

Internet Traffic Characterization Using Packet-Pair Probing

Yu Cheng

Department of Electrical and Computer Engineering
Illinois Institute of Technology
Chicago, Illinois, USA 60616
Email: cheng@iit.edu

Vikram Ravindran and Alberto Leon-Garcia

Department of Electrical and Computer Engineering
University of Toronto
Toronto, Ontario, Canada M5S 3G4
Email: {v.ravindran, alberto.leongarcia}@utoronto.ca

Abstract—This paper presents an edge-based Internet traffic characterization approach. Our objective is to estimate the marginal distribution and the correlation structure of the packet arrival process at a queue, by sending probe packet pairs with a specific dispersion to sample the traffic. The aggregate work load process inferred from the output dispersions is a compound process of the packet arrival process and the packet size distribution. We show that the packet arrival marginal distribution and the packet size distribution can be decoupled by using the probability generating function; given one of the distributions, the other can then be estimated. We use the fact that the Internet packet size follows a known multi-modal distribution. Moreover, multiple series of packet pairs with different input dispersions can be used to estimate the packet arrival process at different time scales, and therefore to estimate the Hurst parameter, which characterizes the long-range dependence, by generating the variance-time plot. While the traffic characterization techniques are developed and validated in a single-queue context, we indicate how the techniques can be applied to a black box system for end-to-end quality of service provisioning along a multi-hop path.

I. INTRODUCTION

Most Internet traffic measurements today use *passive* monitoring. Data traces are collected and recorded by specific hardware in a router, and then forwarded to a measurement collector for analysis to extract statistical characteristics of the Internet traffic [1]. Passive measurement systems have several disadvantages in addition to the requirement for specific data collection hardware. (1) The data collection consumes a large amount of network resources. The high data volume generated by a large network requires large amount of memory in routers, storage and computation resources at the collector, and transmission bandwidth between them [1], [2]. (2) The “collect, transmission, and analyze” procedure is not appropriate for on-line traffic monitoring, which is important for adaptive QoS provisioning and resource allocation in dynamic network conditions. (3) The per-hop traffic characterization is not convenient for end-to-end QoS analysis. For such an objective, we need to characterize both the input and output processes of a queue to analyze concatenated queues, and to characterize the superposition and splitting of packet streams along the path. This is normally not an easy task except for the case of Poisson modeling [3]. (4) The heterogeneous and decentralized nature of the Internet (for example, a path between two end users often passes multiple administrative domains) implies

that one cannot rely on the cooperation of individual servers and routers to aid in network measurement [4]. (5) The Internet service provider (ISP) may regard the collected data traces as confidential. The unavailability of interior network status makes it difficult to design distributed and edge-based admission control and traffic engineering schemes [5].

To avoid the disadvantages of the router-based passive measurement, we propose in this paper *edge-based Internet traffic characterization* using packet-pair dispersion techniques. The *packet-pair dispersion* at a given point is defined as the duration from the time when the last bit of the first probe packet arrives at the point to the time that the last bit of the second probe packet arrives. The basic idea of our measurement approach is that if a closely-spaced probe packet pair is input into a queue, the output dispersion will reflect the aggregate service time of the cross traffic arriving during the input dispersion. The output dispersion measurement can be mapped into a traffic load sample. Internet traffic is commonly measured and modelled as having a wide-sense-stationary increment process, at least during a measurement window [4], [6]–[9], so we can use a series of probe packet pairs to collect statistical information about the traffic load during a time interval. The resulting samples allow us to estimate the marginal distribution of the aggregate traffic arrival process in bits/second or bytes/second.

Our objective is to characterize the *packet arrival process*, which is important for packet-level QoS analysis [10]–[13] and packet-based switch design [14], using packet-pair probing. The challenge is that the byte-level or bit-level work load measured from the output dispersions is a compound process of the packet arrival process and the packet size distribution. We show that the packet arrival marginal distribution and the packet size distribution can be decoupled by using the probability generating function (PGF), under a common simplifying assumption that the packet sizes are independently and identically distributed (iid), and independent of the packet arrival process [11], [15]. We also use the fact that the Internet packet size follows a known multi-modal marginal distribution [2], [7], [16]–[19]. We show how multiple series of probe packet pairs with different input dispersions can be used to sample the packet arrival process at different time scales. By generating the *variance-time plot* [11], [16], we can then

estimate the *Hurst parameter* H that characterizes the long range dependence (LRD) [9]. On the other hand, we also demonstrate that if the packet arrival marginal distribution is available, for example from a measurement collector, the packet size distribution can also be estimated using the PGF tool. This estimation approach avoids the resource-consuming packet-header analysis involved in directly measuring the packet sizes. It has been found that packet size distribution is very application dependent; monitoring packet size distribution is helpful in identifying which network applications are active at a certain time [2], [16], [18].

Packet dispersion techniques, including *packet-pair dispersion* and *packet-train dispersion* [20], [21], are the most common end/edge-based approach for bottleneck capacity or available bandwidth measurement, which fall in the field of network monitoring and inference termed as *Internet tomography* [4]. Internet tomography technologies can greatly facilitate customers' awareness of network status, and bring higher flexibility and scalability to the network management. The edge-based measurement techniques presented in this paper provide new methods for traffic characterization in Internet tomography.

In this paper, we analyze and validate the proposed traffic characterization techniques in the context of a single-hop path. However, we discuss how the edge-based traffic characterization is in fact a *black box* system that can be applied for end-to-end QoS provisioning along a path. Specifically, we can model an end-to-end path as a *virtual single queue* served with the bottleneck capacity of the original path. The probe packet pair is launched at the source end of the path and the output dispersion is measured at the destination end of the path. The marginal distribution and the Hurst parameter estimated from the dispersion measurements can then be used to characterize the virtual arrival process associated with the virtual queue. Such a virtual arrival process aggregates all the buffering effects, superposition effects, splitting effects, and queue concatenation effects along the path. The QoS achieved in the virtual queue would be equivalent to that achieved along the original end-to-end path.

The remainder of the paper is organized as follows. Section II reviews related work. Section III describes the queueing model, that provides the relationship between the input dispersion, the cross traffic, and the output dispersion. Section IV presents the traffic characterization techniques. Section V presents simulation results to demonstrate the performance of the proposed techniques. Section VI discusses the black-box modeling. Section VII gives the concluding remarks.

II. RELATED WORK

Whereas it is commonly accepted that Internet traffic has a self-similar nature [2], [9], [10], [15], [16], [22], the packet arrival process may have different marginal distributions, e.g. Poisson distribution [10], [23], [24], Gaussian distribution [6], [8], [11], [22], [25] or α -stable distribution [26], [27]. Each different marginal distribution has a different queueing analysis [6], [8], [25], [28]–[30]. The packet pair probing

proposed in this paper, by identifying the marginal distribution of the packet arrival process, can be helpful in determining the proper queueing analysis tool.

Packet-pair dispersion techniques with various modifications, [20], [31], [32] and the references therein, are the most common approach to measure the bottleneck capacity of a path. The bottleneck capacity will be used in the proposed procedure for traffic characterization. Therefore, the Internet traffic characterization can be combined with the capacity estimation to make the packet-pair dispersion technique a versatile framework for Internet measurement.

The performance of packet dispersion techniques is critically dependent on the clock resolution of the end nodes involved in the probing. High resolution timing and accurate clock synchronization well below 1 μs can be achieved with special hardware synchronized to the Global Positioning System (GPS) time [2], [33], or with personal computer (PC) based software clock where the CPU clock cycle (TSC) register is used to keep track of time at very high resolution [34]. The actual dispersion measurement takes place at a single point where the clock synchronization is not of importance.

In this paper, we develop traffic characterization techniques based on the assumption that packet size is iid and independent of the packet arrival process [11], [15], which is termed as the *iid packet-size assumption* for convenience. However, some Internet traffic measurements [7], [16] show that the sequence of packet sizes is long-range dependent. We nevertheless make the assumption for the following three reasons. (1) The Internet is evolving into a common communication infrastructure for all kinds of multimedia applications, where the aggregate packet stream entering a core router in the backbone will consist of traffic from different wireless/wireline access networks as well as from various applications (e.g. voice, video, email, web browsing). In such a heterogeneous, large-scale, high-multiplexing environment, the iid packet-size assumption can be justified in some degree. (2) It is a common practice in Internet tomography to use simplifying assumptions of spatial and temporal independence to devise practical inference algorithms [4]. We admit that much work remains to be done to incorporate spatial or temporal dependency models into network inference problems. (3) It is valid to model a virtual queueing system with a virtual packet arrival process independent of the packet size distribution, as long as it is an equivalent black box system, with the same input and output statistics, to the original end-to-end path under consideration.

III. PACKET-PAIR DISPERSION BEHAVIOR

In this section, we describe a queueing model to formalize the packet-pair dispersion behavior. We model a store-and-forward router on a path as a first-come-first-serve (FCFS) queue served at a fixed rate of C bits per second (or equivalently bytes per second), the capacity of the link connected to the router. A network path consisting of links from 1 to n is modeled as the concatenation of n queues, where each queue has serving capacity C_i ($1 \leq i \leq n$). We define a *probe packet pair* as two consecutive probe packets that are transmitted

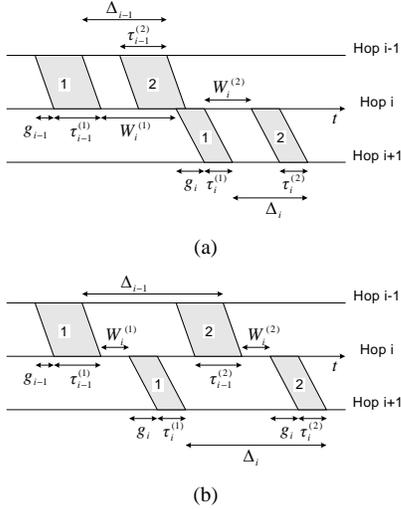


Fig. 1. Two types of dispersion behaviors: (a) the second packet reaches queue i before the first leaves, (b) the second packet reaches queue i after the first has departed.

along a path between the source and the destination. The first packet in the pair has a size $L^{(1)}$ bits, and the second packet has a size $L^{(2)}$. We refer to other packets in the network other than the probe traffic as *cross traffic*. Our objective is to identify the statistical characteristics of the cross traffic.

A. Dispersion Governing Equations

The packet-pair dispersion at a given point is defined as the duration from the time when the last bit of the first packet arrives at the point to the time when the last bit of the second packet arrives. The dispersion of a packet pair leaving queue i is denoted as Δ_i , also termed as the *output dispersion* of queue i . We assume that the propagation delay along a link is constant, and therefore the output dispersion from a queue is equal to the *input dispersion* at the next downstream queue.

The packet-pair dispersion behavior at a queue i can fall in two cases [20], as illustrated in Fig. 1. In one case the second probe packet arrives before the first probe packet leaves the queue, and in the other case the second probe packet arrives after the first probe packet has left. For queue i , the two cases of the output dispersion are described by the following governing equations [20]:

$$\Delta_i = \begin{cases} \sigma_i^{(2)} & \text{if } \Delta_{i-1} \leq \sigma_i^{(1)} \\ \Delta_{i-1} + \sigma_i^{(2)} - \sigma_i^{(1)} & \text{otherwise} \end{cases} \quad (1)$$

where $\sigma_i^{(j)} = \tau_i^{(j)} + W_i^{(j)}$ and $\tau_i^{(j)} = \frac{L^{(j)}}{C_i}$ for $j = 1, 2$. $W_i^{(1)}$ is the queueing delay of first probe packet, and $W_i^{(2)}$ is the queueing delay of packet 2 after packet 1 has departed from the queue.

B. Cross-Traffic Effect

In the dispersion governing equations, the cross traffic affects the output dispersion through the queueing delay $W_i^{(j)}$. To explicitly describe the cross traffic effect, we classify the

packet-pair probing into two cases. The first case is defined as *united-pair* probing, where the two probe packets fall in the same busy period of the queueing process. In this case, the server is always busy before the second packet arrives, while the first probe packet may depart before or after the second probe packet arrives. Let $A(t)$ denote the cross-traffic arrival process (in bytes) to the queue, the output dispersion in this case is

$$\Delta_i = \frac{A(\Delta_{i-1}) + L^{(2)}}{C_i}.$$

The second case is defined as *divided-pair* probing, where the two probe packets fall in different busy periods of the queueing process. In this case, the first probe packet definitely leaves before the second probe packet arrives, as shown in Fig. 1(b). The output dispersion in this case is

$$\Delta_i = \Delta_{i-1} + \frac{L^{(2)} - L^{(1)}}{C_i} + W_i^{(2)} - W_i^{(1)}.$$

In summary, the output dispersion under the cross-traffic effect can be expressed as

$$\Delta_i = \begin{cases} \frac{A(\Delta_{i-1})}{C_i} + \frac{L^{(2)}}{C_i} & \text{for united-pairs} \\ \Delta_{i-1} + \frac{L^{(2)} - L^{(1)}}{C_i} + W_i^{(2)} - W_i^{(1)} & \text{for divided-pairs.} \end{cases} \quad (2)$$

C. Input Dispersion Selection

We will use the united-pair probing for traffic characterization. In the divided-pair case, both idle period(s) and busy periods occur during the input dispersion so the cross traffic cannot be sampled properly. According to (1), a united-pair can be guaranteed if we conservatively select the input dispersion or design the probe packet size according to

$$\Delta_{i-1} \leq L^{(1)}/C_i \leq \sigma_i^{(1)}. \quad (3)$$

Basically, we can use a large size packet for the first probe packet to guarantee the united-pair probing, and a small size packet for the second probe packet to reduce the measurement traffic load.

IV. TRAFFIC CHARACTERIZATION

In this section, we show that the packet-pair dispersion measurements can be used to estimate the statistical characteristics, i.e. the marginal distribution and the autocovariance function of the cross-traffic packet arrival process. To facilitate the estimation, we use the iid packet-size assumption and the fact that Internet packet size follows a known multi-modal distribution.

A. Internet Packet Size Distribution

Measurements of Internet traces have shown that packets of a few sizes tend to make up the vast majority of Internet packets. Exactly which sizes are prominent depends on applications and network protocols used. The prominent sizes have varied over the years, as new applications emerge, new protocols replace old ones, and protocol behaviors change [2], [7], [16]–[19].

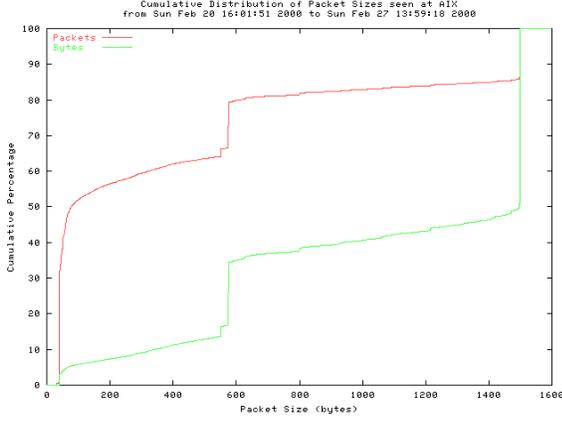


Fig. 2. Packet size distribution at NASA AIX, February 21–27,2000

As an example, we refer to the measurements performed by the Cooperative Association for Internet Data Analysis (CAIDA) on the traffic passing through the NASA Ames Internet Exchange (AIX) between February 21st and 27th, 2000 [17]. The cumulative distribution function of the packet sizes is re-plotted in Fig. 2, where it can be seen that the vast majority of packets were of a few sizes, i.e. 40 bytes, 576 bytes, and 1500 bytes. The researchers attributed this result to the behavior of the Transport Control Protocol (TCP), with the 40-bytes packets corresponding to a TCP acknowledgements, the 576-bytes packets to TCP implementations without path discovery, and the 1500-bytes packets to TCP implementations with packet discovery. The 1500-bytes mode also corresponds to the Ethernet payload size limit. As CAIDA does not provide the histogram, we produce an approximate numerical probability mass function (PMF) from Fig. 2:

$$P[S = k] = \begin{cases} 0 & k \in \{0, 1, \dots, 39\} \\ 0.50 & k = 40 \\ \frac{0.13}{575-40} & k \in \{41, 42, \dots, 575\} \\ 0.17 & k = 576 \\ \frac{0.03}{1499-576} & k \in \{577, \dots, 1499\} \\ 0.17 & k = 1500 \end{cases} \quad (4)$$

Such a PMF will be used in Section V for performance evaluation.

B. Packet Arrival Marginal Distribution Estimation

We now develop a method to estimate the marginal distribution of packet arrival process, based on the probability generating function (PGF). Consider a united-pair with dispersion Δ that is sent to probe a queue. Cross-traffic can be completely described by two random variables, which are the random variable $N(\Delta)$ to denote the number of packets that arrive in a given interval Δ , and the random variable S to denote the packet sizes. The total amount of cross-traffic $A(\Delta)$ is:

$$A(\Delta) = \sum_{n=0}^{N(\Delta)} S_n \quad (5)$$

where S_n are iid random variables.

1) *PGF Analysis:* Let us define the PGFs of $N(\Delta)$ and X as

$$G_{N(\Delta)}(z) = E[z^{N(\Delta)}] = \sum_{n=0}^{\infty} P[N(\Delta) = n]z^n \quad (6)$$

$$G_S(z) = E[z^S] = \sum_{k=0}^{\infty} P[S = k]z^k. \quad (7)$$

According to the iid packet-size assumption, it is well-known that [35]

$$\begin{aligned} G_{A(\Delta)}(z) &= E[z^{A(\Delta)}] = E\left[E[z^{A(\Delta)}|N(\Delta)]\right] \\ &= E\left[(G_S(z))^{N(\Delta)}\right] = E[z^{N(\Delta)}]_{z=G_S(z)} \\ &= G_{N(\Delta)}(G_S(z)). \end{aligned} \quad (8)$$

With united-pair probing, the aggregate cross-traffic arrivals during the input dispersion can be obtained from the output dispersion measurement. Again, using Δ_{i-1} and Δ_i to represent the input and output dispersions associated with the queue under consideration respectively, we have

$$A(\Delta_{i-1}) = C_i \Delta_i - L^{(2)}. \quad (9)$$

By sending a sufficient number of probe packet pairs, the histogram of $A(\Delta)$ (the subscription of Δ_{i-1} is dropped for convenience) can be estimated, and therefore the corresponding PGF $G_{A(\Delta)}(z)$ can be calculated. Using $q_k(\Delta)$ to denote the histogram probability $P[A(\Delta) = k]$, the PGF $G_{A(\Delta)}(z)$ is

$$G_{A(\Delta)}(z) = \sum_{k=0}^{\infty} q_k(\Delta)z^k. \quad (10)$$

Given $G_{A(\Delta)}(z)$, our objective is to estimate the marginal distribution $P[N(\Delta) = n]$, briefly denoted as $p_n(\Delta)$, for $n = 0, 1, \dots, \infty$.

We now use the fact that Internet packet size takes only a few, say m , dominant values, by which $G_S(z)$ can be simplified as

$$G_S(z) = a_1 z^{l_1} + a_2 z^{l_2} + \dots + a_m z^{l_m} \quad (11)$$

where l_1, l_2, \dots, l_m , arranged as $l_1 < l_2 < \dots < l_m$, are the m dominant packet sizes and a_1, a_2, \dots, a_m are the corresponding probabilities. Define the probability vector $\mathbf{p} = (p_0, p_1, \dots, p_{l_m})$, where $p_{l_k} = a_k$ ($k = 1, \dots, m$) and other values in the vector are equal to 0. Use $\mathbf{p}^{(n)} = (p_0^{(n)}, p_1^{(n)}, \dots, p_{nl_m}^{(n)})$ to denote the probability vector of the random variable $S^{(n)} = \sum_{i=1}^n S_i$, and \otimes the convolution operation, we have

$$\mathbf{p}^{(n)} = \mathbf{p} \otimes \mathbf{p} \cdots \otimes \mathbf{p} \text{ (with } n \text{ fold)}. \quad (12)$$

For the PGF, we have

$$G_{S^{(n)}}(z) = (G_S(z))^n = \sum_{k=0}^{nl_m} p_k^{(n)} z^k. \quad (13)$$

TABLE I
AN EXAMPLE OF MATRIX \mathbf{P}

k_i	$p_{k_i}^{(0)}$	$p_{k_i}^{(1)}$	$p_{k_i}^{(2)}$	$p_{k_i}^{(3)}$	$p_{k_i}^{(4)}$	$p_{k_i}^{(5)}$	$p_{k_i}^{(6)}$	$p_{k_i}^{(7)}$	$p_{k_i}^{(8)}$	$p_{k_i}^{(9)}$	$p_{k_i}^{(10)}$
0	1	0	0	0	0	0	0	0	0	0	0
40	0	0.5000	0	0	0	0	0	0	0	0	0
80	0	0.0002	0.2500	0	0	0	0	0	0	0	0
1580	0	0	0.0001	0.1276	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
2156	0	0	0.0000	0.0001	0.0868	0.0002	0.0002	0.0002	0.0001	0.0001	0.0001
2196	0	0	0.0000	0.0001	0.0002	0.0723	0.0002	0.0002	0.0001	0.0001	0.0001
2236	0	0	0.0000	0.0001	0.0002	0.0002	0.0543	0.0002	0.0001	0.0001	0.0001
2276	0	0	0.0000	0.0001	0.0002	0.0002	0.0002	0.0380	0.0001	0.0001	0.0001
3776	0	0	0	0.0000	0.0000	0.0001	0.0001	0.0002	0.0259	0.0002	0.0002
4352	0	0	0	0.0000	0.0000	0.0000	0.0001	0.0001	0.0001	0.0199	0.0002
4392	0	0	0	0.0000	0.0000	0.0000	0.0001	0.0001	0.0001	0.0002	0.0166

Based on (13), (8) can be unwrapped as

$$\begin{aligned}
 G_{A(\Delta)}(z) &= G_{N(\Delta)}(G_S(z)) = \sum_{n=0}^{\infty} (G_S(z))^n p_n(\Delta) \\
 &= \sum_{n=0}^{\infty} \sum_{k=0}^{nl_m} p_k^{(n)} z^k p_n(\Delta) = \sum_{k=0}^{\infty} \left[\sum_{n=0}^{\infty} p_k^{(n)} p_n(\Delta) \right] z^k.
 \end{aligned} \tag{14}$$

Comparing (10) and (14), we can get

$$q_k(\Delta) = \sum_{n=0}^{\infty} p_k^{(n)} p_n(\Delta). \tag{15}$$

2) *Marginal Distribution Estimation*: It is possible to solve $p_n(\Delta)$ from (15). However, it is impractical to estimate $p_n(\Delta)$ up to $n \rightarrow \infty$. In practice, the cross-traffic input to a queue is limited by the upstream link capacity. Letting L_{min} denote the minimum packet size allowed in the network, so we can obtain an upper bound of $N(\Delta)$ as

$$N(\Delta) \leq N_{max}^{\Delta} = C_{i-1} \Delta / L_{min} \tag{16}$$

and (15) reduces to

$$q_k(\Delta) = \sum_{n=0}^{N_{max}^{\Delta}} p_k^{(n)} p_n(\Delta). \tag{17}$$

With vectors $\mathbf{p}^{(n)}$ ($n = 1, \dots, N_{max}^{\Delta}$) available from (12), define $\mathbf{p}_{(k)} = (p_k^{(0)}, p_k^{(1)}, \dots, p_k^{(N_{max}^{\Delta})})$. Assume that we can find $N_{max}^{\Delta} + 1$ points of k , i.e. $k_0, k_1, \dots, k_{N_{max}^{\Delta}}$, so that the $(N_{max}^{\Delta} + 1) \times (N_{max}^{\Delta} + 1)$ matrix \mathbf{P} , with $\mathbf{p}_{(k_i)}$ for $i = 0, 1, \dots, N_{max}^{\Delta}$ as the row vectors, is invertible. Moreover, from the histogram of $A(\Delta)$, we can have the values of $(q_{k_1}(\Delta), q_{k_2}(\Delta), \dots, q_{k_{N_{max}^{\Delta}}}(\Delta))$ to form a vector $\mathbf{q}(\Delta)$. Define the cross-traffic marginal distribution vector $\mathbf{p}(\Delta) = (p_0(\Delta), p_1(\Delta), \dots, p_{N_{max}^{\Delta}}(\Delta))$, which can be estimated as

$$\mathbf{p}(\Delta)^T = \mathbf{P}^{-1} \mathbf{q}(\Delta)^T. \tag{18}$$

We would like to emphasize that the solvability of $\mathbf{p}(\Delta)^T$ depends on the invertibility of matrix \mathbf{P} , which is closely related to the packet size distribution. However, in practice, the number, position, and magnitude of the dominant packet size modes vary with the applications. It is not an easy task to theoretically derive a general approach to construct

an invertible \mathbf{P} . In our experiments, we use the following heuristic approach to construct the matrix \mathbf{P} .

Step 1: As it is obvious that $p_0^{(0)} = 1$ and $p_0^{(i)} = 0$ for $i = 1, \dots, N_{max}^{\Delta}$, select $k_0 = 0$ and use $\mathbf{p}_{(0)}$ as the zeroth row of the matrix.

Step 2: For $i = 1, \dots, N_{max}^{\Delta}$, select k_i according to

$$k_i : p_{k_i}^{(i)} = \max_k (\mathbf{p}^{(i)}) = \max_{k=0,1,\dots,i \cdot l_m} (p_k^{(i)}) \tag{19}$$

Step 3: For $i = 1, \dots, N_{max}^{\Delta}$, use the vector $\mathbf{p}_{(k_i)}$ as the i -th row of matrix \mathbf{P} .

The intuitive idea behind the construction approach is to generate a matrix \mathbf{P} , the diagonal of which has dominant values much larger than other elements in the matrix; therefore the matrix is invertible. Our experiments show that the construction approach is robust in producing an invertible \mathbf{P} from the available packet size distributions. As an example, consider the CAIDA packet size distribution. If the input dispersion is set small enough to limit $N_{max}^{\Delta} = 10$, the obtained matrix \mathbf{P} is shown in Table I. We are carrying on theoretical investigations of the heuristic matrix construction approach.

C. Hurst Parameter Estimation

A wide-sense stationary $X(t)$ in discrete time is said to exhibit *long-range dependence*, if its autocorrelation function $r_X(k)$ decays with time lag k taking the form

$$r_X(k) \sim k^{-\beta}, \quad \text{as } k \rightarrow \infty \tag{20}$$

where $0 < \beta < 1$ and “ \sim ” denotes that the expressions on the two sides are asymptotically proportional to each other [9], [11], [36]. Note that LRD implies nonsummability of the correlations, i.e., $\sum_k r_X(k) = \infty$. The *Hurst parameter* H is commonly used to measure the degree of LRD, and is related to the parameter β in (20) by $H = 1 - \beta/2$.

Let the aggregated process $X^{(m)} = \{X_k^{(m)}\}$ be obtained by averaging the original traffic process X over non-overlapping intervals, with each interval being m time units in length, i.e. $X_k^{(m)} = \frac{1}{m} \sum_{t=m(k-1)+1}^{mk} X(t)$. It has been proved [37] that the long-range dependence can also be characterized by

$$\text{var}[X^{(m)}] \sim m^{-\beta}, \quad \text{as } m \rightarrow \infty, \quad 0 < \beta < 1. \tag{21}$$

Denoting the autocorrelation function of $X^{(m)}$ by $r_X^{(m)}(k)$, we also have

$$r_X^{(m)}(k) \sim k^{-\beta}, \quad \text{as } m \rightarrow \infty, k \rightarrow \infty, 0 < \beta < 1. \quad (22)$$

The expression (22) means that the correlation structure of $X(t)$ is asymptotically preserved under the time aggregation, so $X(t)$ is also defined to be *asymptotically second-order self-similar*. In fact, by the restriction $0 < \beta < 1$ or $\frac{1}{2} < H < 1$, asymptotic second-order self-similarity implies long-range dependence, and vice versa. Therefore in practice, *self-similarity* and *long-range dependence* are often used interchangeably [9].

Long-range dependence or self-similarity can be detected, and the corresponding value of H can be estimated in many ways [7], [36]. One simple and popular approach is the *variance-time plot* which makes use of the property that the “ $\log(\text{var}[X^{(m)}])$ vs. $\log(m)$ ” curve asymptotically tends to be linear with a slope of $-\beta$ [10], [11], [16].

The variance-time plot approach can be readily exploited by the packet-pair probing. We first determine the minimum usable input dispersion Δ_{min} , according to the clock resolution of the end nodes involved in the measurement. Define Δ_{min} as a time unit. By sending a series of probe packet pairs with input dispersions set as Δ_{min} , we can estimate the marginal distribution and therefore the variance of $X^{(1)}$ using the traffic characterization techniques presented in Section IV-B. Similarly, $\text{var}[X^{(m)}]$ can be estimated by packet-pair probing with input dispersion of $m\Delta_{min}$. The variance-time plot can then be applied to estimate the Hurst parameter.

D. Packet Size Distribution Estimation

We can also estimate the packet size distribution based on (8), if the packet arrival marginal distribution is available from traffic modeling or from historical data measurements. The estimation is straightforward as

$$G_S(z) = G_{N(\Delta)}^{-1}(G_{A(\Delta)}(z)). \quad (23)$$

For example, let's assume that the self-similar Internet traffic can be well modeled by an $M/G/\infty$ input process [10], [23]. The Marginal distribution is then Poisson with average rate λ packets/second, giving $G_{N(\Delta)} = e^{\lambda\Delta(z-1)}$. Using this result,

$$G_S(z) = 1 + \frac{1}{\lambda\Delta} \ln(G_{A(\Delta)}(z)). \quad (24)$$

With the PGF known, the probability vector \mathbf{p} can be obtained with an inverse fast Fourier transform (IFFT) operation.

V. PERFORMANCE EVALUATION

In this section, we present some computer simulation results to illustrate the performance of the proposed traffic characterization techniques. We develop a single-queue simulator using the *event-driven simulation* technique in the MATLAB language. Due to the paper length limit, here we focus on evaluating the accuracy in estimating the marginal distribution of the packet arrival process and the Hurst parameter for the LRD traffic. Simulation results demonstrating the accuracy of the packet-size distribution estimation are presented in

[41]. In all the simulation examples, we use the iid packet-size assumption, and the cross-traffic packet size distribution follows the CAIDA PMF given in (4).

A. Estimation with Poisson Arrival

We first use a simple example to illustrate the performance of the marginal distribution estimation technique proposed in Section IV-B. We simulate a FCFS queue with a serving capacity of 10 Mbps. The cross packet arrivals are generated as a Poisson process. The CAIDA PMF is used to generate iid packet sizes, corresponding to an average packet size of 444.1338 bytes. Packet pairs with size of $(L^{(1)}, L^{(2)}) = (1500, 40)$ bytes are sent for probing. The packet-pair input dispersion is set as 0.0012s so that united-pair measurement is guaranteed according to (3). The inter-pair interval is set as 0.05s. Two load scenarios with utilization $u = 0.8$ and $u = 0.99$ are simulated, where the packet arrival rates are 2.2516×10^3 and 2.7863×10^3 packets/s or 2.7019 and 3.3436 packets/dispersion, respectively. In this example, we assume that a proper N_{max}^Δ can be set according to the historical traffic measurements; the N_{max}^Δ according to (16) is normally a conservative configuration incurring unnecessary calculations. Specifically, set $N_{max}^\Delta = 12$ so that $\sum_{n=0}^{N_{max}^\Delta} p_n(\Delta) \approx 1.0000$ in both load scenarios.

We estimate the packet arrival marginal distribution from the output dispersion measurements by the following procedure:

- Calculate $\mathbf{p}^{(n)}$ according to (12), and construct the matrix \mathbf{P} using the approach given in Section IV-B.
- Send a reasonable number of probe packet pairs, 10000 and 30000 pairs in this example, with the selected input dispersion. Measure the output dispersions and convert the dispersions to traffic loads in bytes according to (9).
- Generate the histogram of $A(\Delta)$, where the estimated value of $\hat{q}_k(\Delta)$ is estimated as the frequency of the dispersions corresponding to $A(\Delta) = k$, for $k = 0, 1, 2, \dots, 1500N_{max}^\Delta$.
- Estimate the packet arrival marginal distribution $\hat{p}_n(\Delta)$, $n = 0, 1, \dots, N_{max}^\Delta$ according to (18).

The estimated marginal distributions in both load scenarios are given in Table II and compared with the true Poisson probabilities. We can have the following observations. (1) In both load scenarios, the probabilities as small as the order of 10^{-2} can be estimated with a reasonable accuracy. (2) It is desirable to capture the distributional discrepancy between the estimated probabilities and the real values by the standard *chi-square test* [38]. However, the marginal distribution is indirectly inferred rather than directly estimated by counting in different categories; therefore, chi-square test can not be applied. Instead, we use the metric of weighted mean relative difference (WMRD) proposed in [1] to measure the distributional discrepancy. Specifically,

$$\text{WMRD} = \frac{\sum_n |\hat{p}_n(\Delta) - p_n(\Delta)|}{\sum_n (\hat{p}_n(\Delta) + p_n(\Delta))/2}. \quad (25)$$

(3) The matrix inversion calculation in the estimation can lead to negative estimate values, if the to-be-estimated prob-

TABLE II
TRAFFIC CHARACTERIZATION WITH POISSON ARRIVAL PROCESS

$N(\Delta)$	$u = 0.8$					$u = 0.99$				
	10000 pairs		30000 pairs		True probabilities	10000 pairs		30000 pairs		True probabilities
	Estimated	Normalized	Estimated	Normalized		Estimated	Normalized	Estimated	Normalized	
0	0.0680	0.0692	0.0673	0.0676	0.0671	0.0376	0.0383	0.0361	0.0368	0.0353
1	0.1762	0.1793	0.1781	0.1789	0.1812	0.1110	0.1131	0.1121	0.1141	0.1181
2	0.2382	0.2424	0.2470	0.2482	0.2448	0.1959	0.1997	0.1950	0.1986	0.1974
3	0.2230	0.2269	0.2219	0.2230	0.2205	0.2088	0.2128	0.2150	0.2190	0.2200
4	0.1516	0.1543	0.1489	0.1496	0.1490	0.1756	0.1789	0.1783	0.1815	0.1839
5	0.0780	0.0794	0.0845	0.0849	0.0805	0.1276	0.1301	0.1359	0.1384	0.1230
6	0.0338	0.0344	0.0332	0.0333	0.0362	0.0722	0.0736	0.0648	0.0660	0.0685
7	0.0140	0.0142	0.0113	0.0114	0.0140	0.0344	0.0351	0.0344	0.0350	0.0327
8	-0.0007	0	0.0019	0.0019	0.0047	0.0181	0.0185	0.0104	0.0106	0.0137
9	-0.0004	0	0.0012	0.0012	0.0014	-0.0008	0	-0.0008	0	0.0051
10	-0.0005	0	-0.0005	0	0.0004	-0.0010	0	0.0031	0	0.0017
11	-0.0006	0	0.0045	0	0.0001	-0.0012	0	-0.0012	0	0.0005
12	-0.0007	0	-0.0009	0	0.0000	-0.0016	0	-0.0015	0	0.0001
WMRD		2.80%		2.27%			4.90%		4.07%	

abilities are too small. When calculating WMRD, we first set those *ineffective probability estimates* $\hat{p}_n(\Delta) = 0$, $n = n_{neg}, \dots, N_{max}^\Delta$, where n_{neg} is the smallest index of a negative estimate value, and then normalize the distribution to $\sum_{n=0}^{N_{max}^\Delta} \hat{p}_n(\Delta) = 1$. Estimation accuracy can be improved, reflected as a decreased WMRD, by collecting more dispersion samples. (4) The WMRD is larger for $u = 0.99$. The reason is that the random variable $N(\Delta)$ has a larger variance under a larger arrival rate.¹ Estimating with the same number of samples, the larger variance leads to the larger distributional discrepancy.

B. Estimation with fractional Brownian motion (FBM) arrival

In this example, we consider the marginal distribution estimation and the Hurst parameter estimation with a more practical input process, i.e. the FBM process, which has been extensively used to model the self-similar Internet traffic [6], [8], [11], [22], [25]. The standard (normalized) FBM process $\{Z(t) : t \geq 0\}$ with Hurst parameter $H \in [0.5, 1)$ is a centered Gaussian process with stationary increments that possesses the following properties [9]: (a) $Z(0) = 0$, (b) $\text{var}\{Z(t)\} = t^{2H}$, and (c) $Z(t)$ has continuous sample paths. The self-similar FBM input $\{A(t) : t \geq 0\}$ can be represented by

$$A(t) = \lambda t + \sigma Z(t) \quad (26)$$

where the mean arrival rate $E\{A(t)/t\} = \lambda$, and the variance $\text{var}[A(t)] = \sigma^2 t^{2H}$. Note that σ^2 is the variance of traffic in a time unit. The FBM is *exactly self-similar* according to the property (b) [9]. When $0.5 < H < 1$, the self-similar FBM is long-range dependent.

We again simulate a 10 Mbps queue. The CAIDA PMF is used as the packet size distribution. Four scenarios with different utilization and Hurst parameter combinations, i.e. $(u, H) = (0.6, 0.6)$, $(0.6, 0.7)$, $(0.8, 0.6)$, and $(0.8, 0.7)$, are simulated. The FBM process is generated using the modified Random Midpoint Displacement (RMD) algorithm [39], [40], where one time unit (t-unit) is set as 0.0006s. In different simulation scenarios, FBM processes are generated with λ tuned to the target utilization and $\sigma^2 = 2\lambda$. As the number

¹With Poisson distribution, mean and variance of the random variable are equal.

of packets in a time unit generated from the RMD algorithm may take fractional values, we ceil all the fractional values to integer values. Variance-time plot is used to calibrate the Hurst parameter of the ceiled FBM process.

In each scenario, we use two 25000-pairs probing series, one with the input dispersion of 1 t-unit and the other of 5 t-units, to estimate the marginal distributions of $N(1)$ and $N(5)$, respectively. For simplicity, we ignore the limit of the maximum packet size, and set $L^{(1)} = 750$ (3750) bytes when $\Delta = 1$ (5) t-units to guarantee the united-pair measurement. The inter-pair interval is set as 0.05s. Since FBM is exactly self-similar, the Hurst parameter can be estimated from $\text{var}[A(t)] = \sigma^2 t^{2H}$ as

$$\hat{H} = \frac{1}{2} \frac{\log(\hat{\text{var}}[N(5)]) - \log(\hat{\text{var}}[N(1)])}{\log(5)}. \quad (27)$$

In practice, Internet traffic is asymptotically self-similar, and the variance-time plot or other approaches should be used to estimate the Hurst parameter. This example with FBM input is to illustrate the validity of the multiple-timescale probing in estimating the correlation structure.

We set $N_{max}^\Delta = 25$ for the estimation, and discard ineffective probability estimates and normalize the estimated distributions as that done in the previous example. Since here we are estimating a continuous Gaussian distribution, we can apply the *Kolmogorov-Smirnov (K-S) test for Normality with the mean and standard deviation unknown* [38] to measure the estimation accuracy.

The K-S test statistic values D and the estimated Hurst parameters in all the four scenarios are presented in Table III, where we can have the following observations. (1) In all the scenarios, the K-S test static values (D_s) calculated from the effective probability estimates are smaller than the corresponding critical values (D_c) at the 1% level of significance (obtained from Table 9 on page 336 in [38]). Therefore, the hypothesis that the packet arrival process is Gaussian can be accepted at the 1% level of significance. (2) The D_s values when $\Delta = 5$ t-units are smaller than the corresponding values when $\Delta = 1$ t-unit. The reason is that the higher aggregation over a longer input dispersion makes the discrete distribution closer to the continuous Gaussian distribution. As an illustration, the sample cumulative distribution function and

TABLE III
TRAFFIC CHARACTERIZATION WITH FBM ARRIVAL PROCESS

u	H	0.6 (0.64 calibrated)		0.7 (0.73 calibrated)	
0.6	\hat{H}	0.61		0.67	
	K-S test	$\Delta = 1$	$\Delta = 5$	$\Delta = 1$	$\Delta = 5$
	D_s	0.1388	0.0727	0.1318	0.0906
	D_c	0.2940	0.2124	0.3110	0.2248
0.8	\hat{H}	0.62		0.66	
	K-S test	$\Delta = 1$	$\Delta = 5$	$\Delta = 1$	$\Delta = 5$
	D_s	0.1386	0.0752	0.1310	0.1200
	D_c	0.2940	0.2248	0.3110	0.2350

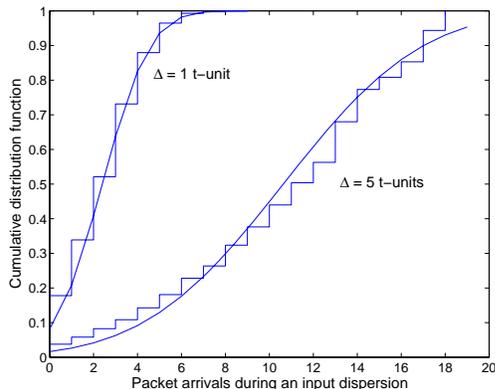


Fig. 3. Kolmogorov-Smirnov test of the marginal distribution estimated with the FBM input, with $u = 0.8$ and $H = 0.7$ (calibrated as 0.73).

the hypothesized Gaussian distribution function in the K-S test for $(u, H) = (0.8, 0.7)$ are plotted in Fig. 3. (3) In all the scenarios, the estimated Hurst parameters \hat{H} are close to the true values. However, in all the cases, $\hat{H} < H$ (calibrated). The reason is that in the estimation procedure we have to give up some ineffective probability estimates, which are associated with those large values of $N(\Delta)$, due to the limited number of samples; therefore, the estimate value of the variance is smaller than the true value. We define the variance underestimation as the *smoothing effect* for convenience. With the same number of samples, the smoothing effect is more severe when the to-be-estimated variance is larger, which then leads to the underestimation of the Hurst parameter according to (27). A straightforward approach to improve estimation accuracy is to collect more measurement samples. However it is well known that such an approach is not very efficient in the presence of LRD [7]. Combining the edge-based traffic characterization approach with other efficient estimation tools for LRD traffic, e.g. the wavelet analysis, is very interesting and being under our investigation.

In the previous discussions and examples, we have emphasized that N_{max}^{Δ} should be selected properly for the estimation. In practice, N_{max}^{Δ} can be adaptively adjusted. At first, a large enough initial value is selected according to historical traffic measurement. After a measurement period, N_{max}^{Δ} can be reduced to $n_{neg} - 1$ to avoid ineffective estimations. After a measurement period where all the estimate probabilities are effective, N_{max}^{Δ} can be increased by 1 for more accurate estimation.

VI. BLACK-BOX MODELING

With edge-based network measurement and inference, the network is treated as a black-box system. The capacity and available bandwidth estimation techniques in the literature and the traffic characterization technique in this paper are developed to infer the internal status of the black box by analyzing its input and output statistics. Seen from the system perspective, it is desirable to derive a mathematical description of the black box system, ignoring the details of how the system is constructed. For example, in signal processing, a transfer function is usually used to describe a linear system, ignoring what circuits are used to construct the system.

We deem that the edge-based traffic characterization can also be used to derive a mathematic description of an end-to-end path. We define the bottleneck link capacity along a path as the *path capacity*. In a QoS context, the path capacity may be determined by a scheduler. The end-to-end path can be modeled as a virtual single-hop FCFS queue that is fed with a virtual packet arrival process and served with the path capacity. With black-box modeling, it is required to determine the marginal distribution and the correlation structure of the virtual packet arrival process so that the input and output statistics of the virtual queue are exactly the same as those associated with the original end-to-end path. The traffic characterization techniques developed in this paper can be used to determine the virtual packet arrival process modulated by iid packet-sizes. Such a system model conceals the service details along the path into the virtual queueing system, such as the actual packet arrival process at each hop, the statistical relationship between the packet arrival process and the packet size distribution, correlation structure of the packet size sequence, queueing effect at each hop, and the traffic multiplexing/splitting effect along the path.

The black-box modeling has the potential to bring great convenience to end-to-end QoS provisioning and edge-based admission control. When a new flow comes in, queueing analysis can be applied to the virtual queue to check whether a target end-to-end QoS measure, e.g. the outage probability of a delay bound, can still be guaranteed after admitting the new flow. In addition, the virtual packet arrival process needs to be re-characterized periodically by packet-pair probing to trace the dynamics of the network status.

VII. CONCLUSIONS

In this paper, we propose an edge-based approach to estimate the statistical characteristics of the Internet traffic. For the estimation, a series of probe packet pairs with a selected dispersion are sent to sample the cross traffic. From the output dispersion measurements, the aggregate work load process can be inferred, which is a compound process of the packet arrival process and the packet size distribution. Under the iid packet-size assumption, we develop a probability generating function tool to decouple the packet arrival marginal distribution and the packet size distribution; given one of the distributions, the other can then be estimated. The fact that Internet packet size follows a known multi-modal distribution is utilized to

facilitate the estimation. When self-similarity or LRD presents, multiple series of packet pairs with different input dispersions are used to perform a multi-timescale sampling and estimation, and an estimate of the Hurst parameter is obtained by generating the variance-time plot. Computer simulation results in a single-queue context demonstrate that the proposed traffic characterization techniques have a reasonable estimation accuracy. In addition, we present a system perspective that the probing-based traffic characterization can be used to establish a black-box model for an end-to-end path, where all the service details along the path are concealed into an equivalent single-hop virtual queue.

For future work, we are investigating how to apply some advanced statistical analysis techniques, e.g. the wavelet analysis, to the dispersion samples to improve estimation accuracy when LRD presents. We also plan to experiment with practical Internet traces to further validate the edge-based traffic characterization. The black-box modeling also needs to be examined with simulations and experiments along multi-hop paths.

REFERENCES

- [1] N. Duffield, C. Lund, and M. Thorup, "Learn more, sample less: control of volume and variance in network measurement," *IEEE Trans. Inform. Theory*, vol. 51, pp. 1756–1775, May 2005.
- [2] C. Fraleigh *et al.*, "Packet-level traffic measurements from the Sprint IP backbone," *IEEE Network*, vol. 17, pp. 6–16, Nov./Dec. 2003.
- [3] S. Song, J. K.-Y. Ng, and B. Tang, "Some results on the self-similarity property in communication networks," *IEEE Trans. Commun.* vol. 52, pp. 1636–1642, Oct. 2004.
- [4] M. Coates, A. O. Hero III, R. Nowak, and B. Yu, "Internet tomography," *IEEE Signal Processing Mag.*, vol. 19, pp. 47–65, May 2002.
- [5] F. P. Kelly, P. B. Key and S. Zachary, "Distributed admission control", *IEEE J. Select. Areas Commun.*, vol. 18, pp. 2617–2628, Dec. 2000.
- [6] I. Norros, "A storage model with self-similar input," *Queueing Systems*, vol. 16, pp. 387–396, 1994.
- [7] P. Abry and D. Veitch, "Wavelet analysis of long-range-dependent traffic," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2–15, Jan. 1998.
- [8] H. S. Kim and N. S. Shroff, "Loss probability calculations and asymptotic analysis for finite buffer multiplexers," *IEEE/ACM Trans. Networking*, vol. 9, pp.755–768, Dec. 2001.
- [9] K. Park and W. Willinger, "Self-similar network traffic: an overview," in *Self-Similar Network Traffic and Performance Evaluation*, edited by K. Park and W. Willinger, John Wiley & Sons, Inc. 2000, pp. 1–38.
- [10] V. Paxson and S. Floyd, "Wide area traffic: the failure of Poisson modeling," *IEEE/ACM Trans. Networking*, vol. 3, pp. 226–244, Jun. 1995.
- [11] A. Erramilli, O. Narayan, and W. Willinger, "Experimental queueing analysis with long-range dependent packet traffic," *IEEE/ACM Trans. Networking*, vol. 4, pp. 209–223, Apr. 1996.
- [12] S. Nananukul, "Multiplexing of periodic arrival processes with different packet sizes," *IEEE Trans. Commun.*, vol. 50, pp. 1055–1057, Jul. 2002.
- [13] J.M. Pitts, J.A. Schormans, C.I. Phillips, and J.M. Griffiths, "Accurate delay bounds for real-time IP services in presence of variable packet size," *IEE Electronic Letters*, vol. 27, pp. 166–167, Feb. 2001.
- [14] Y. Ganjali, A. Keshavarzian, and D. Shah, "Cell switching versus packet switching in input-queued switches," *IEEE/ACM Trans. Networking*, vol. 13, pp. 782–789, Aug. 2005.
- [15] W. Leland, M. Taqqu, W. Willinger, and D. Wilson, "On the self-similar nature of Ethernet traffic (extended version)," *IEEE/ACM Trans. Networking*, vol. 2, pp. 1–15, Feb. 1994.
- [16] L. Yao, M. Agapie, J. Ganbar, and M. Doroslovacki, "Long range dependence in Internet backbone traffic," in *Proc. IEEE ICC*, vol. 3, May 2003, pp. 1611–1615.
- [17] Cooperative Association for Internet Data Analysis, NASA Ames Internet Exchange Packet Length Distributions, [Online]. Available: http://www.caida.org/analysis/AIX/plen_hist/
- [18] D.J. Parish, K. Bharadia, A. Larkum, I.W. Phillips, and M.A. Oliver, "Using packet size distribution to identify real-time networked applications," *IEE Proc.-Commun.*, vol. 150, August 2003.
- [19] K. Thompson, G. Miller, and R. Wilder, "Wide area Internet traffic patterns and characteristics," *IEEE Network*, vol. 11, pp. 10-23, Nov-Dec. 1997.
- [20] C. Dovrolis, P. Ramanathan, and D. Moore, "Packet-dispersion techniques and a capacity-estimation methodology," *IEEE/ACM Trans. Networking*, vol. 12, pp. 963–977, Dec. 2004.
- [21] M. Jain and C. Dovrolis, "End-to-end available bandwidth: measurement methodology, dynamics, and relation with TCP throughput," *IEEE/ACM Trans. Networking*, vol. 11, pp. 537–549, Aug. 2003.
- [22] W. Willinger, M. S. Taqqu, R. Sherman, and D. V. Wilson, "Self-similarity through high-variability: statistical analysis of Ethernet LAN traffic at the source level," *IEEE/ACM Trans. Networking*, vol. 5, pp. 71–86, Feb. 1997
- [23] M. Parulekar and A.M. Markowski, "M/G/ ∞ input processes: A versatile class of models for network traffic," in *Proc. IEEE INFOCOM*, Apr. 1997, pp. 419–426.
- [24] N. Likhanov and B. Tsybakov, "Analysis of an ATM buffer with self-similar ("fractal") input traffic," in *Proc. IEEE INFOCOM*, vol. 3, Apr. 1995, pp. 985–992.
- [25] J. Choe and N. Shroff, "A central-limit-theorem-based approach for analyzing queue behavior in high-speed networks," *IEEE/ACM Trans. Networking*, vol. 6, pp. 659–671, Oct. 1998.
- [26] A. Karasaridis and D. Hatzinakos, "Network heavy traffic modeling using α -stable self-similar processes," *IEEE Trans. Commun.*, vol. 49, pp. 1203-1214, Jul. 2001.
- [27] T. Mikosch, S. Resnick, H. Rootzén, and A. Stegeman, "Is network traffic approximated by stable Lévy motion or fraction Brownian motion?" *Ann. Appl. Probab.*, vol. 12, no. 1, pp. 23–68, 2002.
- [28] M. Parulekar and A. Makowski, "Tail probabilities for a multiplexer driven by M/G/ ∞ input processes (i): preliminary asymptotics," *Queueing Syst.*, vol. 27 pp. 271–296, 1997.
- [29] N. G. Duffield, "Queueing at large resources driven by long-tailed M/G/ ∞ -modulated processes," *Queueing Syst.*, vol. 28, pp. 245-266, 1998.
- [30] F. C. Harmantzis, D. Hatzinakos, and I. Lambadaris, "Effective bandwidths and tail probabilities for Gaussian and stable self-similar traffic," *Proc. IEEE ICC*, vol. 3, May 2003, pp. 1515-1520.
- [31] V. Jacobson, "Congestion avoidance and control," in *Proc. ACM SIGCOMM*, Aug. 1988, pp. 314–329.
- [32] N. Kapoor, L.-J. Chen, M. Y. Sanadidi, and M. Gerla, "CapProbe: A simple and accurate capacity estimation technique," in *Proc. ACM SIGCOMM*, 2004, pp. 67–78.
- [33] J. Micheel, S. Donnelly, and I. Graham, "Precision timestamping of network packets," in *Proc. 1st ACM SIGCOMM Workshop on Internet Measurement*, 2001, pp. 273–277.
- [34] A. Pasztor, "Accurate active measurement in the Internet and its applications," PhD Thesis, Department of Electrical and Electronic Engineering, The University of Melbourne, Australia, 2003
- [35] A. Leon-Garcia, *Probability and Random Processes for Electrical Engineering*, Addison Wesley Longman, Reading, MA, USA, 2nd edition, 1994.
- [36] J. Beran, R. Sherman, M.S. Taqqu, and W. Willinger, "Long-range dependence in variable-bit-rate video traffic," *IEEE Trans. Commun.*, vol. 43, pp. 1566–1579, Feb./March/Apr. 1995
- [37] D.R. Cox, "Long-range dependence: a review," in *Statistics: An Appraisal*, H.A. David and H.T. David, Eds., The Iowa State University Press, Ames, Iowa, 1984, pp. 55–74.
- [38] A.O. Allen, *Probability, Statistics, and Queueing Theory: With Computer Science Applications*, Academic Press, INC., NY, USA, 1978.
- [39] W.-C. Lau, A. Erramilli, J. Wang, and W. Willinger, "Self-similar traffic generation: the random midpoint displacement algorithm and its properties," in *Proc. IEEE ICC'95*, pp. 466–472, June 1995.
- [40] I. Norros, P. Mannersalo, and J. L. Wang, "Simulation of fractional Brownian motion with conditional random displacement," *Advances in Performance Analysis*, vol. 2, no. 1, pp. 77–101, 1999.
- [41] V. Ravindran, "Real-time measurement of service levels on IP networks", Master Thesis, Department of Electrical and Computer Engineering, University of Toronto, Jun. 2005.