ is the sequence arrival rate).

Probing sequences with three types of inter-arrival time distributions are commonly used: a Poisson probe sequence [15], a periodic probe sequence [16], and an exponential probe sequence [3]. Each probe

has applications in the measurement practice. A Poisson probe sequence can provide unbiased near-continuous detection while a periodic probe sequence is easier to implement without having to concern about the stochastic robustness of a Poisson sequence. Nevertheless, mathematically it is more difficult to analyze the departure process of a periodic stream in a series of tandem queues. An exponential probe sequence was recently proposed to capture the long-term dependence of the Internet traffic; yet, it lacks an accurate analysis for understanding its stochastic behaviors [17]. In our work, we focus on the Poisson and periodic probing sequences.

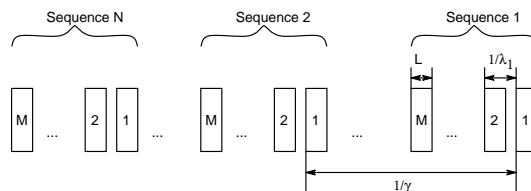


Fig. 3. Probe sequences in one measurement session.

2) *Characteristics of the Cross Traffic:* The Internet traffic is a superposition of many flows, based on many protocols. Tremendous research effort has been made both in regard to theory and measurement so as to present an understanding of Internet traffic. Traffic on the Internet has been reported, in numerous studies, to be self-similar, [18], [19]; in large networks, however, it has been found [20] that the overall arrival rate at nodes within the core of the network is Poisson.

Knowledge of the packet distributions on the Internet provides a guideline in understanding the service time distribution. Several works on the measurement of packet size distributions of the Internet have been published [19], [21], [22], and all show very common features. It is not possible, in practice, to obtain the complete packet size distribution; however, with tremendous effort in Internet measurement [19], [21], [22], it is fair to assume that the first and second moments of the packet size distribution of the Internet traffic can be determined. Then, the first moment matching or two moment matching schemes can be applied [8].

In what follows, we illustrate a simple first order mapping scheme for the packet size distribution using a negative exponential distribution. Based on the Internet measurement, it is assumed that the Internet traffic consists of three types of packet sizes, $\{40, 550, 1500\}$ bytes, with a distribution of $\{40\%, 50\%, 10\%\}$. Because there is only one parameter in an exponential distribution, we map the first moment so that the average packet size (441 bytes) in the mapped negative exponential distribution is the same as that

in the above typical Internet packet distribution. We used the negative exponential distribution assumption of packet sizes in our simulation study. However, our probing method is not limited to exponentially distributed packet sizes. In [8], we proposed mapping the first and second moments of the packet size distribution into a generalized distribution. This two-moment mapping scheme achieved a better accuracy than the first moment mapping scheme.

3) *Knowledge about the Network Services:* It is also a prerequisite to understand how packets are transmitted in the network, such as the queueing discipline at the routers. One common assumption used in current measurement work is the First-In-First-Out (FIFO) queueing scheme deployed in Internet routers. It is still largely true for most current Internet routers; however, various Active Queue Management (AQM) and different packet scheduling algorithms, i.e., Weighted Fair Queueing (WFQ), have started to be deployed in routers. In the loss-based model [6], [7], the reason for the packet dropping is assumed to be due to buffer overflow only. With the application of AQM, more packet dropping is possibly observed. This indicates the potential inaccuracy of the loss-based available bandwidth estimation by overestimating the packet loss probability and, hence, underestimates the available bandwidth.

4) *Model Construction:* The system illustrated in Figure 2 is modelled as a single server, a finite capacity queue operating in continuous time, which accepts two classes of customer arrivals, the probing packet arrivals and the cross traffic packet arrivals. The probing packet arrival process is assumed to be independent and probing packets are referred to as customers of class 1. The arrival rate of the probing stream is denoted as λ_1 and the service rate is μ_1 . The cross-traffic packets are referred to as customers of class 2, and the arrival rate of cross traffic is denoted λ_2 and the service rate is as μ_2 . Assume that the queueing system is in a steady state; therefore, the traffic intensity $\rho = \rho_1 + \rho_2 = \frac{\lambda_1}{\mu_1} + \frac{\lambda_2}{\mu_2} < 1$. The load of the cross traffic is also denoted as $\rho_{CT} = \rho_2$. Denote the buffer size of K packets ($K \geq 1$) including the packet in service.

C. The Inference Problem

1) *Metric Selection:* The inference problem involves an appropriate selection of an observable metric set Y and the resolution of the inverse of F . A “good” metric ($y \in Y$) ought to be observable in the sense that it is sensitive to X , and, yet, the inference procedure $F^{-1}(\hat{Y}, \Theta)$ leads to a stable estimated \hat{X} , being robust to measurement noises.

The IPPM working group of the IETF defined some end-to-end performance metrics [15], including

delay (RFC2679 and RFC2681), delay variation (RFC3393), throughput (RFC3148) and loss (RFC2680 and RFC3357). These candidate metrics are delay-related or loss-related. The throughput-based measurement has turned out to be too intrusive because it injects too much traffic into the network, and it can also be considered as a loss-related metric because the transmitted traffic is the total input traffic minus the traffic dropped in the network.

The available bandwidth measurement, based on the analysis of the loss process of the probing stream, was discussed in [6] and [7]. The work in [1], [2] and [8] rely on the measurement of the delay performance of the probing stream. To the best of our knowledge, the model-based available bandwidth measurement using delay-variation has not been investigated. Of particular interest in our investigations, is the SCV of the inter-departure time between two consecutive probing packets. In a real measurement system, given the measured SCV of the probing stream, inverting the function F helps to infer the load of the cross traffic on the end-to-end path between two end nodes.

2) *Inference Procedures:* With the assumption that cross traffic is a simple Poisson arrival process and only a single bottleneck exists, the active Poisson-type probing problem can be formulated into a queueing analysis problem in regard to the performance metrics of the probing stream. These assumptions lead to a tractable mathematical treatment; however, it is also important to examine the robustness of the proposed model when these assumptions do not hold.

In the loss-based model, the loss process of the probing stream was analyzed using $M_1 + M_2/M/1/K$ and $M_1 + M_2/D/1/K$ in [6]. In the delay-based model, the departure process (delay process) of the probing stream can be analyzed using an $M_1 + M_2/GI_i/1$ queueing system [23] to capture the impact of cross traffic on a Poisson probing stream. A heavy traffic approximation approach [24] can also be employed to analyze the delay process of the probing stream in the above queueing system.

IV. MODEL-BASED AVAILABLE BANDWIDTH INFERENCE: LOSS-MODELS VS. DELAY-MODELS

In Section IV-B, the end-to-end available bandwidth is inferred based on the measurement of the loss process and we highlight the ineffectiveness of the loss-based models. In Section IV-C, we present a delay-gap model using queueing analysis and evaluate the impact of the buffer size on the delay-gap model using simulations. In Section IV-D, the queueing analysis and simulation results show that the SCV-based model provides a promising light-weight probing methodology in the end-to-end available bandwidth inference.

A. Simulation Configurations

The network model used in this section is shown in Figure 4. It is a single-hop path. (In spite of the simplicity, the analysis and simulation can still illustrate the insights. We will present the experiment results in a more general multi-hop topology in later sections.) The link is modelled using a single-server drop-tail queue with a processing rate equal to the link bandwidth. This queue is assumed to deploy the FIFO scheduling scheme. The maximum queue size is equal to the buffer size of the router. This link has a bandwidth capacity of 2 Mbps and the buffer size is 500 Kbytes. There are 10 homogeneous cross traffic sources, and the packet size of the cross traffic follows a negative exponential distribution with a mean value of 441 bytes. For each set of simulation points for the evaluation of the loss models, there are 11 batches in one simulation run and the length of each batch is 10^5 probing packets. The first batch of probing packets are not used to collect statistics in order to avoid the effect of the transient states. The latter 10 batches simulation results are used to calculate the 95% confidence intervals, indicated by the error bars in the figures. The delay information is much easier to measure compared to the loss measurement and, hence, the length of each batch is set to 10^4 probing packets for the evaluation of the delay model while the accuracy and robustness of the measurement are still good. The whole simulation model was constructed using SimLib 2.2 [25].

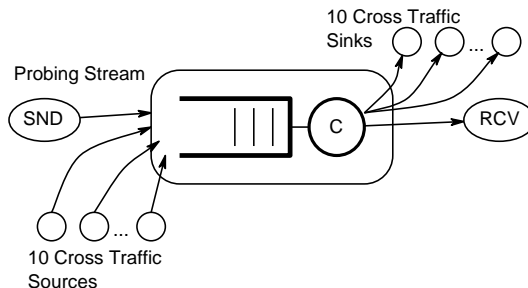


Fig. 4. A single hop path.

B. Loss Model using $M_1 + M_2/M/1/K$

In [6], the network characteristics were inferred via moment-based estimators. The inference (queueing) models were developed for a single node case based on the single bottleneck assumption. Two models were used, $M + M/M/1/K$ and $M + M/D/1/K$, to simultaneously estimate the cross traffic intensity and the buffer size of the bottleneck link given the bottleneck capacity. Multiple inference schemes were proposed based on the moment estimators of the packet loss probability (P_L), server utilization (U), the

expected response time (R), the conditional loss probability (q_L) (the probability that two consecutive losses occur), and the conditional non-loss probability (q_N) (the probability that a probing packet arrives to find room in the buffer given that the previous probing packet was also admitted into the queue). The best scheme was shown to be the inference via the estimators of the packet loss and the expected response time using the simulation study. (A brief summary of the equations used in the related inference schemes can be found in Appendix I.)

Considering the end-to-end measurement situation, the packet loss is not easily measured during the measurement period (there is, possibly, no loss at all) and the slow convergence is within expectation; the server utilization cannot be measured without the cooperation of the network; due to the possible clock drifting between the sender and the receiver, the estimation of the expected response time of the probing packets is, possibly, inaccurate based on the arriving time instant to the queueing system at the sender side and the departure time instant out of the queueing system at the receiver side. Since the packet loss probability, the conditional loss probability, and the conditional non-loss probability are the only three “good” candidate metrics in the end-to-end measurement scenario, the schemes of $P_L - q_L$, $P_L - q_N$, $q_L - q_N$ are evaluated in the later part of this section.

In $M_1 + M_2/M/1/K$, the probing stream is generated with the arrival rate of λ_1 and the cross traffic is generated with the arrival rate of λ_2 . The average service rates of two streams are $\mu_1 = \mu_2 = \mu$. This indicates that the configuration of the probing stream ought to ensure that the probing packet size is the average packet size of the cross traffic packets. The aggregate traffic intensity is $\rho = \frac{\lambda_1 + \lambda_2}{\mu}$. In the remaining of this section, we discuss the impact of the buffer size using the $M_1 + M_2/M/1/K$ model as well as simulations.

Figure 5 shows the relationship between the loss probability P_L and the normalized probing load¹ ρ_1 for different buffer sizes. Only when the probing stream congests the path ($\rho \rightarrow 1$), loss events start to be observable in practice (e.g., $P_L \leq 0.01$). For measurement purposes, it is important to really detect the loss events in order to provide information for the available bandwidth inference. To obtain a loss measurement of $P_L = 0.01$, at least 100 probing packets have to be sent out to detect 1 packet loss. For a practical statistically-reliable measurement of $P_L = 0.01$, 10 times (or even higher) the required number of the packets should be sent out (i.e., at least 1000 probing packets). This implies a heavy measurement

¹This normalized probing load is $\rho_1 = \text{probe_arrival_rate} * \text{probing_packet_size} / \text{bottleneck_link_capacity}$.

overhead and long measurement intervals.

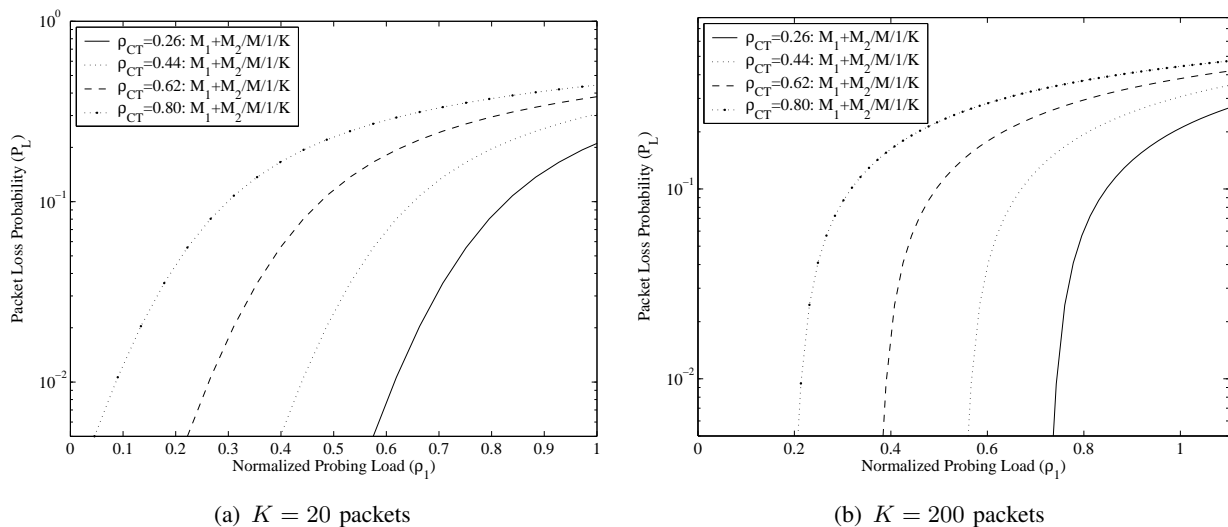


Fig. 5. The relationship between the loss probability P_L and the Poisson probing load $\rho_1 = \frac{\lambda_1}{\mu}$ with different buffer sizes (K) using (5).

In Figure 6, we can observe that the measurement operating point ensures an aggregate load approaching to 1 when the buffer size increases in order to have a practical measurable $P_L = 0.01$. When $K = 10$ packets, the aggregate load (ρ) is larger than 0.71 for the measurement point $P_L = 0.01$. When $K = 80$ packets, the required aggregate load (ρ) increases very quickly to 0.99 for detecting $P_L = 0.01$. This indicates that even if the cross traffic load is small, i.e., $\rho_{CT} = 0.1$, the probing load has to be as high as $\rho_1 = 0.89$ so that a loss probability $P_L = 0.01$ is practically measurable when $K = 80$ packets.

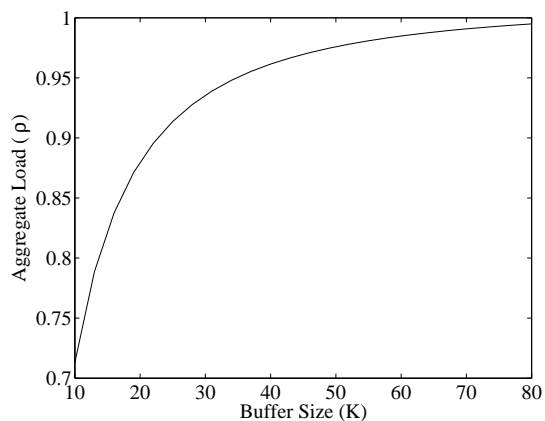


Fig. 6. The sensitivity of the loss probability when $P_L = 0.01$ on the Poisson probing load ρ_1 with Poisson-type cross traffic and different buffer sizes (K) using (5).

Figure 7 shows the relationship between the conditional loss probability q_L and the normalized probing load ρ_1 for different buffer sizes. The q_L curves appear to be insensitive to the buffer size. Since a packet

drop is a rare event when $\rho < 1$, the conditional loss probability is even harder to obtain in the actual measurement. This was verified by the $M_1 + M_2/M/1/K$ simulation results, which are presented in Figure 8. In Figure 8(a) ($K = 20$), the measured conditional loss probability has a large variation indicated by a large confidence interval when $\rho_{CT} = 0.26$ and the probing load ρ_1 is small. As shown in Figure 8(b), the situation is even worse. Besides the large measurement variation, there are practically no conditional loss events recorded during the simulation when the probing load is small in four different cross traffic load scenarios. This implies that measuring the conditional loss probability for the available bandwidth inference does not provide a reliable result.

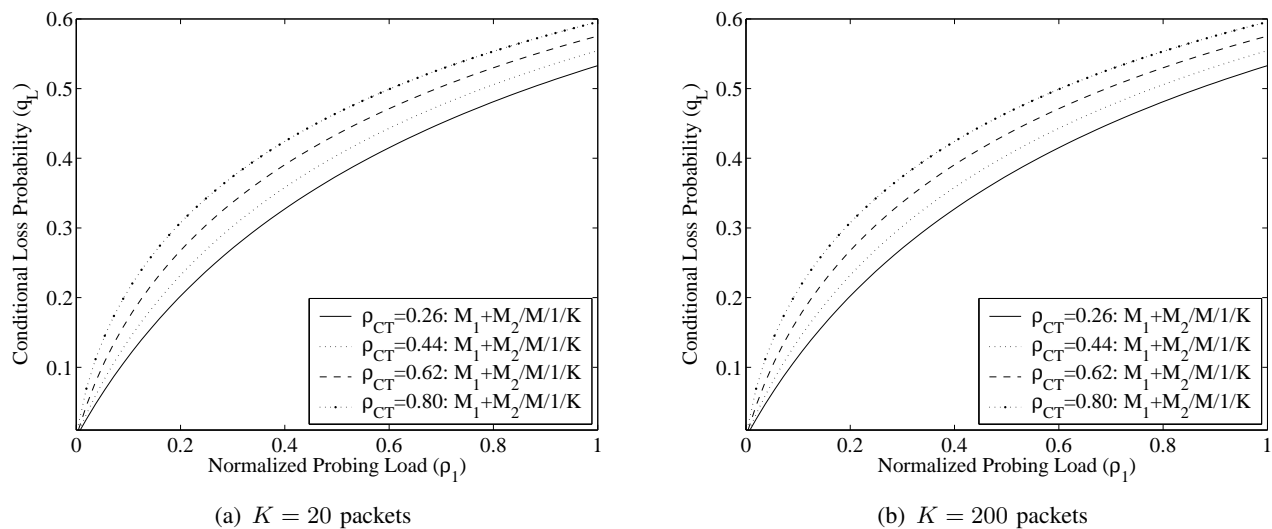


Fig. 7. The relationship between the conditional loss probability q_L and the Poisson probing load ρ_1 with Poisson-type cross traffic and different buffer sizes (K) using (7).

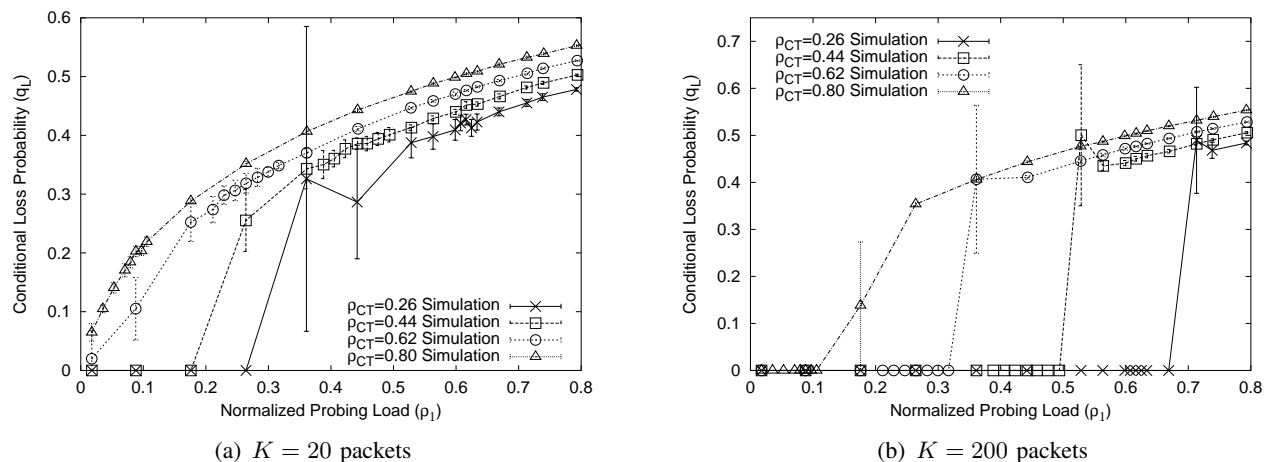


Fig. 8. The relationship between the conditional loss probability q_L and the Poisson probing load ρ_1 with Poisson-type cross traffic and different buffer sizes (K) in simulation.

From the above discussions, all of the $P_L - q_L$, $P_L - q_N$, $q_L - q_N$ inference schemes have practical difficulties when the probing load is small. To ensure a more observable loss measurement, Salamatian, Baynat and Bugnazet [7] proposed to further increase the probing load so that the aggregate load $\rho \geq 1$. However, this negates the advantage of the model-based measurement without congesting the network.

C. Delay-Gap Model using $GI_1 + GI_2/GI_i/1/\infty$

The analysis of the delay process of the probing stream in $GI_1 + GI_2/GI_i/1/K$, depicted in Figure 2, is difficult because packet dropping may destroy the delay pattern. Under the assumption that the buffer size is infinite, the original queue becomes $GI_1 + GI_2/GI_i/1/\infty$. In this section, we will also discuss the impact of packet dropping on the delay pattern of the probing stream when the buffer size is finite and the cross traffic is modelled as an aggregate Poisson process.

Figure 9 depicts the time diagram of the delay process in $GI_1 + GI_2/GI_i/1/\infty$. Denote the n -th probing packet arrival as C_n of the class-1 customer arrival (the probing stream). Its service time is X_n . Let I_n denote the inter-arrival time (i.a.t.) between the n -th and $(n + 1)$ -st probing packets and D_n the corresponding inter-departure time (i.d.t.). N_y denotes the number of arrivals from the cross traffic stream which occur during the inter-arrival time C_n and C_{n+1} . Z_n is the waiting time of C_n . Y_j is the service time for the j -th packet from the cross traffic stream between the arrival of C_n and C_{n+1} . When probing packets traverse the network along a path in the multi-class queue as depicted in Figure 4, the departure process of the probing stream is determined by its arrival process and the disturbance from the cross traffic and the network service process.

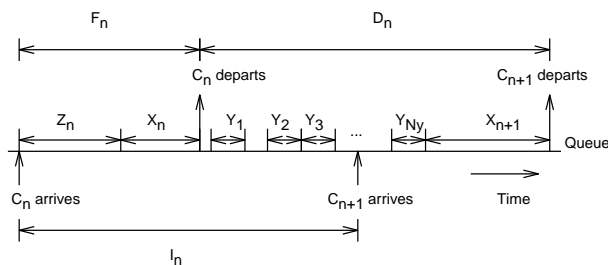


Fig. 9. The time diagram of the delay process in $GI_1 + GI_2/GI_i/1/\infty$.

The delay gap ratio is defined as the ratio between the average inter-departure time D_n and the average inter-arrival time I_n of the probing packets. This ratio essentially provides first-order delay information of the delay process.

With the FIFO scheduling algorithm in a queue with an infinite size, $E[N_y] = \lambda_2/\lambda_1$, $E[I_n] = 1/\lambda_1$. When $\rho < 1$, for a queue with an infinite buffer, the waiting time of probing packets is finite. These packets eventually leave the queue after their service. Due to the flow conservation law, the departure rate of the probing stream equals its arrival rate. Note that the inverse of the packet arrival rate is $E[I_n]$ and the inverse of the packet departure rate is $E[D_n]$; therefore, $E[D_n] = E[I_n]$. Thus,

$$GapRatio = \frac{E[D_n]}{E[I_n]} = 1.$$

If $\rho \geq 1$, the server is always busy; therefore,

$$D_n = \sum_{j=1}^{N_y} Y_j + X_{n+1}.$$

N_y is the number of packet arrivals from the cross traffic stream during the inter-arrival time I_n . Since $\rho_{CT} < 1$; therefore, $E[N_y] < +\infty$. All these random variables are assumed to be independent, and $E[Y_j] < +\infty$ and $E[X_{n+1}] < +\infty$; thus

$$E[D_n] = E[N_y]E[Y_j] + E[X_{n+1}] < +\infty.$$

By considering the $GI_1 + GI_2/GI_i/1/\infty$ queueing system, the above equations can be simplified as follows: $E[X_{n+1}] = \frac{1}{\mu_1}$, $E[Y_j] = \frac{1}{\mu_2}$, $E[D_n] = E[N_y]E[Y_j] + E[X_{n+1}] = \frac{\lambda_2}{\lambda_1\mu_2} + \frac{1}{\mu_1}$; hence,

$$GapRatio = \frac{E[D_n]}{E[I_n]} = \frac{\lambda_1}{\mu_1} + \frac{\lambda_2}{\mu_2} = \rho_{probe} + \rho_{CT}.$$

Therefore,

$$GapRatio = \begin{cases} 1, \rho < 1 \\ \rho_{probe} + \rho_{CT}, \rho \geq 1 \end{cases} \quad (1)$$

Note that the above independence assumptions are fair. N_y and Y_j are the parameters of the cross traffic. X_{n+1} is the service time of probing packet $n + 1$. The cross traffic stream and the probing stream are assumed to be two independent streams. Hence, N_y , Y_j and X_{n+1} are independent.

Intuitively, Equation 1 deals with the first order delay information of the arrival process and the departure process of the probing sequence. The cross traffic burstiness does not impact the measurement of the average input gap and the average output gap for an infinitely-long probing sequence. However, in practice, we cannot send infinitely-long probing sequences. The impact of the traffic burstiness on the accuracy of a finite-length probing sequence should be investigated carefully. Nevertheless, according to our simulation study, when the probing sequence consists of more than 4000 probing packets, the measured SCV of the

probing sequence converges to the theoretical value. With a few hundred probing packets, the average gap of the probing sequence also converges to the theoretical value.

From (1), there exists a transition point when the aggregate load $\rho = 1$. When $\rho < 1$, the average inter-departure time ($E[D_n]$) equals the average inter-arrival time ($E[I_n]$) of the probing packets. In [2], the IGI/PTR algorithms essentially utilize this property of the delay process of the probing stream to identify this transition point. Note that the identification of this transition point requires that the probing rate has to be large enough to congest the network ($\rho \geq 1$) for a period of time. The existence of this gap transition point also explains the basic principle used in Pathload [1]. When the delay gap is larger than 1, it equivalently indicates that the one way delay of the probing packets is exhibiting a strong increasing trend. When the delay gap equals 1, this one way delay increasing trend is practically not observable.

The delay gap in (1) is obtained under the assumption of infinite buffer size. In practice, probing packets are prone to being dropped because of buffer overflow or active buffer management. This packet dropping can destroy the structure of the delay gap pattern. In case of packet dropping, only a delay gap of two consecutive probing packets are collected for statistical computation, namely, a conditional gap ratio. To this end, each probing packet is assigned a unique sequence number when it is sent out. This sequence number increases by one whenever a new probing packet is sent out. By checking the sequence number of received packets, the receiver extracts those consecutive probing packets for data collection. In Figure 10 are plotted this conditional gap ratio for two different buffer size (K) with increasing probing load. In Figure 10, when the buffer size is finite, the gap ratio is not increasing linearly with the increasing probing load as predicted using (1). When the buffer size is small ($K = 20$ packets), the packet dropping is more severe and the location of the transition point is shifted. When the buffer size becomes larger ($K = 200$ packets), this delay gap transition point is preserved to some extent. In both cases, in order to ensure the identification of this transition point, the overall load should be $\rho > 1$, at least temporarily. This partially explains why Pathload and IGI tend to overestimate the available bandwidth [11]. Another reason might be that this transition point exists at $\rho = 1$, the link is already congested and packet dropping occurs; hence, those elastic TCP flows might detect packet dropping and slow down their transmission rate. As a result, the probing stream grabs more available bandwidth.

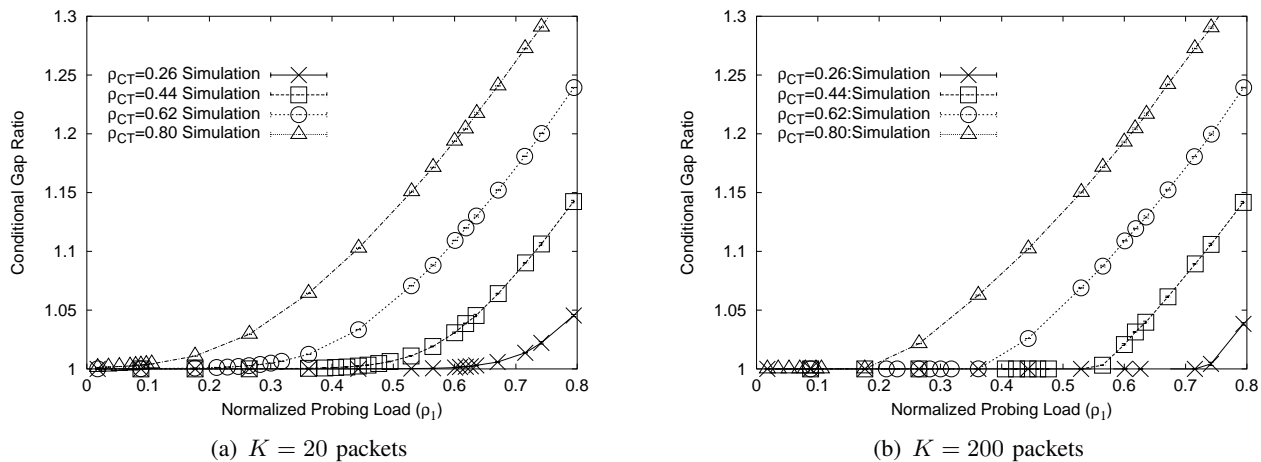


Fig. 10. The relationship between the conditional gap ratio and the Poisson probing load ρ_1 with Poisson-type cross traffic and different buffer sizes (K) in simulation.

D. SCV-based Model using $GI_1 + GI_2/GI_i/1/\infty$

1) *Overview*: The delay variation is defined as $J_n = D_n - I_n$. Note that there are possible negative jitters as well as positive jitters. Negative variation corresponds to a clustering of packets. This may result in buffer overflow; a positive variation corresponds to a dispersion of the packets and this may result in excessive delay. Because I_n is known in advance (the customers of class 1 are probing packets under the control), the task of characterizing J_n is reduced to find the property of D_n . For different probing streams, the queuing models can be constructed correspondingly. When the path load is less than 1, the average output gap equals the average input gap regardless of the types of arrival processes of the cross traffic and traffic intensity. Therefore, this first order characteristics of the delay process cannot provide information about the exact load of the cross traffic when $\rho < 1$. Therefore, we turned our investigation to the second order characteristics of the delay process of the probing stream. The SCV of the inter-probing packet arrival time at the receiver is defined in (2). The SCV is a dimensionless metric used to describe the delay variation of the probing stream. Intuitively, SCV_{D_n} is a normalized version of the variance of D_n with respect to I_n because $E[D_n] = E[I_n]$ when $\rho < 1$. It shows how the departure process of the probing stream is disturbed with respect to its arrival process due to the impact of the cross traffic. The absolute variance of D_n cannot offer an accurate information about the load of the cross traffic; only the relative variance of D_n with respect to I_n can serve our purpose of estimating this load.

$$SCV_{D_n} = \frac{Var[D_n]}{E^2[D_n]} \quad (2)$$

2) *Poisson Probing*: For a Poisson probing stream and a Poisson cross traffic stream, the $GI_1 + GI_2/GI_i/1/\infty$ queue becomes $M_1 + M_2/GI_i/1/\infty$. SCV_{D_n} in $M_1 + M_2/GI_i/1/\infty$ is analytically solvable using the procedure reported in [23] (see Appendix II). SCV_{D_n} is influenced by various factors, including the variance of the probing stream, the variance of the packet size of the cross traffic, and the load of the cross traffic. The first two are determinable parameters and the last one is the targeted measurement metric. The relationship between SCV_{D_n} and the normalized probing load ρ_1 with Poisson-type cross traffic and different probing packet sizes using $M_1 + M_2/GI_i/1$, is shown in Figure 11. It is different from the delay gap ratio in that, when $\rho < 1$, SCV_{D_n} is significantly influenced by the load of the cross traffic, and this effect can serve as the basis for the inference of available bandwidth.

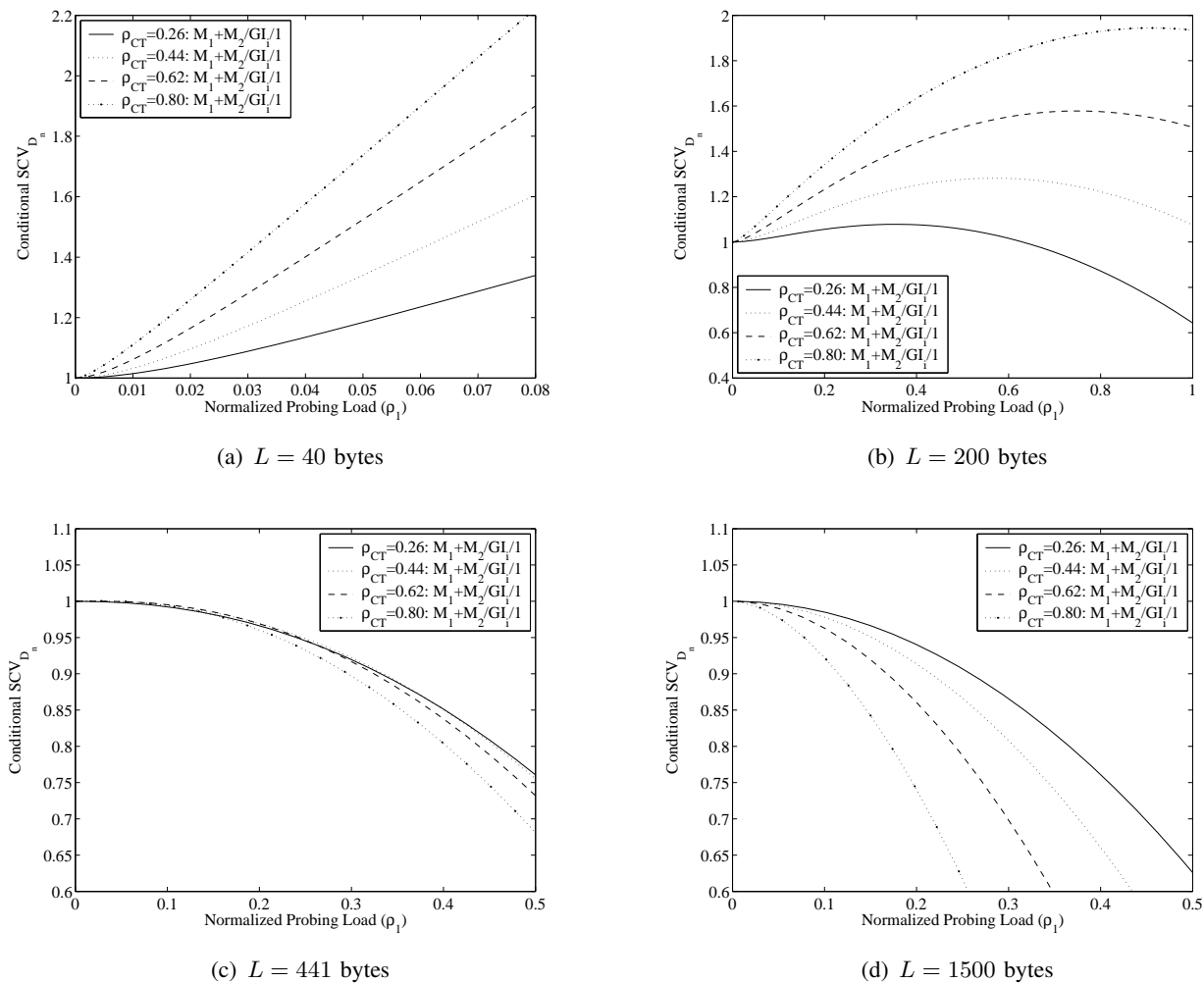


Fig. 11. The relationship between SCV_{D_n} and the Poisson probing load ρ_1 with Poisson-type cross traffic and different probing packet sizes (L) using $M_1 + M_2/GI_i/1/\infty$ ($GI_1 = D$ and $GI_2 = M$).

Figure 11 depicts the relation between the load of the cross traffic and the conditional SCV_{D_n} . We

solved the $M_1 + M_2/GI_i/1$ queueing model using a numerical method, because there is no closed-form solution for $M_1 + M_2/GI_i/1$. The plots in Figure 11 are obtained using an iterative algorithm. Based on these plots, we can observe that the SCV is much more differentiable with respect to the load of cross traffic when the probing packet size is small (e.g., 40 bytes) than when the probing packet size is large (441 bytes, 1500 bytes). Of particular interest is that SCV_{D_n} exhibits an almost linear relationship to the probing load when the probing packet size is $L = 40$ bytes. This promising result indicates that a simple relationship can be established between the available bandwidth and SCV_{D_n} . When the probing packet size ($L = 441$ bytes) equals the average packet size of the cross traffic, SCV_{D_n} is not sensitive to the load of the cross traffic. This implies that the probing packet size should be much smaller than the average packet size of the cross traffic. In our probing method, we recommend to use the probing packet size of 40 bytes to ensure an invertible function so as to infer the load of cross traffic. Sending small probing packets also minimizes the probing load and 40 bytes is the smallest IP packet size when TCP/UDP protocols are used.

As presented in the previous sections, packet dropping may destroy the pattern of SCV_{D_n} . We are also interested in the conditional SCV_{D_n} in that only D_n of consecutive probing packets are collected for the computation of SCV_{D_n} . Figure 12 depicts the second-order property of the delay process of the probing stream obtained by simulation with the different types of arrival processes of the cross traffic and traffic intensities. The SCV proves to increase when the normalized probing load increases. For different types of the arrival processes of the cross traffic (including Pareto, Hyper-Exponential, Poisson, Erlang), the probing stream has a similar SCV, especially when the cross traffic is lightly loaded. When the load of the cross traffic increases, the impact of the higher order statistics of the arrival process starts to occur; however, the effect is still small. Therefore, the analysis of the per-class departure process for the Poisson-type cross traffic can be used to infer the non-Poisson type cross traffic without much loss on accuracy. In the simulation experiments of a multi-hop path in Section V, the arrival process of the cross traffic is Pareto-type with a shape value of $\alpha = 1.9$, the inference accuracy is still quite good even though the model assumes Poisson-type cross traffic. What we conjecture is that the SCV of the probing stream is dominated by the low frequency components of the cross traffic. This may be explained by the queueing analysis in [26] and [27].

Figure 13 shows the relationship between the conditional SCV_{D_n} and the normalized probing load ρ_1

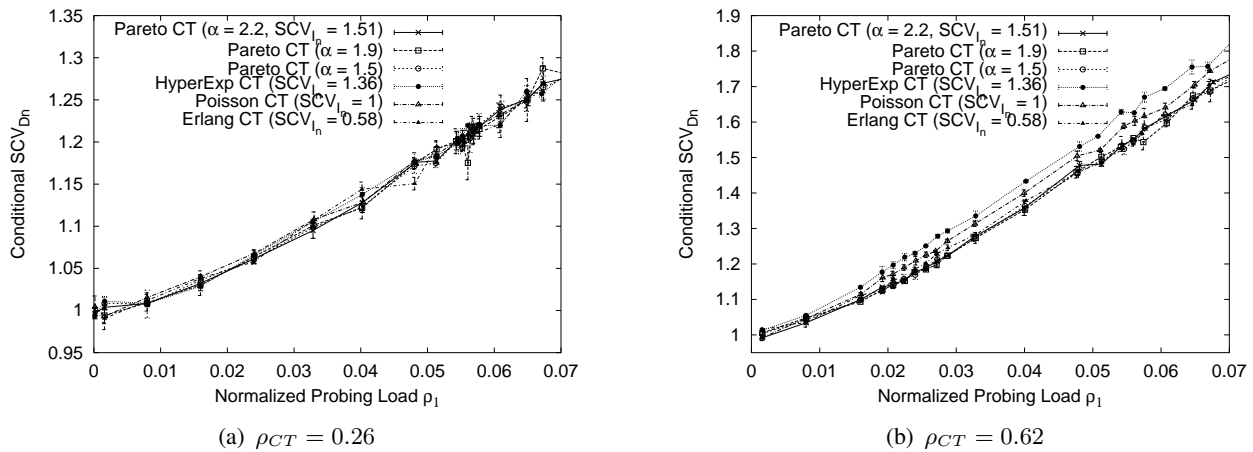


Fig. 12. Relationship between the conditional SCV_{D_n} and the normalized probing load with various type cross traffic arrival processes and the probing packet size is 40 bytes.

for different buffer sizes (K). We can observe that the impact of the buffer size on SCV_{D_n} is very small in both cases. This is due to the deployment of a small probing packet size ($L = 40$ bytes), and the perturbation of the probing stream on the network is quite small; yet, SCV_{D_n} is quite differentiable in various load situations. In addition, the model and simulation results match very well. This provides a firm basis for accurate available bandwidth inference.

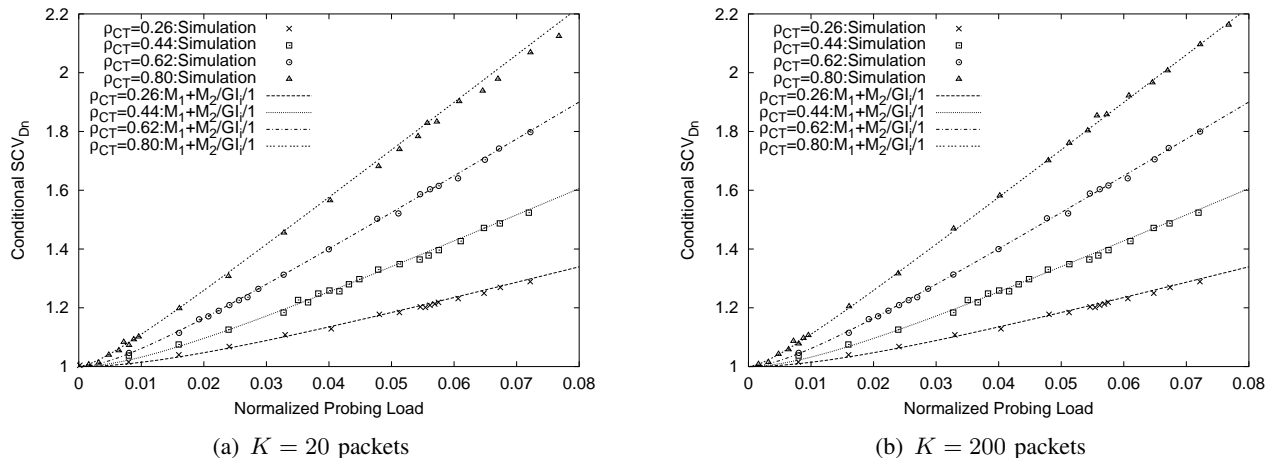


Fig. 13. The relationship between the conditional SCV_{D_n} and the Poisson probing load ρ_1 with Poisson-type cross traffic and different buffer sizes (K) in simulation.

3) *Periodic Probing*: For a periodic probing stream and a Poisson cross traffic stream, the $GI_1 + GI_2/GI_i/1/\infty$ queue becomes $D + M/GI_i/1/\infty$. An exact calculation of SCV_{D_n} is difficult; hence, we proposed a hybrid approximation for calculating SCV_{D_n} for periodic probing.

As in [28] and also shown in (3), for a two-class queueing model of $GI_1 + GI_2/GI_i/1/\infty$,

$$C_1^D = \rho_1^2 C_1^S + \rho_2^2 \left(\frac{p_1}{p_2}\right) (C_2^S + C_2^A) + (1 - \rho_1 \rho + \rho_1^2) C_1^A, \quad (3)$$

where C_i^D is the SCV of the departure process for class i , C_i^S is the SCV of the service process for class i , C_i^A is the SCV of the arrival process for class i , ρ_i is the traffic intensity for class i and p_i is the probability that an arrival is from class i . Equation (3) is an approximation which can be solved by simple iterative algorithms. Nevertheless, for our measurement practice, numerical solution still serves our purpose.

Replace the periodic probing stream with a Poisson probing stream and maintain the other parameters unchanged. The $D + M/GI_i/1/\infty$ queue becomes an $M_1 + M_2/GI_i/1/\infty$ queue, which can be exactly analyzed in [23]. Our idea of analyzing the departure process of the periodic stream in $D + M/GI_i/1/\infty$ is to identify the relationship between these two queueing systems using (3). We propose an approximate formula for calculating the SCV_{D_n} for a periodic stream as in (4), namely, “hybrid approximation”.

For a periodic stream with a constant probing packet size, $C_1^S = 0$ and $C_1^A = 0$.

$$\begin{aligned} SCV_{D_n}^{Periodic} &= \rho_1^2 C_1^S + \rho_2^2 \left(\frac{p_1}{p_2}\right) (C_2^S + C_2^A) \\ &\quad + (1 - \rho_1 \rho + \rho_1^2) C_1^A \\ &= \rho_2^2 \left(\frac{p_1}{p_2}\right) (C_2^S + C_2^A). \end{aligned}$$

For a Poisson stream with a constant probing packet size, $C_1^S = 0$ and $C_1^A = 1$.

$$\begin{aligned} SCV_{D_n}^{Poisson} &= \rho_1^2 C_1^S + \rho_2^2 \left(\frac{p_1}{p_2}\right) (C_2^S + C_2^A) \\ &\quad + (1 - \rho_1 \rho + \rho_1^2) C_1^A \\ &= \rho_2^2 \left(\frac{p_1}{p_2}\right) (C_2^S + C_2^A) + (1 - \rho_1 \rho + \rho_1^2). \end{aligned}$$

Therefore,

$$SCV_{D_n}^{Periodic} = SCV_{D_n}^{Poisson} - (1 - \rho_1 \rho + \rho_1^2) \quad (4)$$

As shown in Figure 14, the SCV_{D_n} of the periodic probing stream is sensitive to the load of the cross traffic; however, for the Poisson-type and Pareto-type cross traffic, the periodic probing stream has a similar SCV_{D_n} . In addition, the proposed hybrid approximation predicts the SCV_{D_n} of the probing stream quite accurately. In an inverse process, given the measured SCV_{D_n} , the intensity of the cross traffic can be inferred accordingly.

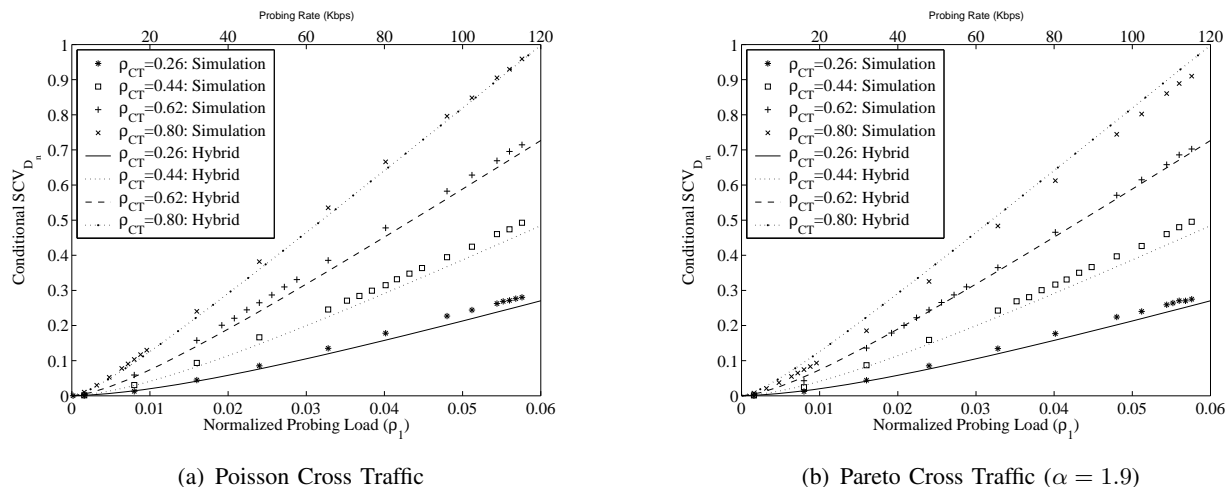


Fig. 14. The relationship between SCV_{D_n} and the Periodic probing load ρ_1 with two types of cross traffic and probing packet sizes $L = 40$ bytes using the hybrid approximation.

4) *Model Inference*: The exact inference can be accomplished numerically using a standard non-linear equation solver, since F^{-1} is already known. In our experiments, we used “fsolve” available in Matlab to achieve the inversion of F^{-1} . The inversion process in general only requires dozens of iterations in our practice and this numerical inversion finishes in less than 1 second using a Pentium III PC.

In order to demonstrate the estimation accuracy, we introduce the notion of “Relative Estimation Error”, which is defined as the ratio between the absolute estimation error and the true value of the cross traffic intensity. Figure 15 shows the accuracy of the intensity inference of the cross traffic using the $M_1 + M_2/GI_i/1$ queue for Poisson-type cross traffic. The estimation accuracy improves with the normalized probing load increasing. When the normalized probing load is larger than 0.05, the relative estimation error is less than 5%.

In Figure 16, the accuracy of the intensity inference of the cross traffic using the hybrid approximation for $D + M/GI_i/1$ is depicted for Pareto cross traffic ($\alpha = 1.9$) in the single-hop path depicted in Figure 4. The estimation accuracy improves with the normalized probing load increasing. However, further increasing the probing load does not improve the inference accuracy significantly. The reasons can be two fold. One is that the Pareto-type cross traffic already breaks the assumption of Poisson cross traffic in $D + M/GI_i/1$. The other may be due to the nature of the approximate calculation for SCV_{D_n} of the periodic probing stream. Nevertheless, when the normalized probing load is larger than 0.02, the relative estimation error is less than 10%.

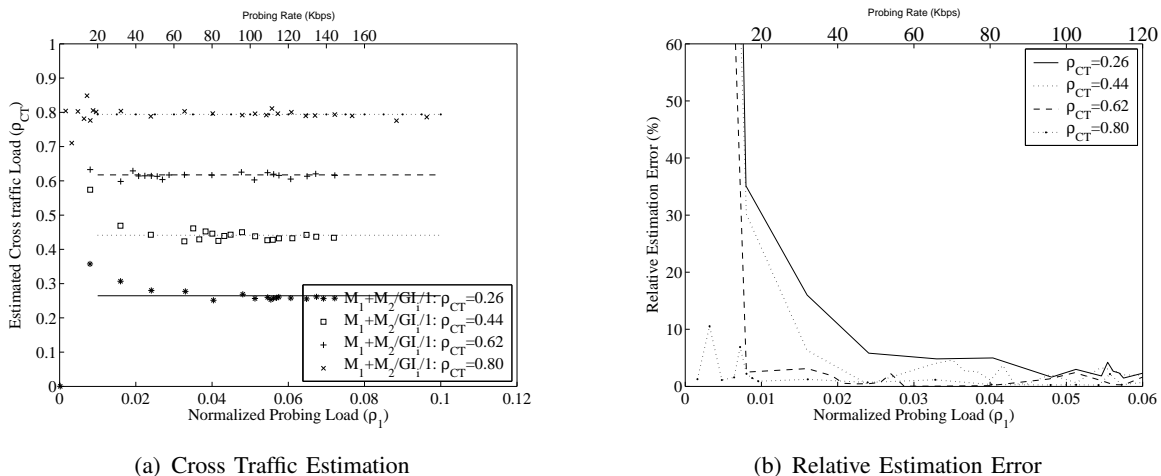


Fig. 15. Accuracy of the Poisson probing using $M_1 + M_2/GI_i/1$ with Poisson-type cross traffic and the probing size $L = 40$ bytes in the single-hop path depicted in Figure 4.

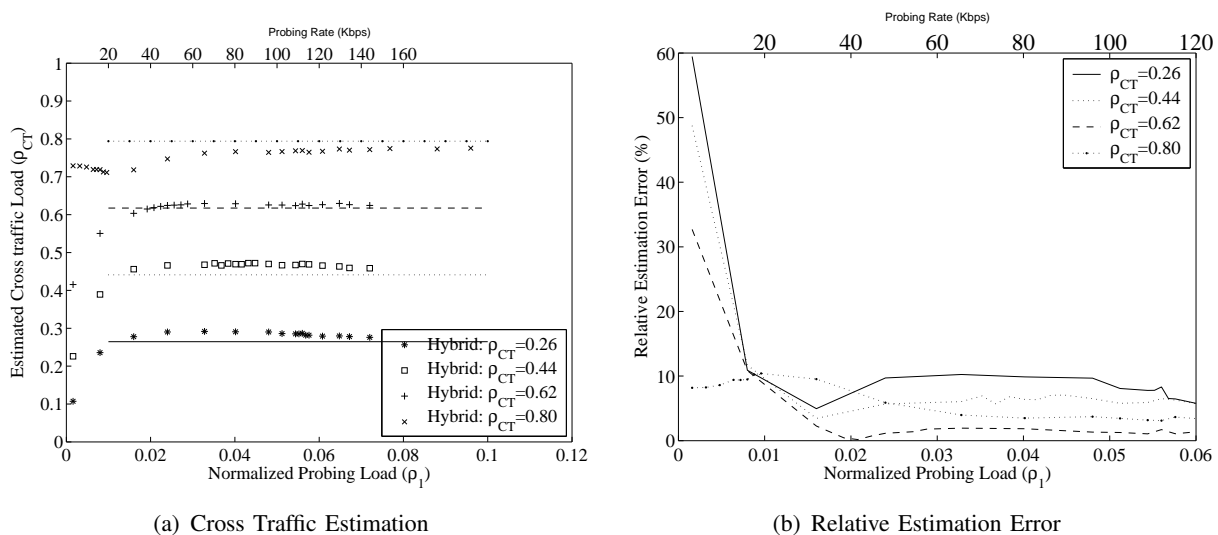


Fig. 16. Accuracy of the Periodic probing using the hybrid approximation for $D + M/GI_i/1$ with Pareto-type cross traffic ($\alpha = 1.9$) and the probing size $L = 40$ bytes in the single-hop path depicted in Figure 4.

E. Summary

The network topology used in the simulation experiments of this section has only a single-hop. Despite the simplicity of the simulation model, the loss-based model exhibits the difficulty with loss event detection when the probing load is small and, hence, it is not suitable for providing a light-weight available bandwidth measurement. We continue the performance evaluation of the SCV-based inference scheme along more general multi-hop paths in Section V.

V. MULTI-HOP EFFECTS

The loss process and the delay process of the probing stream in Figure 2 were analyzed using $M_1 + M_2/M/1/K$, $M_1 + M_2/GI_i/1/\infty$ and $D + M/GI_i/1/\infty$ presented in Section IV-B, Section IV-C and Section IV-D, respectively. These models assume a single congested link along an end-to-end path in the sense that this link produces most of the losses and significant queuing delays on this end-to-end path. Because this single-congested link assumption is still widely believed to be true [29], we focus on the evaluation of our SCV-based inference scheme on multi-hop paths with only one congested link in this section. However, with multiple tight links, underestimation of the available bandwidth may occur because SCV of the probing stream tends to be cumulative. This issue will be investigated further in our future work.

The network path used in this section is a 3-hop path ($R_1 - R_2 - R_3 - R_4$), as in Figure 17. The probing packets enter the network at router R_1 and exit at router R_4 . The links along the path are classified as the tight link and non-tight links. The tight link has the smallest link capacity along the path and the rest of the links along the path are referred as non-tight links [1]. The tight link has capacity C_t and available bandwidth A_t . The non-tight links have the same capacity C_{nt} and available bandwidth A_{nt} . Cross traffic is generated at each link from 10 random sources. These cross traffic connections only traverse one hop on this path. The arrival process of the cross traffic follows a Pareto distribution with a shape value of $\alpha = 1.9$. The cross-traffic packet sizes follow an exponential distribution with a mean value of 441 bytes. By controlling the arrival rate of the cross traffic, the load of the links can be varied. The probing packet size is 40 bytes. The path tightness factor β is defined as the ratio between the available bandwidth in the non-tight links A_{nt} and that in the tight link A_t (i.e., $\beta = \frac{A_{nt}}{A_t} \geq 1$).

Shown in Figure 17, link 2 is the tight link ($C_t = 10$ Mbps) and link 1 and 3 are non-tight links ($C_{nt} = 50$ Mbps). Those access links have a bandwidth capacity of 100 Mbps. Four cross traffic load scenarios at the tight link (link 2) are simulated: $\rho_{CT} = 0.80$, $\rho_{CT} = 0.62$, $\rho_{CT} = 0.44$, $\rho_{CT} = 0.26$. For $\rho_{CT} = 0.80$ at link 2, the available bandwidth of the links along this path from R_1 to R_4 are $\{42.0, 2.0, 42.0\}$ Mbps, respectively. The cross traffic rates at other links are configured the same as the one at link 2. Therefore, only link 2 is mostly congested and link 1 and 3 have a light load 1/5 of the load of link 2. The path tightness factor $\beta = 21$. The buffer sizes of link 1 and 3 are set to 500 packets and the buffer size of link 2 is set to 200 packets. The access links have buffers of 500 packets each. In this section, simulation

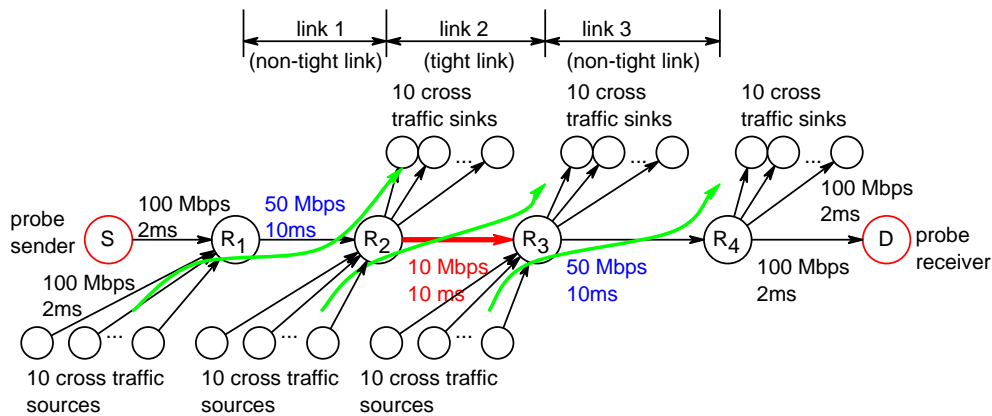


Fig. 17. A 3-hop path used in ns-2 simulation.

experiments are conducted using ns-2 [30] to include the effects of various network components on the performance of the proposed SCV-based inference scheme, including protocols, propagation delay, and so on.

In Figure 18, the conditional SCV_{D_n} of a Poisson probing stream at the output of each link is plotted. At link 1, the overall load is light; therefore, the SCV_{D_n} of the probing stream is around 1. SCV_{D_n} at the output of link 1 is almost the same as the SCV of the arrival process of this Poisson probing stream ($SCV_{I_n} = 1$). This can be explained using the light approximation in [31]. After the congested link at link 2, a significant SCV_{D_n} is observed and this SCV_{D_n} traverses links 3, and is finally measured at the probe receiver with little distortion.

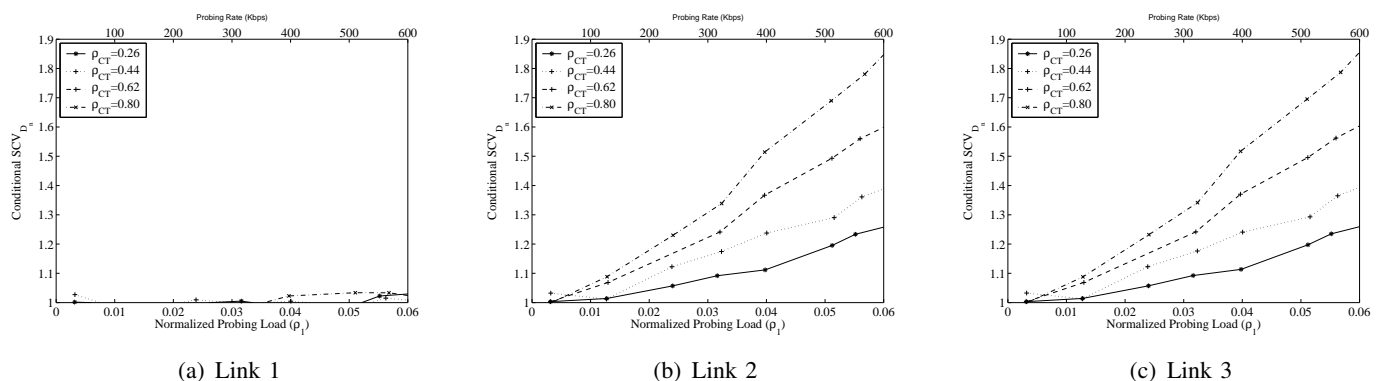


Fig. 18. The relationship between the conditional SCV_{D_n} and the Poisson probing rate with Pareto-type ($\alpha = 1.9$) cross traffic along the 3-hop path depicted in Figure 17.

Figure 19 shows the accuracy of the load inference of the cross traffic using the $M_1 + M_2/GI_i/1$ model for the cross traffic of Pareto-type with a shape value of $\alpha = 1.9$. The estimation accuracy improves with an increasing probing rate for four cross traffic load scenarios. When the probing rate ρ_1 is larger than 400

Kbps, the relative estimation error is below 10%. However, further increasing the probing rate does not improve the inference accuracy significantly. Note that the normalized probing load on the congested link capacity (10 Mbps) is only 0.04. In contrast, using self-congestion based available bandwidth measurement [1], [2], the probing load during the measurement period is a few Mbps in order to fill in the residual capacity of link 2 for the purpose of measurement.

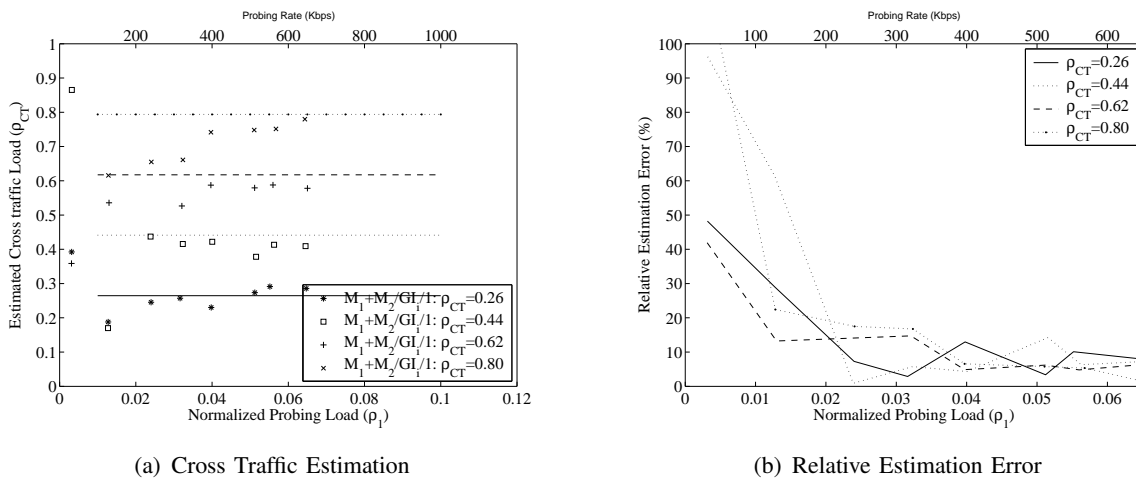


Fig. 19. Accuracy of the Poisson probing using $M_1 + M_2/GI_i/1$ with Pareto-type cross traffic ($\alpha = 1.9$) and the probing size $L = 40$ bytes along the 3-hop path depicted in Figure 17.

The convergence speed of the inference scheme using Poisson probing is illustrated in Figure 20 for four cross traffic load scenarios when the probing rate is $\lambda_1 = 2025$ packets/s. After approximately 4000 probes, the estimation process converges and the relative estimation error is below 10%. Note that the probing rate is $\lambda_1 = 2025$ packets/s; therefore, within a few seconds, the probing process converges. In addition, the packet size is 40 bytes and the bottleneck link capacity is 10 Mbps; hence, the normalized probing load is quite light, approximately 6.5% of the bottleneck capacity.

We implemented the proposed SCV-based probing scheme, namely, SCVProbe, in ns-2. SCVProbe has two variants: Poisson and periodic. SCVProbe is evaluated against Pathload in terms of probing accuracy, convergence time and overhead. The probing accuracy is evaluated by comparing the available bandwidth estimates against the real value; the convergence time is calculated considering the total time for the tool to provide an estimate; the probing overhead is the total data bytes used by the measurement tool during one measurement session.

The simulation configuration is the same as in Figure 17. The simulation duration lasts for 300 seconds. The cross traffic connections start randomly between $t = 0$ and $t = 10$. The measurement tools start at

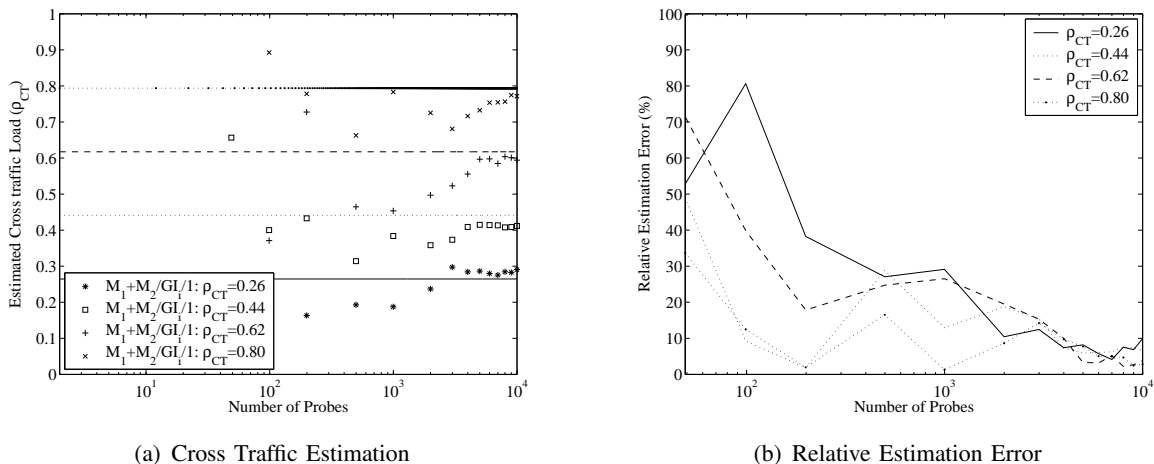


Fig. 20. Evolution of the estimated cross traffic load vs. the number of probes with a Poisson probing rate $\lambda_1 = 2025$ packets/s and the probing size $L = 40$ bytes and Pareto-type cross traffic ($\alpha = 1.9$) along the 3-hop path depicted in Figure 17.

$t = 20$. When the tools terminate and return with an available bandwidth estimate, we restart the tool immediately. We run the periodic SCVProbe with the probing sequence of 4000 packets and probing packet size of 40 bytes and the probing rate $\lambda_1 = 1599$ packets/s. We configure Pathload using the recommended parameters in [1], as shown in Table I.

TABLE I
SIMULATION PARAMETERS OF PATHLOAD FOR NS-2 EXPERIMENTS

Fleet Length	12 streams
Stream Length	100 packets
Bandwidth Resolution	$\omega = 400$ Kbps
Grey Region Resolution	$\chi = 400$ Kbps
PCT threshold	0.55
PDT threshold	0.4
Aggregate threshold	0.7

In Figure 21 are shown the measurement results of SCVProbe and Pathload. Pathload returns with a range of available bandwidth, and the middle point of this range tends to be larger than the real available bandwidth. The available bandwidth estimates measured by SCVProbe are quite close to the real value and they have relative error rates smaller than 20% for the four scenarios with different cross traffic loads. More importantly, these estimates by the periodic SCVProbe tend to be smaller than the real available bandwidth. This conservative estimation helps not to over-use the network available bandwidth for actual applications.

Our current implementation of SCVProbe uses a constant number of probing packets for one measure-

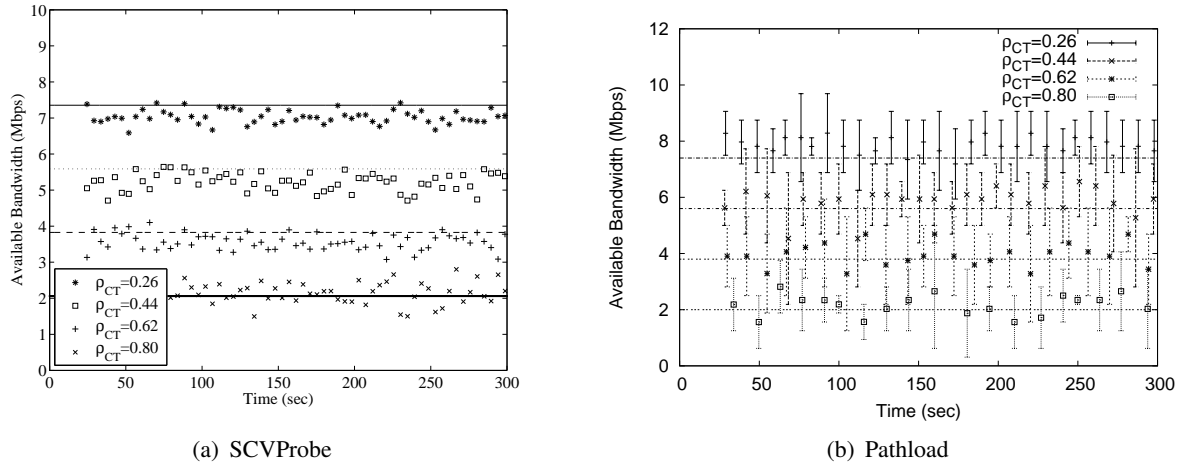


Fig. 21. Comparison of measurement accuracy along the 3-hop path depicted in Figure 17.

ment session. Therefore, the overhead and convergence time are constant. For periodic SCVProbe with the probing rate $\lambda_1 = 1599$ packets/sec and the probing sequence of 4000 packets and the packet size of 40 bytes, the convergence time is around 2.5 seconds and the overhead of each SCVProbe measurement session is 160 Kbytes. Figure 22 shows the probing overhead and the convergence time of Pathload. Because Pathload is an iterative algorithm, the actual convergence time depends on the dynamic behavior of the network. In Figure 22(a), the convergence time of most measurement sessions are shown to be larger than 10 seconds. In particular, when the cross traffic is heavily loaded ($\rho_{CT} = 0.8$), Pathload takes longer time to converge. Correspondingly, the probing overhead is over 1 Mbps mostly in Figure 22(b).

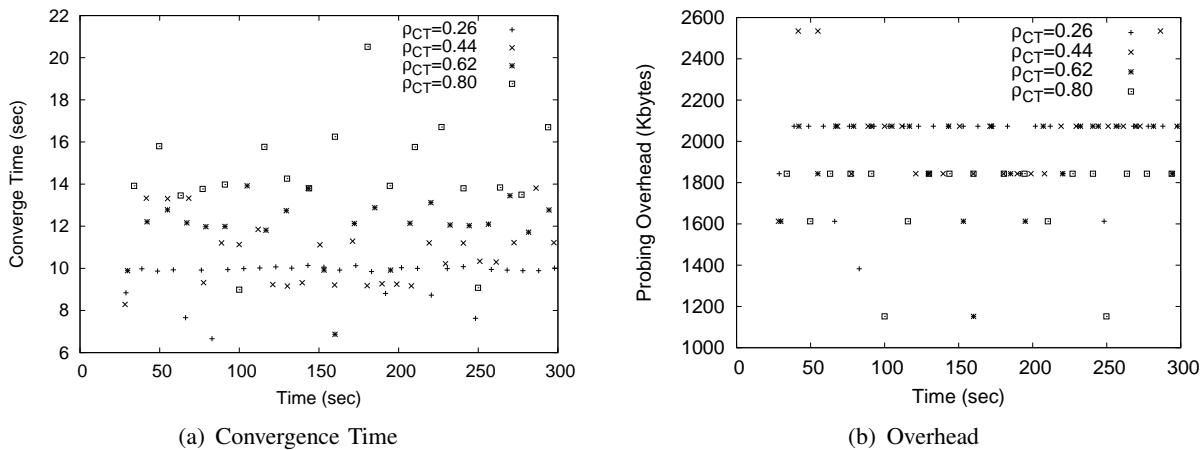


Fig. 22. Probing convergence time and overhead of Pathload along the 3-hop path shown in Figure 17.

The above comparison study indicates that SCVProbe achieves similar or even better measurement accuracy than Pathload with much less probing time and smaller overhead.

VI. CONCLUSION

In this paper we investigated model-based available bandwidth measurement via the use of an active probing stream. A general end-to-end measurement framework, capable of unifying the current research approaches and highlighting insights in the measurement practice, is proposed. This framework helps to design more efficient probing methodologies and to enhance the understanding of the network performance.

A comparison was conducted between the performance of the available bandwidth measurements based on loss models and on delay models. The queueing analysis and the extensive simulations indicate that the delay-based measurement exhibits more advantages over the loss-based measurements, such as accuracy, overhead, robustness.

The queueing analysis on the delay process of the probing stream shows that a transition point exists for the delay gap ratio when the path load $\rho = 1$. From a queueing analysis point of view, our study confirms the basic self-congestion principle of the delay gap model [1]–[3], [11]. The extensive simulations revealed that the buffer size appears to be part of the reason that Pathload and IGI tend to overestimate the available bandwidth. The SCV-based inference scheme converges within a few seconds and the relative estimation error can achieve to be less than 10% when the normalized probing load is larger than 0.04. However, the current SCV inference scheme assumes that there exists only a single congested link along the end-to-end path. Nevertheless, the basic methodology can be extended for paths with multiple congested links by modelling the path as a series of tandem queues. The delay process of the probing stream along these tandem queues is difficult to analyze in an exact queue analysis approach; therefore, we are investigating an approximate analysis of the departure process of these tandem queues in a decomposition approach [28].

ACKNOWLEDGMENT

The authors would like to thank Manish Jain for sharing his ns-2 simulation codes of Pathload. They would also like to thank the anonymous reviewers' constructive comments, and the editor Jennifer Hou for her efforts in handling the paper.

REFERENCES

- [1] M. Jain and C. Dovrolis, "End-to-end available bandwidth: measurement methodology, dynamics, and relation with TCP throughput," *IEEE/ACM Transactions on Networking (TON)*, vol. 11, no. 4, pp. 537–549, 2003.
- [2] N. Hu and P. Steenkiste, "Evaluation and characterization of available bandwidth probing techniques," *IEEE JSAC Special Issue in Internet and WWW Measurement, Mapping, and Modeling*, vol. 21, no. 6, pp. 879–893, Aug. 2003.

- [3] V. J. Ribeiro, R. H. Riedi, R. G. Baraniuk, J. Navratil, and L. Cottrell, "pathChirp: Efficient available bandwidth estimation for network path," in *Proceedings of PAM2003*, 2003.
- [4] B. Melander, M. Bjorkman, and P. Gunningberg, "A new end-to-end probing and analysis method for estimating bandwidth bottlenecks," in *GLOBECOM 2000 - IEEE Global Telecommunications Conference*, vol. 1, San Francisco, Nov. 2000, pp. 415 – 420.
- [5] V. Sharma and R. Mazumdar, "Estimating traffic parameters in queueing systems with local information," *Perform. Eval.*, vol. 32, no. 3, pp. 217–230, 1998.
- [6] S. Alouf, P. Nain, and D. Towsley, "Inferring network characteristics via moment-based estimators," in *Proceedings of INFOCOM*, vol. 2, Piscataway, NJ, USA, 2001, pp. 1045–54.
- [7] K. Salamatian, B. Bruno, and T. Bugnazet, "Interpretation of losses observed on the Internet by inferring traffic characteristics," in *Proc. of the DIMACS Workshop on Internet and WWW Measurement, Mapping and Modeling*, Rutgers University, Piscataway, NJ, USA, Feb. 13 - 15 2002.
- [8] X. Hei, D. H. K. Tsang, and B. Bensaou, "Available bandwidth measurement using Poisson probing on the Internet," in *Proc. 23rd IEEE International Performance Computing and Communications Conference, IPCCC 2004*, Phoenix, Arizona, 14-17 Apr. 2004, pp. 207–214. [Online]. Available: <http://hdl.handle.net/1783.1/1933>
- [9] —, "On model-based available bandwidth measurement on the Internet," in *Proc. Symposium on Performance Evaluation of Computer and Telecommunication Systems, SPECTS'04*, San Jose, California, July 25 - 29 2004, pp. 423–430.
- [10] X. Hei, B. Bensaou, and D. H. K. Tsang, "A light-weight available bandwidth inference methodology in a queueing analysis approach," in *Proceedings of ICC*, vol. 1, Seoul, Korea, 16-20 May 2005, pp. 120–124.
- [11] J. Strauss, D. Katabi, and F. Kaashoek, "A measurement study of available bandwidth estimation tools," in *Proceedings of the conference on Internet measurement conference*. ACM Press, 2003, pp. 39–44.
- [12] C. Dovrolis, P. Ramanathan, and D. Moore, "What do packet dispersion techniques measure?" in *Proceedings of INFOCOM*, vol. 2, Piscataway, NJ, USA, 2001, pp. 905–14.
- [13] K. Lai and M. Baker, "Nettimer: a tool for measuring bottleneck link bandwidth," in *Proceedings of 3rd USENIX Symposium on Internet Technologies and Systems. (USITS'01)*, Berkeley, CA, USA, 2001, pp. 123–34.
- [14] L. Ciavattone, A. Morton, and G. Ramachandran, "Standardized active measurements on a tier 1 IP backbone," *IEEE Communications Magazine*, vol. 41, no. 6, pp. 90 –97, Jun. 2003.
- [15] V. Paxson, G. Almes, J. Mahdavi, and M. Mathis, "RFC 2330: Framework for IP performance metrics," May 1998, status: INFORMATIONAL. [Online]. Available: <ftp://ftp.internic.net/rfc/rfc2330.txt>, <ftp://ftp.math.utah.edu/pub/rfc/rfc2330.txt>
- [16] V. Raisanen, G. Grotefeld, and A. Morton, "RFC3432: Network performance measurement with periodic streams," 2002.
- [17] K. Park and W. Willinger, *Self-similar network traffic and performance evaluation*. New York : Wiley, 2000.
- [18] W. Leland, M. Taqqu, W. Willinger, and D. Wilson, "On the self-similar nature of Ethernet traffic (extended version)," *IEEE/ACM Transactions on Networking*, vol. 2, no. 1, pp. 1–15, Feb. 1994.
- [19] V. Paxson and S. Floyd, "Wide area traffic: the failure of Poisson modeling," *IEEE/ACM Transactions on Networking (TON)*, vol. 3, no. 3, pp. 226–244, 1995.
- [20] J. Cao, W. S. Cleveland, D. Lin, and D. X. Sun, "Internet traffic tends toward Poisson and independent as the load increases," in *Nonlinear Estimation and Classification*, C. Holmes, D. Denison, M. Hansen, B. Yu, and B. Mallick, Eds. Springer, New York, 2002.
- [21] K. Claffy, G. Miller, and K. Thompson, "The nature of the beast: recent traffic measurements from an internet backbone," 1998. [Online]. Available: <http://www.caida.org/outreach/papers/1998/Inet98/>
- [22] S. McCreary and K. Claffy, "Trends in wide area IP traffic patterns - a view from Ames Internet exchange," in *ITC Specialist Seminar*, vol. 2, Sep. 2000. [Online]. Available: <http://www.caida.org/outreach/papers/2000/AIX0005/AIX0005.pdf>
- [23] D. Stanford and W. Fischer, "The interdeparture-time distribution for each class in the $\sum_i M_i/G_i/1$ queue," *Queueing Systems*, vol. 4, no. 3, pp. 179–91, July 1989.
- [24] W. Fischer and D. Stanford, "Approximations for the per-class waiting time and interdeparture time in the $\sum_i GI_i/G_i/1$ queue," *Performance Evaluation*, vol. 14, no. 2, pp. 65–78, 1992.
- [25] IKR, "University of Stuttgart, IKR Simlib library 2.2." [Online]. Available: <http://www.ind.uni-stuttgart.de/INDSimLib/>
- [26] S. Q. Li and J. D. Pruneski, "The linearity of low frequency traffic flow: an intrinsic I/O property in queueing systems," *IEEE/ACM Transaction on Networking*, vol. 5, no. 3, pp. 429–443, 1997.
- [27] A. Mukherjee, "On the dynamics and significance of low frequency components of internet load," *Journal of Internetworking: Research and Experience*, vol. 5, no. 4, Dec. 1994.
- [28] W. Whitt, "Towards better multi-class parametric-decomposition approximations for open queueing networks," *Annals of Operations Research*, vol. 48, no. 1-4, pp. 221–48, Jan. 1994.
- [29] W. Wei, B. Wang, D. Towsley, and J. Kurose, "Model-based identification of dominant congested links," in *Proceedings of the conference on Internet measurement conference*. ACM Press, 2003, pp. 115–128.
- [30] "The network simulator: ns-2." [Online]. Available: <http://www.isi.edu/nsnam/ns/>
- [31] W. Whitt, "A light-traffic approximation for single-class departure processes from multi-class queues," *Management Science*, vol. 34, no. 11, pp. 1333–1346, 1988.

APPENDIX I

THE $M_1 + M_2/M/1/K$ QUEUE

Following the derivations in [6], the intensity of the cross traffic and the buffer size were estimated using the scheme of $P_L - q_L$ (using (5) and (7)), the scheme of $P_L - q_N$ (using (5) and (8)), the scheme

of $q_L - q_N$ (using (7) and (8)) and the scheme of $P_L - R$ (using (5) and (6)).

$$P_L = \frac{(1 - \rho)\rho^K}{1 - \rho^{K+1}}, \quad (5)$$

$$R = \frac{1}{\mu(1 - \rho)} - \frac{K\rho^K}{\mu(1 - \rho^K)}, \quad (6)$$

$$q_L = \left(\frac{\lambda_1}{\lambda_2}\right) \frac{(1 - a)^{-1} - (b/a)^{K+1}(1 - b)^{-1}}{(b/a)^{K+1} - 1}, \quad (7)$$

$$q_N = 1 - \left(\frac{\lambda_1}{\lambda_2}\right) \left(\frac{1 - \rho}{1 - \rho^K}\right) \left(\frac{1}{b^{K+1} - a^{K+1}}\right) \cdot \left(\frac{a^2(1 - (\rho a)^K)}{(1 - a)(1 - \rho a)} - \frac{b^2(1 - (\rho b)^K)}{(1 - b)(1 - \rho b)}\right), \quad (8)$$

where

$$a = \frac{\lambda_1 + \lambda_2 + \mu + \sqrt{(\lambda_1 + \lambda_2 + \mu)^2 - 4\lambda_2\mu}}{2\lambda_2},$$

$$b = \frac{\lambda_1 + \lambda_2 + \mu - \sqrt{(\lambda_1 + \lambda_2 + \mu)^2 - 4\lambda_2\mu}}{2\lambda_2}.$$

APPENDIX II

THE $M_1 + M_2/GI_i/1/\infty$ QUEUE

In this section, the derivation of the $M_1 + M_2/GI_i/1$ queue is repeated following [23]. Let $\Phi_i(s)$ ($i = 1, 2$) be the Laplace-Stieltjes transform (LST) of the service time distribution of class i and $\eta_i(s)$ be the LST of the distribution of a busy cycle consisting solely of class i customers.

$$\eta_i(s) = \Phi_i(s + \lambda_i(1 - \eta_i(s))) \quad (9)$$

The inter-arrival times (i.a.t.) between successive customers of class 1 are assumed to be i.i.d., denoted as r.v. I. Customers from class 2 arrive at the queue in the form of i.i.d. in the continuous time.

Let W_n be C_n 's waiting time (excluding service), and $F_n = W_n + X_n$ its flow time. The LST of class 1 flow time, $\Phi_{F_n}(s)$, is the product of the waiting time LST $\Phi_{W_n}(s)$ and $\Phi_{X_n}(s) = \Phi_1(s)$. The LST of the composite service time is denoted as $\Phi_Z(s)$. Let B_n represent the busy cycle (possibly null) of the cross traffic packets served after C_n until either C_{n+1} enters service or the server becomes idle. In the busy cycle B_n , the cross traffic packet arrivals are counted as B_n^1, B_n^2, \dots . Let V_n be the interval (possibly null) from C_n 's departure from the system until C_{n+1} enters service.

$$D_n = V_n + X_{n+1}$$

The exact Laplace-Stieltjes transform of the inter-departure time distribution for each class in the $M_1 + M_2/GI_i/1$ queue on a first-come-first-served basis was derived and summarized as follows.

$$\begin{aligned} \Phi_{D_n}(s) &= \Phi_{V_n}(s)\Phi_{X_{n+1}}(s) = \Phi_{V_n}(s)\Phi_1(s), \\ \Phi_{V_n}(s) &= \Theta_1(s) + \Theta_2(s)\Phi_\beta(s), \\ \Theta_1(s) &= E \{ e^{-sB_n}; I_n \leq F_n + B_n \}, \\ \Theta_2(s) &= E \{ e^{-sB_n}; I_n > F_n + B_n \}, \\ \Phi_\beta(s) &= \frac{\frac{\lambda_1}{\lambda+s} + \frac{\lambda_2}{\lambda+s}A(s)[\Phi_2(s) - \eta_2(s + \lambda_1)]}{1 - \frac{\lambda_2}{\lambda+s}\eta_2(s + \lambda_1)}, \\ A(s) &= \frac{\lambda_1}{\lambda - \lambda_2\Phi_2(s)}, \\ \Theta_2(s) &= \Phi_{F_n}(\lambda - \lambda_2\eta_2(s + \lambda_1)), \\ \Theta_1(s) &= A(s)[1 - \Phi_{F_n}(\lambda - \lambda_2\eta_2(s + \lambda_1))], \\ \eta_2(s) &= \Phi_2(s + \lambda_2(1 - \eta_2(s))), \\ \Phi_{F_n}(s) &= \Phi_{W_n}(s)\Phi_{X_n}(s) = \Phi_{W_n}(s)\Phi_1(s), \\ \Phi_{W_n}(s) &= \frac{(1 - \rho)s}{s - \lambda(1 - \Phi_Z(s))}, \\ \Phi_Z(s) &= \frac{\lambda_1}{\lambda}\Phi_1(s) + \frac{\lambda_2}{\lambda}\Phi_2(s). \end{aligned}$$