Modeling Internet backbone traffic at the flow level

Chadi Barakat*, Patrick Thiran, Gianluca Iannaccone, Christophe Diot, Philippe Owezarski

Abstract— **Our goal is to design a traffic model for non congested Internet backbone links, which is simple enough to be used in network operation, while being as general as possible. The proposed solution is to model the traffic at the flow level by a Poisson shot-noise process. In our model, a flow is a generic notion that must be able to capture the characteristics of any kind of data stream. We analyze the accuracy of the model with real traffic traces collected on the Sprint IP (Internet Protocol) backbone network. Despite its simplicity, our model provides a good approximation of the real traffic observed in the backbone and of its variation. Finally, we discuss the application of our model to network design and dimensioning.**

Index Terms— **Traffic modeling, Poisson shot noise, noncongested IP backbone links, measurements.**

I. Introduction

Modeling the Internet traffic is an important issue. It is unlikely that we will be able to understand the traffic characteristics, predict network performance (e.g., for Quality of Service (QoS) guarantees or Service Level Agreement (SLA) definition), or design dimensioning tools without analytical models. The successful evolution of the Internet is tightly coupled to the ability to design simple and accurate models.

The objective of this work is to design a traffic model that can be used by network administrators to assist in network design and management. Such a model needs to be simple, i.e., it has to be fast to compute and to rely on simple parameters that can easily be acquired by a router. Currently, network operators have very basic information about the traffic. They mostly use SNMP [10] that provides average throughput information over 5 minutes intervals. An analytical model could provide more accurate information on the traffic. It could be used in various applications such as detection of anomalies (e.g., denial of service attacks or link failures), prediction of traffic growth, or assessment of the impact on network traffic of a new customer or of a new application. Consequently, a second desired property of the model is to be protocol and application agnostic: it needs to be general enough to evaluate link throughput independently of the application nature and of the transport mechanism.

Packet level models for high speed links are difficult to calibrate, because of the high level of multiplexing of numerous flows whose behavior is strongly influenced by the transport protocol and by the application. In addition, monitoring the traffic at the packet level becomes critical at OC-192 and above link speeds.

Recently, a new trend has emerged, which consists in modeling the Internet traffic at the flow level (see [5] and the references therein). A flow here is a very generic notion. It can be a TCP (Transmission Control Protocol) connection or a UDP (User Datagram Protocol) stream (described by source and destination IP addresses, source and destination port numbers, and the protocol number), or it can be a destination address prefix (e.g., destination IP address in the form a.b.0.0/16). Flows arrive at random times and share the available bandwidth in the network according to certain rules. From a simplicity standpoint, it is much easier to monitor flows than to monitor packets in a router. Tools such as NetFlow already provide flow information in Cisco routers [1].

In this paper, we propose a model that relies on flow-level information to compute the total (aggregate) rate of data observed on an IP backbone link. We are interested in capturing the dynamics of the traffic at short timescales (i.e., in the order of hundreds of milliseconds). For the purpose of modeling, the traffic is viewed as the superposition (i.e., multiplexing) of a large number of flows that arrive at random times and that stay active for random periods. As explained earlier, a flow is a generic notion that must be able to capture the characteristics of any kind of data stream.

In contrast to other works in the literature (e.g., [5], [7], [18]), we choose to model a link that is *not* congested (congestion possibly appears elsewhere on the flow path). This assumption is valid, and is in fact the rule, for backbone links that are generally over-provisioned (i.e., the network is designed so that a backbone link utilization stays below 50% in the absence of link failure [15]). It is driven by our main objective to provide a link dimensioning tool usable in backbone network management.

The contribution of this work is the design of a flow-based Internet traffic model using simple mathematical

* Corresponding author.

[1] http://www.cisco.com/warp/public/732/Tech/netflow

tools (Poisson shot-noise). Thanks to the notion of *shots* we introduce in the purpose of modeling flow transmission rates, our model is able to compute the total rate of data in the backbone using flows' characteristics (i.e., arrivals, sizes, durations). Once the model is introduced, the paper focuses on its confrontation to real data collected on the Sprint IP backbone network. This confrontation illustrates the efficiency of the model in computing the traffic in the backbone and its variation. We then discuss the application of our model to network design and management. In particular, we study the impact of the different parameters of the model (flow arrival rate, flow size, flow duration) on the characteristics of the traffic in the backbone.

In the next section, we survey the related literature and position our contribution. Section III describes the traces we use throughout the paper for the validation of our model. In Section IV, we present our model and we analyze its performance in Section V. Section VI explains how shots can be determined, and Section VII discusses some issues related to the practical use of our model. In Section VIII, the model is confronted to the real traces. We discuss the use of our model to network dimensioning in Section IX. Conclusions and perspectives on our future work are presented at the end.

II. Related work

Many authors ([11], [14], [21], [24]) have analyzed the Internet traffic and have shown that it behaves in agreement with long range dependent and asymptotically self-similar processes. This finding made a revolutionary step departing from more traditional short-range dependent Markovian models.

The other body of the literature (e.g. [5], [7], [18]) studies fairness issues by modeling Internet traffic at the flow level. The main objective is to show how the capacity of the network is shared among the different flows, or equivalently, to compute the response times of flows. Processor sharing queues [20] are used to model congested links in the network. In [5], an M/G/∞ model is proposed for the number of active flows on a non-congested backbone link. It coincides with a particular case of our model where all flows would have exactly the same rate. In [7], a multi-class processor sharing queue is used to compute the queue length and the packet loss probability in an Active Queue Management buffer crossed by TCP flows of different sizes. The average response time of a TCP flow is obtained. Note that all the above flow-based models make the assumption that flows arrive according to a homogeneous Poisson process.

Our model is different from the above works in that (i) it is designed for non congested links as those found in the backbone, (ii) it uses any flavor of flow definition to model the variation of the traffic, and (iii) it focuses on the variation of the traffic, a performance measure of particular interest for network engineering (i.e., provisioning, SLA definition, anomaly detection, etc.).

Date	Length	Avg. Link Utilization
Nov 8th, 2001	7h	243 Mbps
Nov 8th, 2001	10h	180 Mbps
Nov 8th, 2001	6h	262 Mbps
Nov 8th, 2001	39h 30m	26 Mbps
Sep 5th, 2001	10h	136 Mbps
Sep 5th, 2001	7h	187 Mbps
Sep 5th, 2001	16h	72 Mbps

TABLE I
SUMMARY OF OC-12 LINK TRACES

III. Measurement testbed

We consider data collected from OC-12 (622 Mbps) links on the Sprint IP backbone. The monitored links are over-provisioned so that the link utilization does not exceed 50% in the absence of link failures. The utilization is measured over relatively long time intervals, for example the 5 minutes period given by SNMP. In short, the infrastructure we use to collect packet traces consists of passive monitoring systems that tap optical links between access routers and backbone routers (see [15] for details on the monitoring infrastructure). Every packet on those links is timestamped and its first 44 bytes are recorded to disk.

In this paper, we present data from 7 different internal POP (Point-Of-Presence) links collected on September 5th and November 8th 2001 in three different POPs of the backbone. Table I provides a summary of the traces. The traces have different link utilizations (ranging from 26 Mbps to 262 Mbps), resulting in different trace lengths.

We divide each trace into 30 minutes intervals. We tried various intervals and we found that 30 minutes is a good compromise in term of (i) keeping the arrival process stationary, and (ii) giving enough points for the analysis of our model. We discuss later in more details the consequence of this analysis interval on our observations.

We apply the model to each interval and we validate its efficiency in computing the traffic. We focus on the first two moments of the total data rate, namely the mean and the variance. Considering the variance in addition to the mean allows a better characterization of backbone traffic. As we will see, the variability of the traffic on some links of the backbone can be as high as 30% compared to the

mean. The importance of the first two moments of the traffic in dimensioning backbone links will be illustrated in Section IX.

For each 30 minutes interval, we measure the coefficient of variation of the total rate ρ_R (standard deviation divided by the mean), and we compare it to the value given by the model. Our model only requires information on flows, which we derive from the traces (e.g., average arrival rate of flows).

In the measurements, we use two definitions of "flow":

(i) Flow defined by *5-tuple*, which is a stream of packets having the same source and destination IP addresses, same source and destination port numbers, and same protocol number.

(ii) Flow defined by *prefix*, which is a stream of packets having the same /24 destination address prefix (i.e., only the 24 most significative bits of the destination IP address are taken into account).

In both cases, the size of a flow is measured in bytes, while the duration is equal to the time difference between the first and the last packet of the flow. In order to identify the end of a flow, we use a fixed timeout of 60 seconds: if the timeout expires before recording any additional packet, the flow is considered completed. A flow made of only one packet is discarded (the duration would be zero), and that packet is not counted for the purpose of the mean and the variance of the measured total rate. Flows that belong to more than one 30 minutes interval are split over the intervals they overlap. We found that this artificial splitting affects only a small number of flows, as shown in Figure 1. The graph on the left-hand side shows the cumulative number of flows that arrive during one 30 minutes interval. We use the second definition of flow (i.e., /24 prefix) for this graph, since the splitting of flows has more impact with this definition than with the first one (durations of flows are longer in average with the second definition). The second graph is a zoom around 0 of the first one. The arrival rate remains pretty constant throughout the 30 minutes interval, except for the first 0.4 seconds, where we count only around 15,000 extra flows that are the continuation of flows started in the previous interval, out of a total of 680,000 flows. We consider therefore that the splitting of flows on these intervals has a nonzero, yet marginal effect on the arrival process, and in order to keep the model tractable, we do not correct for these effects.

As we mentioned in the Introduction, our model can operate with any definition of flow. The definitions we consider in this paper are no more than two examples of particular interest, corresponding to two different aggregation levels.

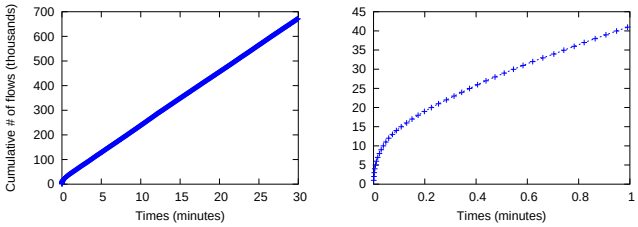

Fig. 1. Cumulative number of flows during one 30 minutes interval

IV. The model

In this section, we describe the model (Poisson shot-noise) used for data flows arriving on a backbone link. It is based on the following two assumptions.

Assumption 1: Flow arrivals follow a homogeneous Poisson process of finite rate λ.

This assumption can be relaxed to more general processes such as MAPs (Markov Arrival Processes) [1], or non homogeneous Poisson processes [6], but we will keep working with it for simplicity of the analysis. Poisson might be the right model if we consider recent findings by [2], [8] about the process of flow arrivals in the backbone of the Internet, where a large number of flows are multiplexed. It is shown in [8] that the distribution of flow inter-arrival times is very well approximated by a Weibull with a shape parameter smaller than 1, and that as the traffic intensity increases, flow inter-arrival times become independent, whereas the Weibull shape parameter gets close to 1. Thus, the flow arrival process tends to be in good agreement with a Poisson process. This limit is explained by well known results on the superposition of marked point processes. The Poisson property is also known to apply to aggregates at the session level [14], [22], [24]. Note that since our model does not depend on a particular definition of flow, one can group packets into sessions that have Poisson arrivals, and apply the model at the session level.

We computed the distribution and auto-correlation of the flow inter-arrival times on the collected traces. Indeed, we found that they are close to those of a homogeneous Poisson process having the same rate. We show the results for one 30 minutes interval in Figure 2. The other 30 minutes intervals provide similar results. This figure corresponds to the two definitions of flow. The graphs on the left-hand side show the quantile-quantile plot (qq-plot) of flow inter-arrival times, and those on the right-hand side show their coefficient of auto-correlation for different lags. The low level of correlation is clear from the graphs. The distribution of flow inter-arrival times still has a slightly heavier tail than exponential, that can be well modeled by a Weibull with shape parameter 0.96 in both figures.

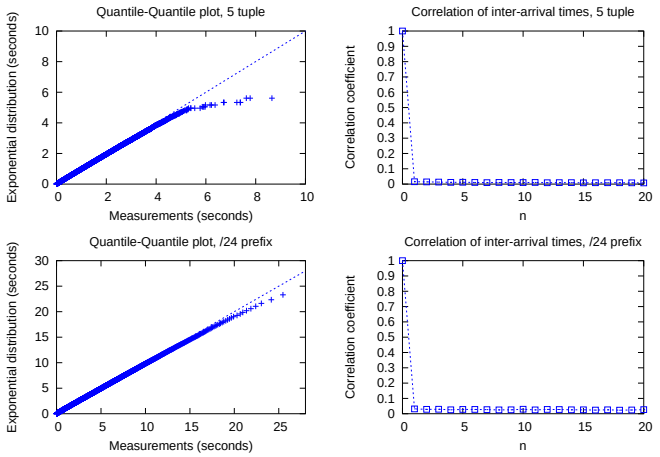

Fig. 2. Distribution and auto-correlation of inter-arrival times $\{T_{n+1} - T_n\}$

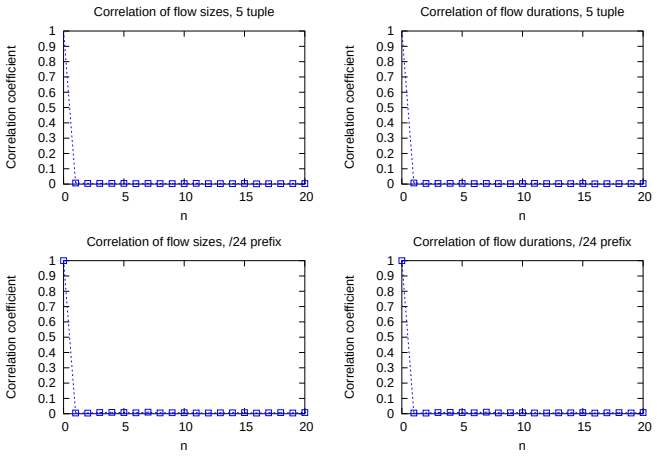

Fig. 3. Correlation of sequences $\{S_n\}$ and $\{D_n\}$

This heavy tail is of small importance for our model given the relatively small number of points that deviate from the diagonal. Although it is a deviation from our modeling assumptions, neglecting this heavy tail strongly simplifies the computations without impacting too much the model accuracy.

Denote by T_n, $n \in \mathbb{Z}$, the arrival time of the n-th flow, by S_n its size (e.g., in bits), and by D_n its duration (e.g., in seconds). A flow is called *active* at time t when $T_n \leq t \leq T_n + D_n$. Define $X_n(t - T_n)$ as the transmission rate of the n-th flow at time t (e.g., in bits/s), with $X_n(t-T_n)$ equal to zero for $t < T_n$ and for $t > (T_n + D_n)$. In other words, $X_n(t - T_n)$ is zero if flow n is not active at time t. We call $X_n(\cdot)$ the *flow rate function* or *shot*. $X_n(\cdot)$ depends on S_n, D_n and on the dynamics governing the flow rate. For example, for TCP flows, the dynamics of the flow rate is a function of the dynamics of the window size, which in turn is a function of the round-trip time of the TCP connection, and of the features of the packet loss process [1], [9], [12], [23]. Note that

$$\int_0^{D_n} X_n(u)\mathrm{d}u = S_n. \tag{1}$$

Our second assumption on $X_n(\cdot)$ is as follows.

Assumption 2: Flow rate functions are independent of each other and identically distributed.

The assumption on the independence of flow rate functions is based on the following facts: (i) The link we consider is a backbone link kept under-utilized by engineering rules. It does not therefore experience congestion, and so it does not introduce dependence among the flow rate functions. (ii) The flows sharing this link have a large number of different sources and destinations, and use many different routes before being multiplexed on the backbone link. The assumption of identical distribution can be relaxed by introducing multiple classes (based on transport protocol, flow size, or any other metric). We keep however a single class in this paper, hence $\{X_n(\cdot)\}$ are iid (independent and identically distributed). A direct consequence of Assumption 2 is that sequences $\{S_n\}$ and $\{D_n\}$ also form iid sequences, although for the same n, S_n and D_n are obviously correlated: the larger S_n, the larger D_n (in general). Finally, we assume that $\mathbb{E}[D_n]$ is finite.

We computed the auto-correlation of sequences $\{S_n\}$ and $\{D_n\}$ on our traces. We found indeed that these sequences exhibit little correlation. The result is illustrated in Figures 3, where we show the auto-correlation coefficients of the two sequences for one 30 minutes interval, using our two definitions of flow. The auto-correlation drops quickly to zero after lag-0.

Define $R(t)$ as the total rate of data (e.g., in bits/s) on the modeled link at time t. It is the result of the addition of the rates of the different flows. We can then write

$$R(t) = \sum_{n \in \mathbb{Z}} X_n(t - T_n). \tag{2}$$

This model is a *Poisson shot-noise process* [6], [13], where the term "shot" is synonymous here of "flow rate function". In the particular case where $X_n(t - T_n) = 1_{\{t \in [T_n, T_n + D_n]\}}$, that is, where shots are rectangles of height 1 and length D_n, the process (2) is the number of clients found at time t in an M/G/∞ queue [19], if clients are identified with flows. We will allow however for "shots" with a more general shape than a rectangle of height 1, and we will see that this is indeed essential to characterize the total data rate on backbone links.

Next, we look for the moments of the process $R(t)$ in the stationary regime. We always assume that we have reached the stationary regime, which exists for finite λ and $\mathbb{E}[D_n]$. We state a result for the Laplace Stieltjes Transform (LST) of $R(t)$, that allows to compute all

moments of $R(t)$, as well as its first order distribution. For the particular shapes of the shot presented in Figure 4, we will see that with only three parameters (λ, $\mathbb{E}[S_n]$ and $\mathbb{E}\left[S_n^2/D_n\right]$), our model is able to compute the average and the variation of the backbone traffic.

V. Performance analysis

A. LST and moments of the total rate

We state in this section the expression of the LST of $R(t)$, which we denote as $\tilde{R}(w) = \mathbb{E}\left[e^{-wR(t)}\right]$, $\mathrm{Re}(w) \geq 0$. We also give the expressions of the average and variance of $R(t)$, which we denote as $\mathbb{E}[R(t)]$ and V_R, respectively.

Let $N(t)$ be the number of active flows at time t. Assumptions 1 and 2 imply that the total data rate $R(t)$ at time t is the sum of a random number $N(t)$ of iid random variables which are the rates of active flows. This leads to the following expression of $\tilde{R}(w)$.

Theorem 1 ([4]): For $w \in \mathbb{C}$ and $\mathrm{Re}(w) \geq 0$, the LST of the total rate is

$$\tilde{R}(w) = \exp\left(\lambda \mathbb{E}\left[\int_0^{D_n} e^{-wX_n(u)}\mathrm{d}u\right] - \lambda \mathbb{E}[D_n]\right).$$

By differentiating with respect to w and then setting w to 0, the LST in Theorem 1 can give us all the moments of the total rate in the stationary regime. In particular, the two first moments are as follows:

Corollary 1: The average of the total rate is $\mathbb{E}[R(t)] = \lambda \mathbb{E}[S_n]$, its variance is $V_R = \lambda \mathbb{E}\left[\int_0^{D_n} X_n^2(u)\mathrm{d}u\right]$.

The mean and variance of the total rate are two important performance measures an ISP needs to know in order to properly dimension the links of its network. A backbone link has to be provisioned so as to absorb the average of the total rate as well as its variations. In contrast to the average, our model tells us that the variance of the total rate is a function of the durations of flows and their rate functions. This requires some assumptions (or more information) on the dynamics of flow rate. Next, we provide approximations of the variance of $R(t)$ for some particular flow rate functions.

B. Two particular shot shapes

Before moving to more general models, let us examine the two particular cases shown in Figure 4a and 4b.

1) Rectangular shots: First, we consider the case where the rate of a flow is constant and equal to S_n/D_n (which gives the rectangular shot of length D_n and height S_n/D_n of Figure 4a). Corollary 1 yields that the variance of $R(t)$ is equal to $V_R = \lambda \mathbb{E}\left[S_n^2/D_n\right]$.

The rectangular assumption is the simplest one; the only generalization from an M/G/∞ model is the height

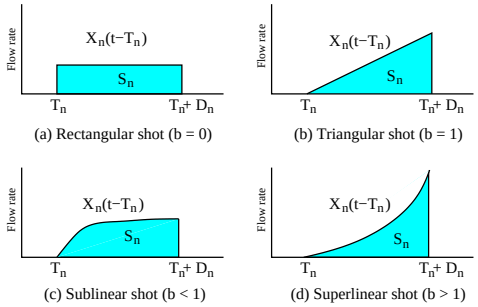

Fig. 4. Simple models for shots

of the "shot" which is now variable. With this assumption, we only capture the variation of the total rate caused by the variation of $N(t)$ and by the variation of the ratio S_n/D_n. It is easy to show that among all possible shot shapes, rectangular shots achieve the lowest variance V_R of the total rate [4, Theorem 3].

2) Triangular shots: Another assumption is to consider that the rate of a flow linearly increases with time (Figure 4b). This assumption is inspired from the dynamics of TCP transfers that form a large majority of the flows in IP backbones [15]. In Section VI-B, we will see that triangular shots are indeed representative of TCP flows under some conditions. For a flow of size S_n and of duration D_n, the rate is assumed to increase linearly from zero to $2S_n/D_n$, with a mean equal to S_n/D_n. At a time t between T_n and $T_n + D_n$, we can write $X_n(t-T_n) = (2S_n/D_n^2)(t-T_n)$. Corollary 1 yields that the variance of $R(t)$ is equal to $V_R = \frac{4\lambda}{3}\mathbb{E}\left[S_n^2/D_n\right]$. Again, the variance is a multiple of $\mathbb{E}\left[S_n^2/D_n\right]$. As expected, the variance is larger than in the rectangular case (by a multiplicative factor $4/3$).

VI. Determination of the shot

Once we have the shot function $X_n(\cdot)$, it is thus easy to compute the moments of the aggregate rate $R(t)$. But what shot function $X_n(\cdot)$ should we choose ? This key question is addressed in this section.

There are two different approaches to compute $X_n(\cdot)$. The first one consists in deriving it directly from measurements, and is developed in Subsection VI-A. The second one uses information from the protocol governing the flow dynamics, and is developed in Subsection VI-B.

A. Measurement-based derivation of shot shapes

The first method is based on *measurements*. It has the advantage of being protocol and application "agnostic", which preserves the generality of the model. The method consists in fitting a parametric model of the shot $X_n(\cdot) = x_\theta(\cdot)$, where $x_\theta(\cdot)$ is an *a priori* chosen function parameterized by a parameter vector θ,

which must satisfy the constraint (1). Vector θ is then computed to minimize some error functional between the experimental value of the distribution (or some moments of $R(t)$), and the value computed by Theorem 1. From now on, we restrict our attention to the variance of $R(t)$, and we compute $x_\theta(\cdot)$ so that

$$\hat{V}_R = \lambda \mathbb{E}\left[\int_0^{D_n} x_\theta^2(u)\mathrm{d}u\right], \tag{3}$$

where $\hat{V}_R$ is the actual empirical variance of the measured aggregate rate.

As we have two equations (1) and (3), we need therefore two parameters: $\theta = (a, b)$. A simple function is a power function $x_\theta(u) = au^b$, with $b \geq 0$, as illustrated in Figure 4. It includes, as particular cases, the rectangular ($b = 0$) and the triangular ($b = 1$) shots.

Solving (1) yields that $a = (b+1)S_n/D_n^{b+1}$, and plugging this value in (3) we get

$$\hat{V}_R = \lambda \frac{(b+1)^2}{2b+1}\mathbb{E}\left[\frac{S_n^2}{D_n}\right].$$

We deduce an estimate of b, based on the measurement of V_R (and clearly of λ and $\mathbb{E}\left[S_n^2/D_n\right]$). We find $b = \kappa - 1 + \sqrt{\kappa^2 - \kappa}$, with $\kappa = \hat{V}_R/(\lambda\mathbb{E}\left[S_n^2/D_n\right])$ (note that $\kappa \geq 1$). Of course, the introduction of a larger number of parameters allows to fit $x_\theta(\cdot)$ to more moments than simply V_R. We will use this expression of b in Section VIII.

B. Protocol-based derivation of shot shapes

In some cases, we can make use of *protocol information* to derive the shape of shots, instead of measurements as in the previous method. The typical example is TCP, whose dynamics shapes the flows and can be captured by analytical models (see [1], [18], [23] for an example of models for long-lived TCP flows). An advantage of this method is that it allows the simultaneous use of different shots for flows having different dynamics. Its drawback is the difficulty to model flows that do not have a well defined dynamics (e.g., uncontrolled UDP flows, flows defined by their address prefixes).

We illustrate this method by modeling the shot of a long-lived TCP flow. Even though long-lived TCP flows are currently not the majority among flows in the Internet, they are known to carry an important part of Internet traffic [15]. Moreover, this type of flows is expected to grow considerably with the arrival of data-greedy applications as Grid and Peer-to-Peer. We present results for the variance of backbone traffic V_R, which is given by Corollary 1.

We consider a fluid model for TCP inspired from [1] – other models, such as [12], could also be used. The transmission rate $X_n(t)$ is governed by the Additive-Increase Multiplicative-Decrease (AIMD) mechanism of TCP: between congestion events (we also call them loss events, since they are usually the times at which a packet loss is detected by the sender), the rate of TCP increases linearly with a slope A_n, which is inversely proportional to the square of the average round-trip time of the connection [1]. A_n is assumed to be time-constant, but is a random variable depending on (S_n, D_n). When a loss event appears, the rate of TCP is divided by two. Let $\overline{T}_l$ denote the time at which the l-th loss event occurred, and let τ_l be the time elapsed between the l-th and the $(l+1)$-th loss events, $\tau_l = \overline{T}_{l+1} - \overline{T}_l$. As in [1], we assume that the sequence of inter-loss times $\{\tau_l\}$ is a stationary, ergodic renewal process, which is independent of D_n and A_n.

As the duration of the nth flow is limited to D_n, we consider the extension of the TCP flow to all $t \in \mathbb{R}$, and denote $Y_n(t)$ its rate. We have thus $X_n(t) = Y_n(t)1_{\{0 \leq t \leq D_n\}}$, where $1_{\{A\}}$ is the indicator that A has occurred. To compute V_R we only need $X_n(t)$ for $0 \leq t \leq D_n$, where it coincides with $Y_n(t)$.

We assume that the AIMD mechanism is the only one to govern the dynamics of $Y_n(t)$, which is then stationary because of the assumptions above [1]. It thus obeys the following equation for all $t \in [\overline{T}_l, \overline{T}_{l+1})$:

$$Y_n(t) = Y_n(\overline{T}_l)/2 + A_n(t - \overline{T}_l), \tag{4}$$

where $Y_n(\overline{T}_l)$ is the rate of the nth TCP flow just before the l-th loss event (i.e. $Y_n(\overline{T}_l) = \lim_{t \to \overline{T}_l, t < \overline{T}_l} Y_n(t)$).

Using this fluid model, we find an expression that upper bounds the variance of Internet backbone traffic in the steady state V_R, and that can be safely used instead of the variance for network provisioning. This expression is stated in Theorem 2, where $\hat{\tau}^{(k)} = \mathbb{E}\left[\tau_l^k\right]/\mathbb{E}^k\left[\tau_l\right]$ denotes the k-th moment ($k \in \mathbb{N}$) of the inter-loss times, normalized by the mean time between loss events. Theorem 2 shows that the variance V_R is upper bounded by $\lambda\mathbb{E}\left[S_n^2/D_n\right]$ multiplied by a coefficient that only depends on the second and third normalized moments of times between loss events $\hat{\tau}^{(2)}$ and $\hat{\tau}^{(3)}$. The knowledge of the transmission rate slope A_n (which is a function of the round-trip time) is not needed in the result. This upper bound on the variance V_R in case of long-lived TCP flows has then the *same* expression as the one obtained with "power-b" shaped shots in Subsection VI-A, which confirms the importance of power-b shots in capturing the dynamics of backbone traffic.

Theorem 2: Assume that the sequence of inter-loss times is a stationary ergodic renewal process. The vari-

ance of the aggregate traffic satisfies

$$V_R \leq \lambda \frac{2+4\hat{\tau}^{(2)}+\hat{\tau}^{(3)}}{3\left(1+0.5\hat{\tau}^{(2)}\right)^2}\mathbb{E}\left[\frac{S_n^2}{D_n}\right]. \quad (5)$$

Proof: Pick any time $t \in \mathbb{R}$, and let l be the index of the last congestion event that occurred before t: $\overline{T}_l \leq t < \overline{T}_{l+1}$. Denote by $\mathbb{E}_d\left[Y_n^k(t)\right] = \mathbb{E}\left[Y_n^k(t)\,|\,D_n = d\right]$ the k-th moment of the transmission rate of the n-th TCP flow, given that $D_n = d$. The Palm inversion formula [1], [3] yields that

$$\mathbb{E}_d\left[Y_n^k(t)\right] = \frac{\mathbb{E}_d^0\left[\int_{\overline{T}_l}^{\overline{T}_{l+1}} Y_n^k(u)\mathrm{d}u\right]}{\tau^{(1)}}, \quad (6)$$

where $\tau^{(k)} = \mathbb{E}\left[\tau_l^k\right]$ is the (non-normalized) k-th moment of the times elapsed between loss events, and where the superscript 0 means that the expectation is taken conditionally to $\overline{T}_l \leq t < \overline{T}_{l+1}$. Inserting (4) in the numerator of the right-hand side of (6), we find that, for $k = 1$,

$$\mathbb{E}_d\left[Y_n(t)\right] = \frac{\mathbb{E}_d^0\left[Y_n(\overline{T}_l)\right]\tau^{(1)} + \mathbb{E}_d\left[A_n\right]\tau^{(2)}}{2\tau^{(1)}}. \quad (7)$$

and, for $k = 2$,

$$\mathbb{E}_d\left[Y_n^2(t)\right] = \frac{\frac{1}{4}\mathbb{E}_d^0\left[Y_n^2(\overline{T}_l)\right]\tau^{(1)} + \frac{1}{2}\mathbb{E}_d^0\left[Y_n(\overline{T}_l)\right]\mathbb{E}_d\left[A_n\right]\tau^{(2)} + \frac{1}{3}\mathbb{E}_d\left[A_n^2\right]\tau^{(3)}}{\tau^{(1)}} \quad (8)$$

Since $\mathbb{E}_d^0\left[Y_n(\overline{T}_{l+1})\right] = \mathbb{E}_d^0\left[Y_n(\overline{T}_l)\right] = \mathbb{E}_d\left[Y_n(\overline{T}_l)\right]$, setting $t = \overline{T}_{l+1}$ in (4) and taking expectations, we find that

$$\mathbb{E}_d^0\left[Y_n(\overline{T}_l)\right] = 2\mathbb{E}_d\left[A_n\right]\tau^{(1)}. \quad (9)$$

Similarly, elevating both sides of (4) to the square and taking expectations, and using (9), we find that

$$\mathbb{E}_d^0\left[Y_n^2(\overline{T}_l)\right] = \frac{4}{3}\left(2\left(\mathbb{E}_d\left[A_n\right]\tau^{(1)}\right)^2 + \mathbb{E}_d\left[A_n^2\right]\tau^{(2)}\right). \quad (10)$$

Inserting (9) in (7), we obtain

$$\mathbb{E}_d\left[Y_n(t)\right] = \mathbb{E}_d\left[A_n\right]\tau^{(1)}(1+0.5\hat{\tau}^{(2)}). \quad (11)$$

Now, taking expectations on both sides of (1) and remembering that $X_n(t) = Y_n(t)$ for $0 \leq t \leq d$, we obtain $\mathbb{E}_d\left[S_n\right] = \mathbb{E}_d\left[\int_0^{D_n} X_n(u)\mathrm{d}u\right] = \int_0^d \mathbb{E}_d\left[Y_n(u)\right]\mathrm{d}u = d\mathbb{E}_d\left[Y_n(t)\right]$, because $Y_n(t)$ is stationary. Therefore, we can write (11) as

$$\mathbb{E}_d\left[A_n\right] = \mathbb{E}_d\left[S_n\right]/(d\tau^{(1)}(1+0.5\hat{\tau}^{(2)})). \quad (12)$$

Likewise, inserting (10) and (9) in (8), we obtain

$$\mathbb{E}_d\left[Y_n^2(t)\right] = \frac{1}{3}\left(2\mathbb{E}_d^2\left[A_n\right]\left(\tau^{(1)}\right)^2 + \mathbb{E}_d\left[A_n^2\right]\tau^{(2)} + 3\mathbb{E}_d^2\left[A_n\right]\tau^{(2)} + \mathbb{E}_d\left[A_n^2\right]\tau^{(3)}/\tau^{(1)}\right). \quad (13)$$

Let us now compute the upper bound on V_R by conditioning on $A_n = a$. Denoting $\mathbb{E}_{ad}\left[\cdot\right]$ the operator of conditional expectation given $A_n = a$ and $D_n = d$, we obtain from (12) and (13) that

$$\mathbb{E}_{ad}\left[Y_n^2(t)\right] = \frac{2+4\hat{\tau}^{(2)}+\hat{\tau}^{(3)}}{3\left(1+0.5\hat{\tau}^{(2)}\right)^2}\frac{\mathbb{E}_{ad}^2\left[S_n\right]}{d^2}.$$

Consequently, Corollary 1 and the stationarity of $Y_n(t)$ imply that

$$\begin{aligned}
V_R &= \lambda\int \mathbb{E}_{ad}\left[\int_0^{D_n} X_n^2(u)\mathrm{d}u\right]\mathrm{d}P_{A_n,D_n}(a,d)\\
&= \lambda\int\left(\int_0^d \mathbb{E}_{ad}\left[Y_n^2(u)\right]\mathrm{d}u\right)\mathrm{d}P_{A_n,D_n}(a,d)\\
&= \lambda\int d\mathbb{E}_{ad}\left[Y_n^2(u)\right]\mathrm{d}P_{A_n,D_n}(a,d)\\
&= \lambda\frac{2+4\hat{\tau}^{(2)}+\hat{\tau}^{(3)}}{3\left(1+0.5\hat{\tau}^{(2)}\right)^2}\int\frac{\mathbb{E}_{ad}^2\left[S_n\right]}{d}\mathrm{d}P_{A_n,D_n}(a,d)\\
&\leq \lambda\frac{2+4\hat{\tau}^{(2)}+\hat{\tau}^{(3)}}{3\left(1+0.5\hat{\tau}^{(2)}\right)^2}\int\frac{\mathbb{E}_{ad}\left[S_n^2\right]}{d}\mathrm{d}P_{A_n,D_n}(a,d)\\
&= \lambda\frac{2+4\hat{\tau}^{(2)}+\hat{\tau}^{(3)}}{3\left(1+0.5\hat{\tau}^{(2)}\right)^2}\mathbb{E}\left[\frac{S_n^2}{D_n}\right]
\end{aligned}$$

where P_{A_n,D_n} is the joint probability measure of A_n and D_n.

□

This theorem enables us to link the power b used in the parametric shot model of Section VI-A with the burstiness of the congestion events. It is interesting to look at some particular sequences of congestion events, to see to which value of b they correspond.

(i) When times between congestion events are equal ($\hat{\tau}^{(i)} = 1$), the variance of backbone traffic V_R is upper bounded by $(28/27)\lambda\mathbb{E}\left[S_n^2/D_n\right]$. This is slightly larger than what we obtain with *rectangular* shots.

(ii) When congestion events follow a homogenous Poisson process ($\hat{\tau}^{(i)} = i!$), the variance of backbone traffic is upper bounded by $(4/3)\lambda\mathbb{E}\left[S_n^2/D_n\right]$, exactly the same variance we obtain with *triangular* shots.

(iii) Burstier congestion processes result in larger values of b.

VII. Practical use of the model

A. Moments of $R(t)$ and averaging interval

In reality, the total measured rate $\hat{R}(t)$ at a certain time t is computed by averaging and sampling the volume of data (e.g., number of bytes) that cross the backbone link during a short time interval δ around t:

$$\hat{R}(t) = \frac{1}{\delta}\int_{k\delta}^{(k+1)\delta} R(s)\mathrm{d}s,$$

with $t \in [k\delta, (k+1)\delta)$, $k \in \mathbb{Z}$. δ denotes the length of the averaging and sampling period. The measured rate $\hat{R}(t)$ appears thus as a piecewise constant function, with segments of length δ. It amounts to convolve the instantaneous rate $R(t)$ by a linear filter of impulse response $1_{\{0 \le t < \delta\}}$ before taking the samples. Except for the first one, the moments of $\hat{R}(t)$ depend on δ: the longer the averaging interval, the smoother the total rate (at least for non self-similar traffic). We can compute that the variance of $\hat{R}(t)$ (the measured variance) is

$$\hat{V}_R = \frac{2}{\delta} \int_0^{\delta} (1 - \tau/\delta) C_R(\tau) \mathrm{d}\tau, \tag{14}$$

with $C_R(\tau) = \mathbb{E}[R(t-\tau)R(t)] - \mathbb{E}\left[R^2(t)\right]$ being the auto-covariance function of the total rate $R(t)$. We give the expression of $C_R(\tau)$ in Theorem 2 in [4].

Since $C_R(\tau) \le V_R$, the above expression of $\hat{V}_R$ is always smaller than V_R. The scaling factor between V_R and $\hat{V}_R$ requires the knowledge of $C_R(\tau)$. Clearly, if $C_R(\tau)$ does not decrease too rapidly in $[0, \delta]$, both variances will remain close to each other. Consequently, we do not take into account the averaging of the data rate in the model, but we rather keep δ small so that $C_R(\tau)$ remains close to $C_R(0) = V_R$ in $[0, \delta]$. V_R can then be safely used as an approximation of $\hat{V}_R$, which models the variance of the measured samples of the total rate. Taking large values of δ amounts to smooth the traffic and hence to make the measured variance $\hat{V}_R$ sensibly smaller than V_R. Note that one can always compute $\hat{V}_R$ by plugging the expression of $C_R(\tau)$ given by Theorem 2 in [4].

Before using our model, an ISP has to choose a value δ of the averaging interval. It can be the longest busy period (i.e., period where the utilization of the link is 100%) allowed by the ISP. It is also the interval below which the ISP does not care about the congestion of the network, possibly because this short-term congestion is absorbed by the buffers at the inputs of links. If the chosen value δ is small enough so that the auto-covariance function $C_R(\tau)$ slowly decreases in $[0, \delta]$, V_R can be used by the ISP as an approximation of traffic variability (for network dimensioning issues), otherwise $\hat{V}_R$ has to be computed and used (using (14) and Theorem 2 in [4]). In what follows, we will choose as averaging interval the (average) round-trip time of flows (200 ms), since we know that most of the flows take more than one round-trip time to end. Our choice is also motivated by the fact that TCP flows update their transmission rates approximately once per round-trip time. Recall that the averaging interval is a parameter that can be set by the ISP to any other value than the round-trip time, depending on the maximum burstiness it tolerates at the inputs of the links of its backbone.

B. Complexity of the model

Our model requires few parameters to characterize the backbone traffic. The first two moments of the traffic can be computed with only three parameters: λ, $\mathbb{E}[S_n]$, and $\mathbb{E}\left[S_n^2/D_n\right]$.

In this paper, we compute the parameters of the model off-line. We infer their values from statistics on the processes $\{S_n\}$ and $\{D_n\}$. The computation is simple and it only requires an averaging over the different samples of the processes. An implementation of the model would require an online computation of these parameters with, for example, an Exponentially Weighted Moving Algorithm, such as the one used by TCP to estimate the average round-trip time.

We leave the problem of the online estimation of the parameters of our model for future research. Our main objective in this paper is to validate the model and to show its usefulness for provisioning and managing IP networks. Given that our model requires few parameters, we believe that it is simpler (in term of computation cost and implementability in an operational environment) than a packet level model that provides the same information about the traffic. The latter could however provide additional, more detailed information.

VIII. Experimental validation

In this section we validate our model using the traces collected on the Sprint IP backbone, and presented in Section III. We compare the real coefficient of variation of the total rate $\hat{\rho}_R = \sqrt{\hat{V}_R}/\mathbb{E}[R]$, with the results obtained from our model $\rho_R = \sqrt{\lambda \mathbb{E}\left[\int_0^{D_n} X_n^2(u)\mathrm{d}u\right]}/(\lambda \mathbb{E}[S_n])$, when the inputs of the model (i.e., flow arrival rate λ and the expectation of S_n^2/D_n) are directly derived from the traces. Samples of the total rate are computed using averaging intervals of 200 ms. This is comparable with the average round-trip time we measure on these links (Section VII-A).

Even if experimental data are in good agreement with Assumptions 1 and 2, the measurement process introduces two differences with the model of Section V. We already addressed these two differences.

(i) The first difference is the averaging and sampling of the measured rate at a periodicity of 200 ms, which will lead to an experimental value of variance $\hat{V}_R$ smaller than the variance of the instantaneous rate V_R, as explained in Section VII-A. We have indeed observed on experimental data that the longer the averaging interval, the smaller

$\hat{V}_R$. Therefore, we expect to find a few occurrences of an empirical value $\hat{V}_R$ smaller than the lower bound on V_R obtained with rectangular shots.

(ii) The second difference is the splitting of flows located on the boundaries of the 30 minutes intervals. As we explained in Section III, the number of these flows is very small compared to the total number of flows that arrive in the intervals, and the splitting has therefore a negligible impact.

These two sources of errors are unavoidable: the first one because traffic is packet-based and not fluid, so that the measurements must be averaged over intervals of some minimal length, and the second one because we need to divide the trace into intervals short enough to keep the arrival process stationary and to reduce the volume of data to manipulate.

A. Results

In this section we do not present results on the first moment of the total rate, since it is computed by our model and by measurements in exactly the same way. We only present results concerning the coefficient of variation of the traffic. All figures presented in this section are plotted using the log-log scale.

In Figure 5 we compare the coefficient of variation computed via measurements ($\hat{\rho}_R$) with that given by our model (ρ_R) with parabolic shots ($b = 2$). These results refer to the first definition of flow using the 5-tuple. Each point in the figure corresponds to a 30 minutes interval. A cross indicates that the average rate during that interval is below 50 Mbps; a triangle is used for those intervals with an average rate between 50 and 125 Mbps; the dots are used for rates above 125 Mbps. The x-axis shows the measured coefficient of variation of the total rate, while the y-axis shows the coefficient of variation given by the model. A point on the diagonal crossing the figure represents a perfect match between the model and the measurements. The two dashed lines identify the bounds for an error in the estimate of 20%. We notice a good match between the model and the measurements. Rectangular and Triangular shots (results not included for lack of space) often under-estimate the real coefficient of variation since they do not capture all the dynamics of flow rates.

The above figure shows three clusters of points, that can be easily distinguished. The interpretation is simple and is related to the fact that we are collecting traces on many diverse links, with three main different utilization levels (Section III). As we will explain in Section IX-.1, backbone traffic becomes smoother when the arrival rate of flows λ increases. An increase in the arrival rate

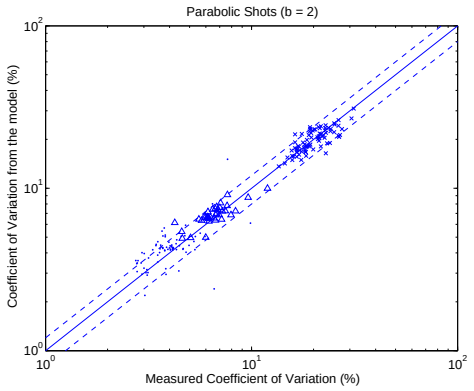

Fig. 5. Coefficient of variation of the total rate with parabolic shots and flows defined by the 5-tuple

of flows is the main responsible for the increase in the utilization among the links, since it is safe to assume that the average file size is the same on all links of the backbone (Corollary 1). Links with higher utilization (above 125 Mbps) exhibit very low variation, and, thus contribute to the first cluster of points at the bottom-left corner of the figure. Those links with a medium utilization (between 50 and 125 Mbps) are represented by the cluster in the middle. Finally, the links with the lowest utilization (below 50 Mbps) exhibit the highest traffic variability (around 30%), and yield the cluster of points on the right-hand side of the figure.

In Section VI-A, we explained how the optimal power b can be computed from a trace so that the variance of the total rate given by our model V_R matches that given by measurements $\hat{V}_R$. For the different 30 minutes traces, we compute this optimal power and we plot its histogram in Figure 6. The average value of b over all the traces is equal to 1.98, which means that parabolic shots are in average the most suited to model traffic when flows are defined by the 5-tuple (from variation point of view). We are currently working on the interpretation of the difference in the value of b among the traces. A possible reason could be the difference in file sizes: small files require a large value of b due to the slow start phase of TCP, and large files require a small value of b due to the slow window increase in TCP congestion avoidance mode.

Figure 7 provides the coefficient of variation for the second definition of flow based on destination address prefixes. We plot the case with rectangular shots ($b = 0$). The use of rectangular shots seems to be able to capture the variability of the traffic aggregate at the level of destination address prefixes. This is probably due to the fact that such a level of aggregation "dilutes" the impact of specific transport protocol mechanisms on the total rate. We also note that some points are above the diagonal, meaning the measured variance is smaller than the

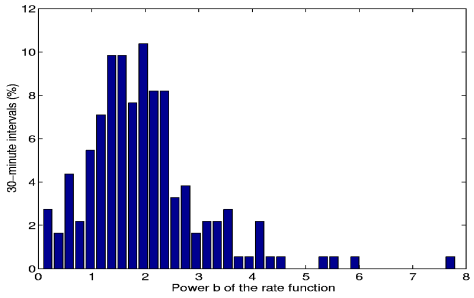

Fig. 6. Power b of flow rate functions with flows defined by the 5-tuple

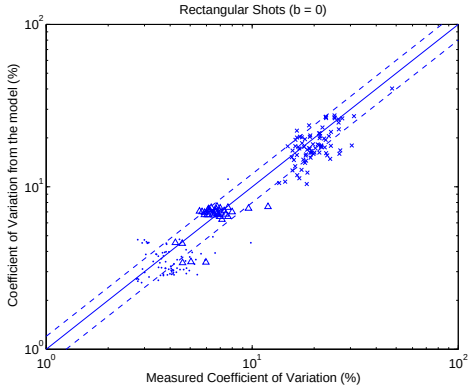

Fig. 7. Coefficient of variation of the total rate with rectangular shots and flows defined by destination address prefix

variance predicted by the model with rectangular shots, in an apparent disagreement with [4, Theorem 3]. This is due to the non-zero averaging interval, as explained in Section VII-A.

This result shows that our model can estimate the total rate and its variance independently of the protocol or application characteristics. The ability of defining a flow through the destination prefix greatly reduces the complexity of a possible implementation. Indeed, on our traces, the number of flows of which a router would need to keep track is reduced on average by one order of magnitude when using a /24 destination prefix. A straightforward extension to this flow definition would be the use of "routable" prefixes (i.e., prefixes present in the forwarding table of the router) to define flows. Such an extension would result in an additional decrease of the burden for the router given the level of flow aggregation (with /8 and /16 prefixes, for example) that could be achieved.

IX. Application of the Model to Network Dimensioning and Management

We discuss in this section some applications of our model to network dimensioning and management. The list is not exhaustive, but it is enough to highlight the role that such a model may have in the engineering of IP backbone networks.

Suppose that an ISP collects statistics on flow sizes, flow durations, and flow arrivals (for example with tools such as Cisco NetFlow). With this sole information, the ISP is able to compute the moments of the total rate. This way, the ISP would have more detailed information than that provided by SNMP (one of the problems of SNMP is that it does not capture traffic variation at short time scales).

The information on flows can be collected on the link we want to monitor. It can also be collected at the edges of the backbone. Combined with the routing information in the edge routers, this will give information on flows on each link of the backbone. Our model can then be used to compute the traffic on the links of the backbone, by only monitoring the edges.

The detailed information provided by our model on the traffic helps to dimension backbone links. Given the characteristics of flows composing the traffic, the links of the backbone network can be dimensioned so as to avoid congestion. Note that for a highly variable traffic, dimensioning the links of the backbone based only on the average utilization is not enough to avoid congestion. Traffic variability should be considered, which is allowed by our model. Rate variation at short time scales are very useful in the definition of the buffer size and in the evaluation of the maximum queuing delay. In the case we collect information on flows at the edges, our model can help in routing flows in the backbone, with the objective to optimize the utilization of the available resources.

Computing the traffic in the backbone using information on flows is not the only application of our model to network dimensioning and management. A key problem the operator faces is the planning of the upgrades of the backbone links, in order to maintain the absence of congestion. What is the impact on the link utilization caused by a change in the distribution of flow sizes, due for example to the arrival of a new application or the addition of a new big cluster of servers resulting in large transfer sizes? What is the impact on the link utilization caused by a change in flow durations, due for example to an increase in the number of users in the congested access networks, resulting in longer flow durations? What is the impact caused by a simultaneous change in flow sizes and durations, due for example to an upgrade of the access networks, resulting in shorter flow durations but larger file transfers? What is the impact on the traffic of a change in the shot shape $X_n(\cdot)$, which may follow a change in the application or in the transport protocol? The model presented in this paper can be used to answer these important questions.

We illustrate this application by the following two examples. The first example shows the impact of a

change in the flow arrival rate λ on the traffic, and hence on the dimensioning of the backbone. The second example shows the impact of the sizes and the durations of flows.

1) Impact of the flow arrival rate: Consider the case when the joint distribution of flow sizes and flow durations is stationary over long time intervals, and does not depend on the flow arrival rate[2]. Suppose that the ISP sets the bandwidth of its links to $\mathbb{E}[R(t)] + A(\epsilon)\sqrt{V_R}$, where $A(\epsilon)$ is the ϵ-quantile of the centered and normalized total rate $R(t)$, i.e., the value such $\mathbb{P}\left\{R(t) > \left(\mathbb{E}[R(t)] + A(\epsilon)\sqrt{V_R}\right)\right\} = \epsilon$, $0 < \epsilon < 1$. ϵ is the congestion probability. The moments of $R(t)$ in this expression of the bandwidth are given by our model (Corollary 1). For a large averaging interval, V_R needs to be corrected using (14). The function $A(\epsilon)$ can be computed using the Gaussian approximation[3], which gives for example $A(0.05) = 1.96$. When the arrival rate of flows increases, the bandwidth of the backbone links has to be increased as well, since the first and second moments of $R(t)$ increase with λ. However, while the first moment of $R(t)$ increases as λ, the standard deviation of $R(t)$ increases as $\sqrt{\lambda}$. This indicates that the coefficient of variation of $R(t)$ decreases as $1/\sqrt{\lambda}$. Concretely, this means that the traffic in the backbone becomes smoother and smoother when more and more flows are multiplexed. The consequence of this smoothing is that the ISP does not need to scale the bandwidth of its links linearly with λ. (S)He can gain in bandwidth by accounting for the smoothing of the traffic.

2) Impact of flow sizes and flow durations: We study in this section the impact of the sizes of flows $\{S_n\}$ and their durations $\{D_n\}$ on the first two moments of the traffic, and hence on the dimensioning of the backbone.

The average rate of the backbone traffic depends only on $\mathbb{E}[S_n]$ (Corollary 1). The study of the variance of the traffic is more complicated since the variance V_R depends on the shot shape, and on the joint distribution of $\{S_n\}$ and $\{D_n\}$ (Corollary 1). We focus on the "power-b" shots of the form $X(u) = au^b$, $b \geq 0$. As shown in Section VI-A, the variance of the traffic in presence of such shots only depends on $\mathbb{E}\left[S_n^2/D_n\right]$ (with a multiplicative factor function of the flow arrival rate λ and the power b). Section VI-B shows that this relationship also holds in case of long-lived TCP flows. For the same average flow size and the same average flow duration, the backbone traffic may have different variation if we consider different joint distributions of $\{S_n\}$ and $\{D_n\}$. To simplify the analysis of the variance, we consider the two extreme cases: (i) S_n and D_n are independent, and (ii) S_n and D_n are strongly positively correlated. These two cases provide respectively upper and lower bounds on the variance of the backbone traffic.

(i) When S_n and D_n are independent, the variance of the traffic V_R is proportional to $\mathbb{E}\left[S_n^2\right]\mathbb{E}[1/D_n]$. This value can be considered as an upper bound on the variance of the traffic in case of negative correlation between S_n^2 and $1/D_n$. We will assume that such a negative correlation holds, which seems a reasonable assumption since the larger the size of a flow, the longer in average its duration. We note here that V_R is proportional to the variance of S_n. V_R can be very large when the sizes of flows are heavy-tailed. Two sets of flow sizes having different variances result in different traffic variability, even if their averages are the same. The tail of D_n does not have an impact on the variance, since D_n is in the denominator, but for the very same reason, small values of D_n can lead V_R to be very large.

We check the correlation between S_n^2 and $1/D_n$ using our traces. The above upper bound is correct if these two random variables are always negatively correlated. For each 30 minutes trace, and using both definitions of flow (/24 prefix and 5-tuple), we compute the coefficient of correlation between S_n^2 and $1/D_n$. The results are plotted in Figure 8. All the traces present negative correlation coefficient, which validates our assumption. We notice in the figure the small value of the correlation coefficient, which is mostly due to the high level of multiplexing in the backbone. The variance of the traffic is then close to that given by the above upper bound.

(ii) The second case, which provides a lower bound on the variance of the traffic, corresponds to a strong

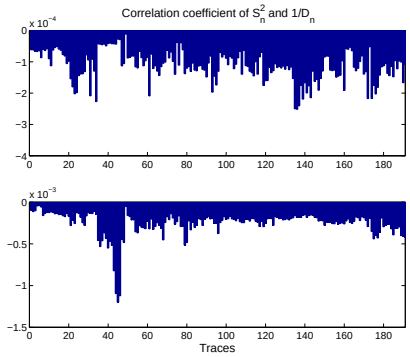

Fig. 8. The coefficient of correlation between S_n^2 and $1/D_n$ for 5-tuple (top) and /24 destination address prefix (bottom) definitions of flow, and for each 30 minutes long trace

[2]In the other case, a model has to be developed for the rest of the Internet, to evaluate the impact of a change in the arrival rate of flows on the joint distribution of flows sizes and flow durations. We will address this problem in a future research.

[3]Since the total rate is the resut of multiplexing of $N(t)$ flows of independent rates, the Central Limit Theorem tells us that the distribution of $R(t)$ tends to Gaussian at high load, which is typical of backbone links.

positive correlation between S_n and D_n. We suppose that these two variables are proportional to each other via a positive constant r, i.e., $S_n = rD_n$, $\forall n$. Note that the correlation coefficient of S_n and D_n is here equal to its maximum value 1.

The quantity r can be seen as the individual throughput of flows. There are many scenarios in which the throughput of a flow can be independent of its size. This is generally the case when the duration of the flow is long compared to its transient phase. In case of TCP, r can be the throughput imposed by the receiver advertised window. r can also be the throughput imposed by the available bandwidth in the network (i.e., Internet access via a slow modem line), or by the congestion control mechanisms of TCP. We refer to [25] for a discussion on the different possible meanings of r.

It is easy to see that a strong positive correlation between S_n and D_n provides indeed a lower bound on the variance of the traffic V_R. Applying Hölder's inequality to the product of the two random variables $S_n/\sqrt{D_n}$ and $\sqrt{D_n}$, we have that

$$\begin{aligned}\mathbb{E}^2[S_n] &= \mathbb{E}^2\left[\frac{S_n}{\sqrt{D_n}}\sqrt{D_n}\right] \\ &\leq \mathbb{E}\left[\left(\frac{S_n}{\sqrt{D_n}}\right)^2\right]\mathbb{E}\left[\sqrt{D_n}^2\right] = \mathbb{E}\left[\frac{S_n^2}{D_n}\right]\mathbb{E}[D_n],\end{aligned}$$

from which we obtain the following lower bound on $\mathbb{E}\left[S_n^2/D_n\right]$ (and therefore on V_R):

$$\mathbb{E}\left[\frac{S_n^2}{D_n}\right] \geq \frac{\mathbb{E}^2[S_n]}{\mathbb{E}[D_n]}.$$

The bound is reached when $S_n = rD_n$ for some $r > 0$ (in which case S_n and D_n have a maximal correlation), and is equal to $\mathbb{E}\left[S_n^2/D_n\right] = r\mathbb{E}[S_n]$. Contrary to the case where S_n and D_n were independent, the variance V_R is now only sensitive to the average flow size and to the individual throughput of flows r. We directly compute that it is equal to $(b+1)^2/(2b+1)r\mathbb{E}[R(t)]$ for power-b shots. This means that when $S_n = rD_n$, the variance changes only if either r or the average traffic $\mathbb{E}[R(t)]$ does. For example, when r increases (due for example to an upgrade of user access lines or to a change in network protocols), the coefficient of variation of the total rate increases as $\sqrt{r}$, even though the average utilization is the same (the traffic in the backbone becomes more variable). The increase in the coefficient of variation is less important than the increase in r due to the statistical multiplexing of flows in the backbone. The ISP can then use this result to anticipate the increase in traffic variability, and to appropriately upgrade the links of its backbone.

To illustrate the impact that the correlation between S_n and D_n can have on the variance of the traffic V_R, we consider the following example, where S_n and D_n are generated from Pareto distributions, but with same average values as those obtained from the traces. Denote by $\overline{S}$ (resp. $\overline{D}$) the average size (resp. the average duration) of flows obtained from measurements. Our idea is to control the correlation between S_n and D_n, while keeping $\mathbb{E}[S_n] = \overline{S}$ and $\mathbb{E}[D_n] = \overline{D}$. This control is not possible without the following artificial construction of flow sizes and durations.

A Pareto random variable V has a Cumulative Distribution Function $\mathbb{P}\{V \leq v\} = 1 - (v/a)^{-\beta}$ [17]. $a > 0$ is the starting point of the variable and $\beta > 1$ its shape parameter. The mean of a Pareto random variable is equal to $\mathbb{E}[V] = a\beta/(\beta-1)$. The variance of a Pareto random variable increases when its shape parameter β decreases, and becomes infinite when $\beta \leq 2$. The Pareto random variable is said to be heavy-tailed, since its tail decreases polynomially rather than exponentially. This variable is often used to model the heavy-tailed nature of the distributions of flow sizes and flow durations in the Internet (see [2], [11], [24] for examples).

First, we assume that the marginal distribution of S_n is Pareto, with shape parameter β_S and of average $\overline{S}$. We consider two values for β_S: 1.5 and 2.5. We define D_n as

$$D_n = w\frac{\overline{D}}{\overline{S}}S_n + (1-w)V_n, \tag{15}$$

where V_n is a Pareto random variable, with shape parameter β_D and of average $\overline{D}$, independent of S_n, and where $w \in [0,1]$. We give two values to β_D: 1.5 and 2.5. The coefficient w is used to vary the correlation between S_n and D_n; when $w = 0$, both variables are independent Pareto variables; when $w = 1$, both variables are maximally correlated. Note that the average value of D_n generated according to (15) is equal to $\overline{D}$. If β_D and β_S are larger than 2, we can compute that

$$w = \frac{COV[D_n, S_n]\overline{S}}{VAR[S_n]\overline{D}}. \tag{16}$$

Second, we give S_n the values we measure on our traces, while generating D_n according to (15). V_n is still a Pareto random variable, with shape parameter β_D and of average $\overline{D}$, independent of S_n.

We plot the variance V_R as a function of w for different values of β_S, β_D, $\overline{S}$ and $\overline{D}$. We consider rectangular shots ($b = 0$), which yields $V_R = \lambda\mathbb{E}\left[S_n^2/D_n\right]$. The plots are shown in Figure 9. The value of the flow arrival rate λ is computed from the traces. Figure 9 shows the plots obtained when both S_n and V_n are generated from Pareto distributions, as well as the plots obtained when only V_n is generated from a Pareto distribution, while S_n is given real flow size values. We remark that

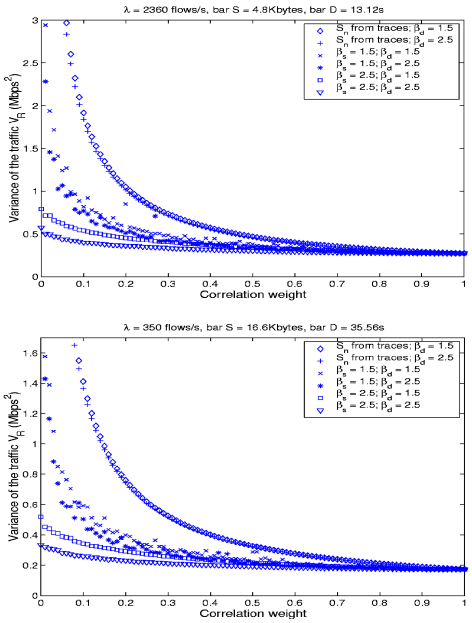

Fig. 9. Variance of the traffic vs. weight w representing the correlation between S_n and D_n. Top: 5-tuple definition of flow. Bottom: /24 prefix definition of flow

the variance V_R (proportional to $\mathbb{E}\left[S_n^2/D_n\right]$) decreases when S_n and D_n become correlated. For $w \simeq 1$ (strong correlation), V_R is insensitive to the marginal distributions of S_n and D_n, and only sensitive to their averages. For $w \ll 1$ (weak correlation), V_R is sensitive to the marginal distributions of S_n and D_n. The heavier the tail of S_n, the larger the variance of the traffic. Our traces indicate that on a backbone link, w is usually small (weak correlation between S_n and D_n), given the high level of multiplexing of flows in the backbone. For the traces considered in Figure 9, the coefficient w computed according to (16) (using the real sizes and real durations of flows) is equal to 0.019 and 0.034, respectively. We also remark in Figure 9 that V_R increases when β_D decreases, for the simple reason that with a small value of β_D, the realization of D_n will sometimes take very small values. The correlation between S_n and D_n is then an important factor impacting the variance V_R. Depending on their correlation, the marginal distributions of S_n and D_n have thus a very different influence on traffic variability, and hence on network dimensioning.

X. Conclusions

We proposed a traffic model for uncongested backbone links that is simple enough to be used in network operation and engineering. The model relies on Poisson shot-noise. With only 3 parameters (λ, arrival rate of flows, $\mathbb{E}\left[S_n\right]$, average size of a flow, and $\mathbb{E}\left[S_n^2/D_n\right]$, average value of the ratio of the square of a flow size and its duration), the model is able to find good approximations for the average traffic on a backbone link and for its variations at short timescales. The model is designed to be general so that it can be easily used without any constraint on the definition of flows, nor on the application or the transport protocol.

We are working on various extensions of our work. We state in [4] a result for the auto-covariance function of the total rate. Using this result, we are investigating the correlation of Internet traffic and its relation with the flow arrival process, the shot shape, and the distributions of flow sizes and flow durations. We are also studying the gain of introducing classes of flows with a different shot for each class. This will solve the problem when the flow rate functions do not have the same distribution. Finally, we are evaluating the worthiness of considering more complex flow arrival processes than Poisson. The challenge is to improve our evaluation of the traffic without much increasing the complexity of the model.

XI. Acknowledgments

We would like to thank the guest editor and the anonymous reviewers for their valuable comments.

References

[1] E. Altman, K. Avratchenkov, C. Barakat, "A stochastic model for TCP/IP with stationary random losses", *ACM SIGCOMM*, September 2000.

[2] S. Ata, M. Murata, H. Miyahara, "Analysis of network traffic and its application to design of high-speed routers", *IEICE Transactions on Information and Systems*, vol. E83-D, pp. 988-995, May 2000.

[3] F. Baccelli and P. Brémaud, "Elements of queueing theory: Palm-Martingale calculus and stochastic recurrences", *Springer-Verlag*, 1994.

[4] C. Barakat, P. Thiran, G. Iannaccone, C. Diot, P. Owezarski, "A flow-based model for Internet backbone traffic", *ACM SIGCOMM Internet Measurement Workshop*, November 2002.

[5] S. Ben Fredj, T. Bonald, A. Proutiere, G. Regnie, J. Roberts, "Statistical Bandwidth Sharing: A Study of Congestion at Flow Level", *ACM SIGCOMM*, August 2001.

[6] P. Brémaud, L. Massoulié, "Power spectra of general shot noises and Hawkes point processes with a random excitation", *Journal of Applied Probability*, to appear.

[7] T. Bu, D. Towsley, "Fixed Point Approximation for TCP behavior in an AQM Network", *ACM SIGMETRICS*, June 2001.

[8] J. Cao, W.S. Cleveland, D. Lin, D.X. Son, "On the Nonstationarity of Internet Traffic", *ACM SIGMETRICS*, June 2001.

[9] N. Cardwell, S. Savage, T. Anderson, "Modeling TCP Latency", *IEEE INFOCOM*, March 2000.

[10] J. Case, M. Fedor, M. Schoffstall, J. Davin, "A Simple Network Management Protocol (SNMP)", RFC 1157, May 1990.

[11] M. Crovella, A. Bestavros, "Self-Similarity in World Wide Web Traffic: Evidence and Possible Causes", *IEEE/ACM Transactions on Networking*, vol. 5, no. 6, pp. 835-846, Dec. 1997.

[12] V. Dumas, F. Guillemin and P. Robert , "A Markovian analysis of AIMD algorithms", *Advances in Applied Probability*, vol. 34, no. 1, pp. 85-111, 2002.

[13] D. Daley, D. Vere-Jones, "An introduction to the theory of point processes", *Springer-Verlag*, 1988.

[14] A. Feldmann, "Characteristics of TCP connection arrivals," in Self-Similar Network Traffic and Performance Evaluation (K. Park and W. Willinger, eds.), John Wiley, 2000.
[15] C. Fraleigh, S. Moon, C. Diot, B. Lyles, F. Tobagi, "Packet-Level Traffic Measurements from a Tier-1 IP Backbone", Sprint ATL Technical Report TR01-ATL-110101, November 2001.
[16] S. Haykin, "Modern filters", *Macmillan publishing company*, 1989.
[17] R. Jain, "The art of computer systems performance analysis", *Wiley*, 1991.
[18] A.A. Kherani, A. Kumar, "Performance Analysis of TCP with Nonpersistent Sessions", *Workshop on Modeling of Flow and Congestion Control*, September 2000.
[19] L. Kleinrock, "Queueing Systems, Vol. I: Theory", *Wiley*, 1975.
[20] L. Kleinrock, "Queueing Systems, Vol. II: Computer Applications", *Wiley*, 1976.
[21] W. Leland, M. Taqq, W. Willinger, D. Wilson, "On the self-similar nature of Ethernet traffic", *ACM SIGCOMM*, September 1993.
[22] C. Nuzman, I. Saniee, W. Sweldens, A. Weiss, "A Compound Model for TCP Connection Arrivals", *ITC workshop*, September 2000.
[23] J. Padhye, V. Firoiu, D. Towsley, J. Kurose, "Modeling TCP Throughput: a Simple Model and its Empirical Validation", *ACM SIGCOMM*, September 1998.
[24] V. Paxson, S. Floyd, "Wide-Area Traffic: The Failure of Poisson Modeling", *IEEE/ACM Transactions on Networking*, vol. 3, no. 3, pp. 226-244, June 1995.
[25] Y. Zhang, L. Breslau, V. Paxson, S. Shenker (ICSI), "On the Characteristics and Origins of Internet Flow Rates", *ACM SIGCOMM*, August 2002.

PLACE PHOTO HERE

Chadi Barakat (cbarakat@sophia.inria.fr) is a permanent research scientist in the PLANETE research group at INRIA - Sophia Antipolis since March 2002. In July 1997, he got his Electrical and Electronics engineering degree from the Lebanese University of Beirut. In June 1998, he got the DEA degree in Networking and Distributed Systems from the University of Nice - Sophia Antipolis, France. After the DEA, he joined the MISTRAL research group at INRIA - Sophia Antipolis to prepare a Ph.D. in Networking. He received his Ph.D. degree in April 2001 and after that, he joined EPFL-Lausanne for a post-doctoral position of ten months. His main research interests are congestion and error control in computer networks, the TCP protocol, voice over IP, Internet measurement and traffic analysis, and performance evaluation of communication protocols.

PLACE PHOTO HERE

Patrick Thiran (Patrick.Thiran@epfl.ch) received the electrical engineering degree from the Universit Catholique de Louvain, Louvain-la-Neuve, Belgium, in 1989, the M.S. degree in electrical engineering from the University of California at Berkeley, USA, in 1990, and the PhD degree from the Swiss Federal Institute of Technology at Lausanne (EPFL), in 1996. He became a professor at EPFL in 1998, and was on leave with Sprintlabs, Burlingame, CA, in 2000-01. His research interests are in communication networks and dynamical systems.

PLACE PHOTO HERE

Gianluca Iannaccone (gianluca@sprintlabs.com) received his B.S. and M.S. degree in Computer Engineering from the University of Pisa, Italy in 1998. He received a Ph.D. degree in computer engineering from the University of Pisa in 2002. From 2000 to 2001, he was a student visitor at Sprint Advanced Techonology Laboratories in Burlingame, California. He joined Sprint as a research scientist in October 2001. His main research interest are network performance measurements, inference methods for packet loss and survivability of IP networks.

PLACE PHOTO HERE

Christophe Diot (cdiot@sprintlabs.com) received a Ph.D. degree in Computer Science from INP Grenoble in 1991. From 1993 to 1998, he was a research scientist at INRIA Sophia Antipolis, working on new Internet architecture and protocols. From 1998 to 2003, he was in charge of the IP research team at Sprint Advanced Technology Labs. Diot recently moved to INTEL research in Cambridge, UK. His current interest is measurement techniques and Internet architecture.

PLACE PHOTO HERE

Philippe Owezarski (owe@laas.fr) is a full time researcher of CNRS (the French center for scientific research), working at LAAS (Laboratory for Analysis and Architecture of Systems), in Toulouse, France. He got a PhD in computer science in 1996 from Paul Sabatier University, Toulouse III. His main interests deal with high speed and multimedia networking and in particular with transport protocols, Quality of Service in the Internet and monitoring of IP networks, focusing especially on actual TCP flows analysis. Philippe Owezarski is one of the main contributor of a monitoring project in France METROPOLIS and leads a French steering group on IP networks monitoring.