

A Flow- and Packet-level Model of the Internet

Ramji Venkataramanan
Dept. of Electrical Engineering
Yale University, USA

Email: ramji.venkataramanan@yale.edu

Min-Wook Jeong, Balaji Prabhakar
Dept. of Electrical Engineering
Stanford University, USA

Email: {minwook, balaji}@stanford.edu

Abstract—Several models of packet-switched networks have been proposed in the literature for analyzing network performance. There are two main types of models, called the packet-level and the flow-level model, each making different assumptions about the operation of the network and providing different insights about its performance. While packet-level models are remarkably accurate, we find, from Internet packet traces and network simulations, that current flow-level models do not accurately model the Internet. This leads us to a new flow-level model that is more accurate and which, moreover, dovetails with a natural packet-level model to provide a complete description of the network. We validate our model via simulations and use it to provide a preliminary analysis of the Internet.

I. INTRODUCTION

It is important for network engineers to have good models to predict the performance of large networks like the Internet since the volume of traffic and the size of these networks make simulation impractical. Typical quantities of interest include flow transfer durations, throughput, queueing delay, queue-size distributions at the buffers etc.

Figure 1 is a high-level depiction of a section of the Internet. Flows arrive at the edge of the network and send their packets into the core through access links. Each packet is routed through a series of routers (where it may be buffered and encounter queueing delays) until it reaches its destination. An end-to-end transport protocol governs the manner in which packets are injected into the core. TCP is the most commonly used transport protocol and regulates the sending rate of sources through a well-studied feedback congestion control mechanism. Each router in the core of the network sees an aggregate of packets sent by several flows, and is ignorant of the sending rate and other characteristics of individual flows.

The models that have been proposed for the Internet can be classified broadly as either packet-level or flow-level models. Packet-level models such as those in [1]–[3], consider a fixed number of ‘long-lived’ flows, each with an infinite stream of packets. These packets are injected into the network according to a process that is governed by TCP’s congestion control mechanism. The packets are often approximated as fluid units to explicitly model the relationships between round-trip times, packet drop probabilities and rate-allocation among the flows. Consequently, they capture the role of buffers as well as queueing and propagation delays. Packet-level models are useful to estimate quantities such as queue sizes at buffers, and throughput. They have been heavily used to study the stability of congestion control algorithms (see [4], for example). Since only long-lived flows are considered, packet-level models are

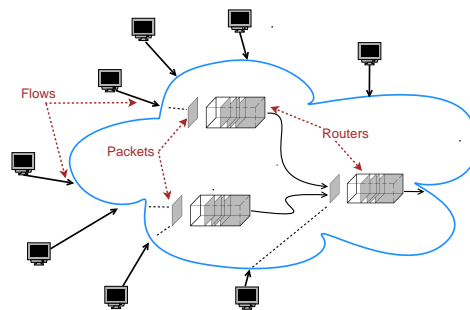


Fig. 1: A high-level depiction of a network

static and do not capture flow-level dynamics such as flow durations or the number of active flows.

In flow-level models, such as those in [5], [6], flows are treated as finite-sized units of data, whose sizes are random and specified by a probability distribution. Flows arrive at the network according to an arrival process with a given distribution. The models aim to capture flow-level dynamics such as flow transfer durations and the number of active flows resulting from different flow bandwidth allocation procedures. It is assumed that there is an underlying mechanism that effects ‘statistical bandwidth sharing’, i.e., the rate of each flow is adjusted *instantaneously* based on the available bandwidth on its path and the number of flows that it shares the bandwidth with at that instant. At each instant of time, a flow transmits data through the network according to the rate allotted to it. When a flow has been served completely, it departs. Flow-level models do not incorporate any packet-level details such as queueing delays at routers or propagation delays. Such details are abstracted away under the assumption of a ‘time-scale separation’: the time-scale of the flow-level dynamics (flow duration) is assumed to be much larger than the time-scale of packet-level dynamics captured in models such as [1], [2].

Using a flow-level model, we would like to answer questions like: What are the number of active flows? What are the flow durations? What do the queues at the buffers look like? In this paper, we first investigate existing flow-level models, and compare their predictions with data from network simulations and internet traces. We find that they do not give accurate estimates of flow durations and the number of active flows, especially when link speeds are high. This is because in existing models, flows arrive and deliver data continuously for their whole duration, ignoring important effects created

by the congestion control protocol, queueing delays etc. We then propose a new flow-level model that is more accurate and which, moreover, dovetails with a natural packet-level model to provide a complete description of the network. The main contributions of this paper are as follows.

- In Section II, we examine the Bonald-Massoulié flow-level model from [5], and explain why it does not accurately capture certain key aspects of flow transfer in the Internet. We present network simulation and Internet trace analysis to illustrate these shortcomings.
- In Section III, we propose a flow- and packet-level model that is more accurate and captures the essential packet-level features of the network as well.
- We provide some preliminary analysis on the new model in Section IV and finally in Section V, validate our model by characterizing the aggregate packet traffic in the core of the Internet.

We note that the approach of packet-level models has been extended to estimate durations of short-lived TCP flows in [7], [8]. We also mention that a joint flow-level and packet-level stochastic network model, different from the one proposed here, was proposed recently in [9].

II. THE BONALD-MASSOULIE MODEL

The Bonald-Massoulié (B-M) flow-level model is one in which all the active flows on a link instantaneously share the link bandwidth according to a specified fairness criterion. For concreteness, we assume that the sharing takes place according to the max-min fairness criterion [10]. The specific details of the sharing mechanism are abstracted away. For illustration, first consider a single bottleneck link of capacity $N\mu$ shared by flows arriving at rate $N\lambda$. Each flow has size S , drawn according to a distribution with mean $E[S]$. The queue is stable if the load $\rho = \frac{N\lambda \cdot E[S]}{N\mu}$ is less than one.

The B-M model for this network corresponds to a processor sharing queue, where all the active flows share the link bandwidth equally. Hence, if the flows arrive according to a Poisson process, the number of active flows in the system is geometrically distributed with parameter ρ [11]. Thus the average number of active flows is $\frac{\rho}{1-\rho}$, and using Little's law, one can show that the duration of a flow with size S is $\frac{S}{N\mu(1-\rho)}$. In other words, as the link speed and arrival rate increase proportional to N such that the load is fixed, the network just speeds up, and flow durations decrease as $1/N$, while the average number of active flows remains the same.

Let us compare this with TCP flows with the same arrival characteristics sharing the bottleneck link. Due to the slow-start mechanism of TCP, the duration of a flow with S packets is at least $\lceil \log_2 S \rceil$ round-trip times. Since the propagation delay (which does not depend on N) is a lower bound on the round-trip time, this implies the duration of a TCP flow of size S is an $O(1)$ quantity, i.e., it does not go to zero, regardless of how large N is. In other words, the B-M flow-level model is too efficient- the flow durations for large N are not accurate since the model does not account for the inefficiency in the network created by the physical constraints of the protocol, viz., round-trip times.

Next consider the number of active flows in the network. As explained above, the average number of active flows in the B-M model is $\frac{\rho}{1-\rho}$, an $O(1)$ quantity. In contrast, in a TCP network, flows arrive at rate $N\lambda$ and stay in the system for $O(1)$ time, so there are $\Theta(N)$ active flows in the system at any given time. We now illustrate this by simulating a network with different routes and multiple bottlenecks.

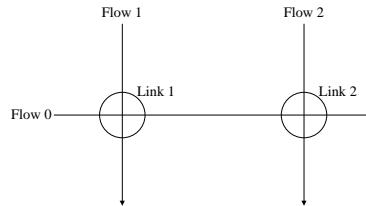


Fig. 2: Two link network

Multiple Bottlenecks: Consider the network shown in Figure 2. There are three types of flows labeled 0, 1, and 2, with arrival rates $0.05N$, $0.04N$ and $0.04N$ flows/s, respectively. Type 0 flows pass through both link 1 and link 2, while type 1 flows go through links 1 only, and type 2 flows go through link 2 only. The size of each flow is drawn iid according to a heavy-tailed distribution with mean 20 packets. The service rate of each link is $2N$ pkt/s, hence the load on each link is 90%. This example is used to model a network where the local and long distance traffic coexist.

In the B-M model, the instantaneous bandwidth allocated to each active flow is determined by the max-min fairness criterion. Thus the bandwidth allocation changes each time a flow arrives or departs the network. The first two rows of Table I show the average number of active flows of type 0 when the network of Figure 2 was simulated with instantaneous bandwidth sharing. We see that the number of active flows is roughly the same for both $N = 1$ and $N = 1000$.

The next two rows of Table I show the number of active flows of type 0 for the same network configuration when the flows use TCP rather than instantaneous bandwidth sharing. (This was simulated using OMNET++ [12], with $N = 1$ corresponding to a link speed of 10Mbps.) The last two rows show the number of active flows in two internet traces - the LBL trace from 1994 (10 Mbps link) used in [13], and a recent trace from a 10Gbps link obtained from [14]. The ratio of the number of active flows in the 10Gbps link to the 10Mbps link is 155.1 for the simulation and 437.0 for the Internet traces. For the 10Mbps and 10Gbps Internet traces, the link utilizations are 3.51% and 26.04%, and the average packet sizes are 176.4B and 578.9B, respectively. Taking these two factors into account, the number of active flows is found to be proportional to the link speed. Thus the number of active flows in an Internet backbone link grows with the link speed, a feature not captured by the B-M model.

In a real network, there is also a limit on the maximum rate at which a flow can send its packets. This is reflected in the maximum receive window size imposed on any TCP connection. There are several reasons for this restriction. First, the constraints of the hardware, operating system and the

TABLE I: Average number of active flows

| Setup | Number of active flows |
|---|------------------------|
| Bonald-Massoulié ($N = 1$) | 18.7 |
| Bonald-Massoulié ($N = 1000$) | 17.5 |
| TCP (OMNeT++, 10Mbps link speed) | 72.3 |
| TCP (OMNeT++, 10Gbps link speed) | 11216.8 |
| TCP (Internet trace, 10Mbps router, 1994) | 327.7 |
| TCP (Internet trace, 10Gbps router, 2009) | 143192.4 |

application at the sender as well as the receiver impose an upper limit on the transmission rate of a flow. Secondly, if the TCP window grows very large, there are a large number of unacknowledged packets at any given time. This leads to a huge number of retransmissions when there is network congestion, resulting in throughput degradation. (In the OMNET++ simulation of Table I, the receive window was set to 64 packets.) Thus the access rate of a flow is limited in practice, even with plenty of link bandwidth available. To summarize, the B-M model has two main shortcomings:

- It ignores the closed control loop of the protocol, and assumes data can be sent continuously by a flow once it arrives at the network. This implies flow durations can get arbitrarily small as link speeds increase, contradicting the fact that a flow of size S packets is active for a minimum of $\log_2 S$ round-trip times.
- It does not account for the fact that access speed of a flow is limited by the maximum window size of TCP and cannot grow arbitrarily large. This imposes another lower bound on the flow duration of a large flow.

Hence the B-M model does not capture the role of round-trip times and queueing in the network. It is also difficult to use the model to make predictions on large, unstructured network topologies. We now propose a flow- and packet-model that attempts to address these shortcomings.

III. THE FLOW- AND PACKET-LEVEL MODEL

Figure 3 shows the network of Figure 1 with flows arriving at backbone link i at the rate of $N\lambda_i$, $i = 1, 2, 3$. Link i is represented by a single-server queue with service rate $N\mu_i$ packets/s. N is the scaling parameter to capture the growth in arrival rates and link speeds. Each flow at the edge is assigned a separate server (representing the access link) which which serves it packet by packet. The served packets enter the queue corresponding to the backbone link. At the access link, different packets of a flow may receive different rates of service, depending on the congestion encountered at the downstream queue by the previous packets. The mechanism to effect this is described in the sequel.

Let us first consider a simple network with a single backbone link. Suppose flows arrive at the edge according to a Poisson process of rate $N\lambda$, and share the downstream link of capacity $N\mu$. Each flow consists of a random number of packets S , drawn independently according to a known heavy-tailed distribution P_S with mean $E[S]$. The sizes of the individual packets are assumed to be iid according a known distribution. The system is depicted in Figure 4.

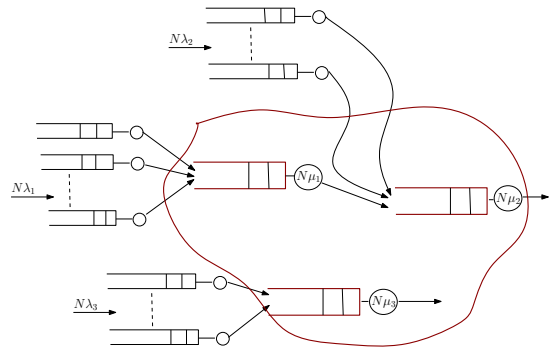


Fig. 3: Flow- and Packet-level model for network of Figure 1

We now specify the mechanism which controls the rate of service the packets of a flow receive at the access link. Recall that the instantaneous rate (or congestion window) of a TCP flow depends on the level of congestion encountered by the previous packets of the flow. This signalling mechanism is modeled as follows. The service rate of a flow at the edge can take one of two values - R_l or R_h , where $0 < R_l < R_h$. A flow starts being served at rate R_l . The service rate is set to R_h if the flow encounters no drops for a certain number of consecutive packets, say C . This models the slow-start phase of TCP. When the buffer of the downstream queue has q packets, an incoming packet is dropped with probability $f(q)$, where $f(q)$ is a non-decreasing function of q . If a packet is dropped and the current service rate of the flow is R_h , the rate is instantly reduced to R_l . This captures the dynamics of the TCP connection, while maintaining the parsimony of the model. As a first step, we ignore delays in the feedback loop and assume acknowledgements and drops are instantly known at the edge. The access rates R_l and R_h are assumed to be $O(1)$ to reflect the fact that TCP imposes a limitation on the maximum window size of a connection.

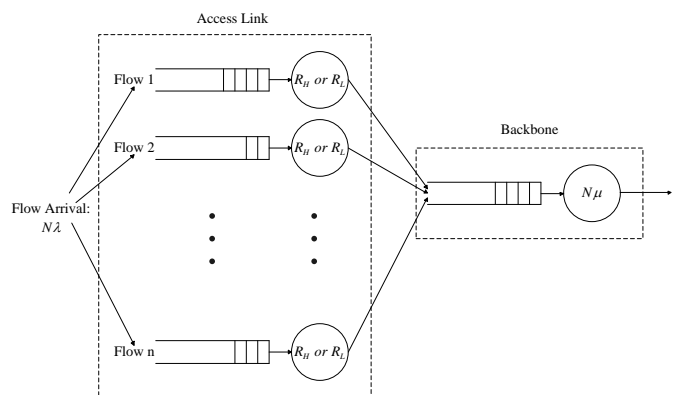


Fig. 4: Flow- and Packet-level model for a single bottleneck link

We note that this network is stable as long as $\lambda E[S] < \mu$. As N grows, we now show that this model qualitatively captures the important features of a TCP network. First, consider flow durations. In the model, a flow at the edge starts off being

served at the low rate R_l . Hence the completion time of flow with S packets is at least $\frac{C}{R_l} + \frac{S-C}{R_h}$ if $S \geq C$, and $\frac{S}{R_l}$, otherwise. Since R_l does not scale with N , the flow durations *do not* go to zero as link bandwidth grows with N , consistent with the behavior of a TCP flow. (Recall from the previous section that a lower bound on the duration of a TCP flow with S packets is $\log_2 S$ round-trip times.)

Next, consider the number of active flows in the model. Since flows arrive at rate $N\lambda$ and flow durations are $O(1)$, the number of active servers at the edge is $\Theta(N)$. This is consistent with the scaling behavior of a network with TCP flows, as illustrated in the previous section.

Finally, we consider the aggregate packet traffic entering the downstream queue. The arrival rate of packets to the queue is $NE[S]\lambda$. This grows linearly with N , while the packet processing time at the downstream queue decreases as $\frac{1}{N\mu}$. In other words, as N increases, the fluctuations in the downstream queue become more rapid, but the average load remains the same.

The inter-arrival time at the downstream queue between two successive packets of the same flow is at least $\frac{1}{R_h}$ sec. Since the downstream queue fluctuates $\Theta(N)$ times per second, the number of fluctuations between the arrival times of two packets of any flow is at least $\Theta(\frac{N}{R_h}) = \Theta(N)$. Since the queue fluctuates a large number of times between the arrivals of the two packets, we expect the queue sizes encountered by different packets of the same flow to be independent. This intuition is verified by simulations in the next section.

IV. THE AGGREGATE PACKET PROCESS FOR LARGE N

A. The Packet Drop Process

Consider the following thinned version of the original flow arrival process. Pick any $K > 0$ and assume each arriving flow is ‘tagged’ with probability $\frac{K}{N}$. Since the original arrival process is Poisson with rate $N\lambda$, the arrival process of tagged flows is a Poisson process of rate $K\lambda$. We refer to packets belonging to tagged flows as tagged packets. Let $P_K(t)$ be the counting process corresponding to arrivals of tagged packets at the downstream queue. Let $M_K(i)$ be an indicator function which is equal to 1 if the i th tagged packet arriving at the downstream queue is dropped, and 0 otherwise. Thus $M_K(\cdot)$ is the binary packet drop process corresponding to packets from the tagged flows only.

The arrival rate of packets at the downstream queue is $N\lambda E[S]$, while the arrival rate of the tagged packets is $\lambda K E[S]$. Thus the number of untagged packets arriving between any two successive tagged packets is $\Theta(\frac{N-K}{K})$. If we keep K fixed and let N grow, the queue fluctuates a large number of times between the arrivals of two successive tagged packets, and we expect the queue sizes encountered by these packets to be independent. This intuition is summarized by the following conjecture.

Conjecture: For a fixed K , $M_K(\cdot)$ converges to an iid Bernoulli process as $N \rightarrow \infty$.

We do not have a proof of the conjecture yet, but experimentally verify it below.

Simulation Set-up: Each flow is chosen according to a heavy-tailed Pareto distribution with parameter α and mean $E[S]$. The packet drop probability $f(q)$ is 1 if the queue-size $Q > Q_{limit}$ and zero otherwise. The value of all the quantities used in the simulation are given in Table II.

TABLE II: Simulation parameters

| | |
|-------------|-----------------------|
| μ | 1.0 |
| $E[S]$ | 20 |
| α | 1.5 |
| λ | $0.4/E[S]$ (40% load) |
| R_h | 4μ |
| R_l | 0.25μ |
| Q_{limit} | 3 |

We fix K , and observe the correlation in $M_K(\cdot)$ for different values of N . We plot two quantities: 1) The autocorrelation function of M_K normalized by the variance and, 2) the pmf of the spacing between successive ones (drops) in M_K . For any iid Bernoulli process, the normalized autocorrelation function is the Kronecker delta function, and the pmf of the time spacing is geometric.

Figure 5 shows these two plots for increasing values of N with $K = 10$. When $N = K = 10$, every flow is tagged and $M_K(\cdot)$ is the drop-process corresponding to *all* arriving packets. In this case, we expect M_K to have a large auto-correlation for small delays since successive packets will encounter similar levels of congestion. As N increases, there is an increasing number of untagged packets arriving between two successive tagged packets. Hence the queue sizes encountered by successive tagged packets look more independent as N increases.

B. The Aggregate Packet Arrival Process

Based on the conjecture, we prove the following result about the aggregate packet process from tagged flows entering the downstream queue.

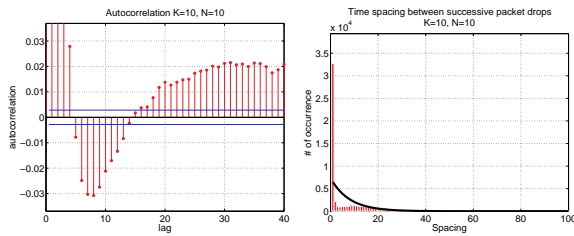
Theorem 1: Let $P_K(t)$ be the arrival process of tagged packets at the downstream queue. Define

$$A_K(t) \triangleq P_K\left(\frac{t}{K}\right).$$

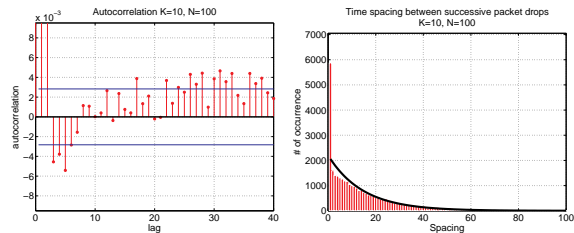
Under the conjecture that the drop process $M_K(\cdot)$ is iid, $A_K(t)$ converges weakly to a Poisson process of rate $E[S]\lambda$ as $K \rightarrow \infty$.

The proof of the Theorem is given in the Appendix. Note that $A_K(t)$ is the arrival process of tagged packets on a time-scale expanded by a factor of K . We demonstrated by simulation that when $N \gg K$, $M_K(\cdot)$ is well-approximated by an iid Bernoulli process. We can satisfy the conditions of the theorem by letting N go to ∞ faster than K ($N = K^2$, for example).

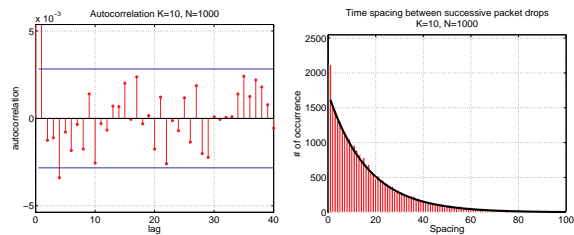
We note that a result related to Theorem 1 is found in [15]. In [15], a fixed number N of infinitely long-lived flows send packets into a single-server queue according to a stationary point process with constant rate. It is shown that as $N \rightarrow \infty$, the queue behaves as if it were fed by a Poisson input process. In our model, we have short-lived flows whose sending rates depend on the congestion encountered in the downstream



(a) $K = 10, N = 10$



(b) $K = 10, N = 100$



(c) $K = 10, N = 1000$

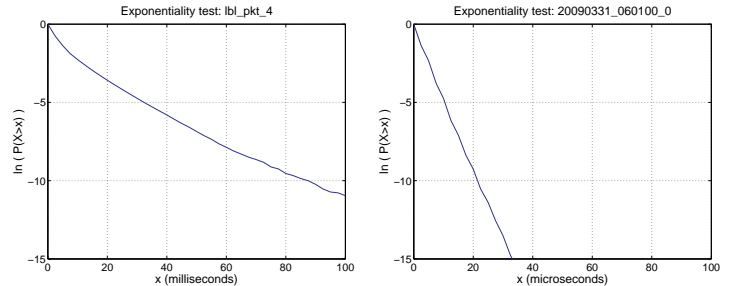
Fig. 5: Analysis of the packet drop process $M_K()$

queue. In both these results, statistical multiplexing of packets from the different flows is responsible for smoothing of the aggregate packet process seen by the downstream queue.

V. DISCUSSION

Figure 3 provides a parsimonious model of a large network in terms of the arrival rates of the flows and the link capacities. The parameters governing the access rate - R_H, R_L , and C - can be adjusted to capture the essential features of TCP: the slow-start mechanism, the growth of the congestion window, and the maximum window size. For the simple case of a single bottleneck link, Theorem 1 says that if we observe the aggregate packet traffic of any subset of arriving flows, it resembles a Poisson process if the link speed is sufficiently large. In fact, simulations and analysis of Internet traces (described below) suggest that a stronger result may hold: the aggregate packet traffic into the downstream queue, when suitably scaled, converges to a Poisson process as $N \rightarrow \infty$. If such a result were true:

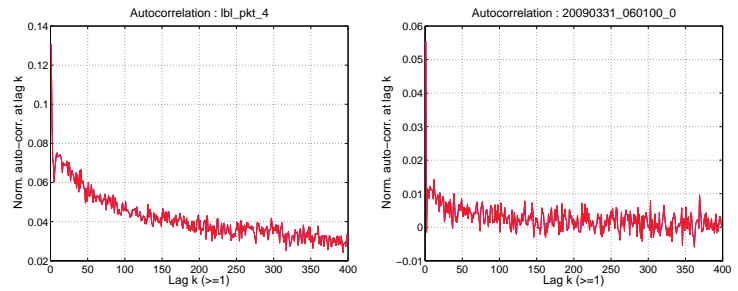
- The network of routers becomes a Jackson network [16] as $N \rightarrow \infty$, leading to a product-form decomposition of the queue-size distributions. The queue-size distributions at the buffers could then be easily estimated.
- From the queue size-distribution and the function $f(q)$ used for dropping, the packet drop probability can be computed. Using this, one can compute the fraction of



(a) LBL trace (1994)

(b) OC192 trace (2009)

Fig. 6: Distribution of the packet interarrival time



(a) LBL trace (1994)

(b) OC192 trace (2009)

Fig. 7: Autocorrelation of the packet interarrival times

packets of a given flow that are served at the higher rate R_h . Using this, the flow durations can be estimated.

There have been extensive efforts to characterize the Internet packet traffic based on trace analysis. Early studies such as [13], [17] found that Internet traffic exhibits long-range dependent behavior, characterized by the sustained correlation of packet counts or packet interarrival times over a long period of time. In contrast, more recent studies such as [18], [19] find evidence that the Internet traffic exhibits Poisson-like behavior, especially on short time scales.

The model of Section III pinpoints the reason for this change in internet traffic characteristics. In the early 90s, both flow arrival rates and link speeds were quite low. In our model, this corresponds to N being small, and hence the aggregate packet traffic does not exhibit Poisson behavior. Both backbone link speeds and flow arrival rates have increased more than a thousand-fold since the 90s, corresponding to a much larger N in our model. Thus there is increased multiplexing of packets from different flows, leading to smoothing of the aggregate packet process.

For illustration, we analyze two traces: 1) an LBL trace from 1994 that was used in [13] to show long-range dependence of internet traffic, and 2) a trace from an OC192 (10 Gps) link in 2009, obtained from CAIDA [14]. Recall that for a Poisson process, inter-arrival times are independent and exponentially distributed. We test the packet inter-arrival times of the traces for both of these properties. Figure 6 plots $\ln P(X > x)$ against x , where X is the inter-arrival time and $P(X)$ its empirical distribution from the traces. For the 2009 OC192

trace, the plot is close to a straight line, indicating that the distribution of the packet inter-arrival time is exponential. The plot is not a straight line for the 1994 LBL trace. Figure 7 plots the autocorrelation of the packet inter-arrival times (normalized by the variance) for various values of lag k . For LBL trace, autocorrelation does not go to zero even for large values of k , showing the long range dependence of the inter-arrival time process. For the OC192 trace, autocorrelation is very small for large k , and takes both positive and negative values. The two tests indicate that the distribution of packet inter-arrival times for the OC192 traces resemble an i.i.d exponential random variable much more than the distribution for the LBL trace. Finally, we note that [18] contains extensive trace analysis showing Poisson-like behavior of packet traffic for OC48 (2.5Gbps) backbone links.

An important factor we have ignored in this paper is the role of delay (round-trip time) in the feedback loop. It is important to study how lags affect the smoothing of the packet process caused by multiplexing of packets from different flows. The ultimate goal is to prove a convergence result for the aggregate packet process entering the downstream queue without the help of the conjecture. The challenge here is to explicitly take into account the instantaneous correlation in the access rates of the different flows.

REFERENCES

- [1] V. Misra, W. Gong, and D. Towsley, "Fluid-based analysis of a network of AQM routers supporting TCP flows with an application to red," in *In Proc. of ACM SIGCOMM*, 2000.
- [2] J. Mo and J. Walrand, "Fair end-to-end window-based congestion control," *IEEE/ACM Trans. Networking*, vol. 8, no. 5, pp. 556–567, 2000.
- [3] F. P. Kelly, A. Maulloo, and D. Tan, "Rate control in communication networks: shadow prices, proportional fairness and stability," *Journal of the Operational Research Society*, vol. 49, pp. 237–252, 1998.
- [4] R. Srikant, *The Mathematics of Internet Congestion Control*. Birkhauser, 2004.
- [5] T. Bonald and L. Massoulié, "Impact of fairness on internet performance," in *In Proc. of ACM SIGMETRICS*, pp. 82–91, 2000.
- [6] S. B. Fredj, T. Bonald, A. Proutiere, G. Régnié, and J. Roberts, "Statistical bandwidth sharing: A study of congestion at flow level," in *In Proc. of ACM SIGMETRICS*, 2001.
- [7] N. Cardwell, S. Savage, and T. Anderson, "Modeling TCP latency," in *In Proc. of IEEE INFOCOM*, 2000.
- [8] U. Ayesta, K. Avrechenkov, E. Altman, C. Barakat, and P. Dube, "Multilevel approach for modeling short TCP sessions," in *In Proc. of ITC-18*, Berlin, 2003.
- [9] C. Moallemi and D. Shah, "On the flow-level dynamics of a packet-switched network," in *In Proc. of ACM SIGMETRICS*, 2010.
- [10] D. Bertsekas and R. Gallager, *Data Networks*. Prentice Hall, 1987.
- [11] L. Kleinrock, *Queueing Systems*, vol. 2. J. Wiley and Sons, 1975.
- [12] "OMNeT++ network simulator." <http://www.omnetpp.org/>.
- [13] V. Paxson and S. Floyd, "Wide-area traffic: The failure of poisson modeling," *IEEE/ACM Trans. Networking*, pp. 226–244, June 1995.
- [14] C. Walsworth, E. Aben, K. Claffy, and D. Andersen, "The caida anonymized oc192 traces 2009." http://www.caida.org/data/passive/passive_2009_dataset.xml, 2009.
- [15] J. Cao and K. Ramanan, "A poisson limit for buffer overflow probabilities," in *In Proc. of IEEE INFOCOM*, 2002.
- [16] F. P. Kelly, *Reversibility and Stochastic Networks*. Wiley, 1979.
- [17] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, "On the self-similar nature of ethernet traffic," *IEEE/ACM Trans. Networking*, pp. 1–15, February 1994.
- [18] T. Karagiannis, M. Molle, M. Faloutsos, and A. Broido, "A nonstationary poisson view of internet traffic," in *In Proc. of INFOCOM*, 2004.
- [19] Z.-L. Zhang, V. J. Ribeiro, S. Moon, and C. Diot, "Small-time scaling behaviors of internet backbone traffic: An empirical study," in *In Proc. of INFOCOM*, 2003.
- [20] D. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes*. Springer, 2003.

APPENDIX

PROOF OF THEOREM 1

We will first give the proof for fixed flow-size S , and then extend to arbitrary flow-size distributions.

Deterministic flow size S

The arrival process of tagged packets to the downstream queue $P_K(t)$ can be written as

$$P_K(t) = P_K^1(t) + P_K^2(t) + \dots + P_K^S(t), \quad t \geq 0 \quad (1)$$

where $P_K^j(t)$ is the counting process corresponding to arrival of the j th packet of each tagged flow for $j = 1, \dots, S$, and hence

$$A_K(t) = P_K^1\left(\frac{t}{K}\right) + P_K^2\left(\frac{t}{K}\right) \dots + P_K^S\left(\frac{t}{K}\right), \quad t \geq 0 \quad (2)$$

For $j = 1, \dots, S$, let R_j be the rate at which the j th packet of a given flow is served at the access link; R_j can take value either R_h or R_l . By assumption, there is a fixed probability p of a tagged packet getting dropped which is independent of the current access rates of all the tagged flows. This induces a stationary joint distribution $P(R_1, R_2, \dots, R_S)$ which governs the access rates of the S packets in each flow. This joint distribution is determined by p and the congestion control algorithm. Denoting the Poisson flow arrival process of tagged flows by $T_K(t)$, observe that $P_K^1(t)$ is a process formed by shifting to the right each arrival in F_K independently, where the shifts are i.i.d realizations of $\frac{1}{R_1}$, with $R_1 \sim P(R_1)$. ($P(R_1)$ is the marginal from the joint distribution.) From [20, Problem 2.3.4 (b)], $P_K^1(t)$ is a Poisson process with rate $K\lambda$.

Similarly, for $1 \leq i \leq S$, $P_K^i(t)$ is a process formed by shifting to the right each arrival in F_K independently, where the shifts are i.i.d realizations of $\frac{1}{R_1} + \dots + \frac{1}{R_i}$, with $(R_1, \dots, R_i) \sim P(R_1, \dots, R_i)$. Hence $P_K^i(t)$ is a Poisson process of rate $K\lambda$ for $1 \leq i \leq S$. We now show that $A_K(t)$ defined in (2) satisfies both the stationary increments and independent increments properties as $K \rightarrow \infty$.

Stationary increments. Consider the number of points of the process $A_K(t)$ in the interval $t \in (a, a + \Delta)$, denoted $A_K(a, a + \Delta)$. This is equal to the sum of the number points of P_K^1, \dots, P_K^S that occur in the interval $(\frac{a}{K}, \frac{a+\Delta}{K})$.

$$\begin{aligned} \Pr(A_K(a, a + \Delta) = n) &= \sum_{k_S=0}^n \sum_{k_{S-1}=0}^{n-k_S} \dots \sum_{k_2=0}^{n-(k_S+\dots+k_3)} \Pr \left[P_K^S \left(\frac{a}{K}, \frac{a+\Delta}{K} \right) = k_S, \right. \\ &\quad \left. \dots, P_K^1 \left(\frac{a}{K}, \frac{a+\Delta}{K} \right) = n - (k_S + \dots + k_2) \right]. \end{aligned} \quad (3)$$

Since P_K^S is a Poisson process of rate $K\lambda$, we can write the above using conditioning as

$$\begin{aligned} & \Pr(A_K((a, a + \Delta)) = n) \\ &= \sum_{k_S=0}^n \sum_{k_{S-1}=0}^{n-k_S} \dots \sum_{k_2=0}^{n-(k_S+\dots+k_3)} \frac{e^{-\lambda\Delta}(\lambda\Delta)^{k_S}}{k_S!} \\ & \cdot \Pr\left[P_K^{S-1}\left(\frac{a}{K}, \frac{a+\Delta}{K}\right) = k_{S-1} \mid P_K^S\left(\frac{a}{K}, \frac{a+\Delta}{K}\right) = k_S\right] \dots \\ & \cdot \Pr\left[P_K^1\left(\frac{a}{K}, \frac{a+\Delta}{K}\right) = n - (k_S + \dots + k_2) \mid P_K^S\left(\frac{a}{K}, \frac{a+\Delta}{K}\right) = k_S\right] \end{aligned} \quad (4)$$

Consider the first conditional probability term in the above equation. Let the k_S points in $P_K^S\left(\left(\frac{a}{K}, \frac{a+\Delta}{K}\right)\right)$ occur at times t_1, t_2, \dots, t_{k_S} , $\frac{a}{K} < t_i < \frac{a+\Delta}{K}$. Let the corresponding S th packet service times for these points be D_1, D_2, \dots, D_{k_S} , respectively. Note that D_1, D_2, \dots, D_{k_S} are independently distributed $\sim \frac{1}{R_S}$.

Then, the points in $P_K^S\left(\left(\frac{a}{K}, \frac{a+\Delta}{K}\right)\right)$ exactly correspond to k_S points in the process $P_K^{S-1}(t)$ at times $(t_1 - D_1, \dots, t_k - D_{k_S})$. Since

$$\frac{1}{R_h} \leq D_i \leq \frac{1}{R_l}, \quad \forall i,$$

these k points all lie within the interval $\left(\frac{a}{K} - \frac{1}{R_l}, \frac{a+\Delta}{K} - \frac{1}{R_h}\right)$. If

$$\frac{a+\Delta}{K} - \frac{1}{R_h} < \frac{a}{K}, \quad (5)$$

then $\left(\frac{a}{K}, \frac{a+\Delta}{K}\right)$ and $\left(\frac{a}{K} - \frac{1}{R_l}, \frac{a+\Delta}{K} - \frac{1}{R_h}\right)$ are disjoint intervals. Thus if (5) holds, the points of $A_K((a, a + \Delta))$ from P_K^S and P_K^{S-1} correspond to points of P_K^{S-1} in *disjoint* intervals of time. Let E denote the event that $R_h < \frac{K}{\Delta}$, i.e., (5) holds. Then we have

$$\begin{aligned} & \Pr\left[P_K^{S-1}\left(\frac{a}{K}, \frac{a+\Delta}{K}\right) = k_{S-1} \mid P_K^S\left(\frac{a}{K}, \frac{a+\Delta}{K}\right) = k_S\right] \\ &= \Pr\left[E, P_K^{S-1}\left(\frac{a}{K}, \frac{a+\Delta}{K}\right) = k_{S-1} \mid P_K^S\left(\frac{a}{K}, \frac{a+\Delta}{K}\right) = k_S\right] \\ & \quad + \Pr\left[E^c, P_K^{S-1}\left(\frac{a}{K}, \frac{a+\Delta}{K}\right) = k_{S-1} \mid P_K^S\left(\frac{a}{K}, \frac{a+\Delta}{K}\right) = k_S\right] \end{aligned} \quad (6)$$

We have

$$\begin{aligned} & \Pr\left(P_K^{S-1}\left(\frac{a}{K}, \frac{a+\Delta}{K}\right) = k_{S-1} \mid E, P_K^S\left(\frac{a}{K}, \frac{a+\Delta}{K}\right) = k_S\right) \\ &= \frac{e^{-\lambda\Delta}(\lambda\Delta)^{k_{S-1}}}{(k_{S-1})!} \end{aligned}$$

because conditioned on E , the points in $P_K^S\left(\frac{a}{K}, \frac{a+\Delta}{K}\right)$ all correspond to points of P_K^{S-1} in a disjoint interval from $\left(\frac{a}{K}, \frac{a+\Delta}{K}\right)$.

Also note that for any fixed Δ , $R_h < \frac{K}{\Delta}$ for large enough K . Hence,

$$\lim_{K \rightarrow \infty} P(E) = 1$$

and (6) becomes

$$\begin{aligned} & \lim_{K \rightarrow \infty} \Pr\left[P_K^{S-1}\left(\frac{a}{K}, \frac{a+\Delta}{K}\right) = k_{S-1} \mid P_K^S\left(\frac{a}{K}, \frac{a+\Delta}{K}\right) = k_S\right] \\ &= \frac{e^{-\lambda\Delta}(\lambda\Delta)^{k_{S-1}}}{k_{S-1}!} \end{aligned} \quad (7)$$

We can similarly find the limiting value of each term in the expansion (4) to obtain

$$\begin{aligned} & \lim_{K \rightarrow \infty} \Pr(A_K(a, a + \Delta) = n) \\ &= \sum_{k_S=0}^n \dots \sum_{k_2=0}^{n-(k_S+\dots+k_3)} \frac{e^{-\lambda\Delta}(\lambda\Delta)^{k_S}}{k_S!} \cdot \frac{e^{-\lambda\Delta}(\lambda\Delta)^{k_{S-1}}}{k_{S-1}!} \\ & \quad \dots \frac{e^{-\lambda\Delta}(\lambda\Delta)^{n-k_S-\dots-k_2}}{(n-k_S-\dots-k_2)!} \\ &= \frac{e^{-S\lambda\Delta}(\lambda\Delta)^n}{n!} \sum_{\substack{k_1, \dots, k_S: \\ \sum k_i = n}} \binom{n}{k_1 k_2 \dots k_S} \\ &= \frac{e^{-S\lambda\Delta}(S\lambda\Delta)^n}{n!}. \end{aligned} \quad (8)$$

Thus $A_K(a, a + \Delta)$ has a Poisson distribution with parameter Δ for all a , which is the stationary increments property.

Independent Increments. Consider two disjoint intervals $(b, b + \delta)$ and $(a, a + \Delta)$, with $a > b + \delta$. We need to show that $A_K((a, a + \Delta))$ is independent of $A_K((b, b + \delta))$ as $K \rightarrow \infty$. We have

$$\begin{aligned} A_K(a, a + \Delta) &= P_K^1\left(\frac{a}{K}, \frac{a+\Delta}{K}\right) + \\ & \quad \dots + P_K^{S-1}\left(\frac{a}{K}, \frac{a+\Delta}{K}\right) + P_K^S\left(\frac{a}{K}, \frac{a+\Delta}{K}\right), \\ A_K(b, b + \delta) &= P_K^1\left(\frac{b}{K}, \frac{b+\delta}{K}\right) + \\ & \quad \dots + P_K^{S-1}\left(\frac{b}{K}, \frac{b+\delta}{K}\right) + P_K^S\left(\frac{b}{K}, \frac{b+\delta}{K}\right). \end{aligned} \quad (9)$$

We now show that each term in $A_K(a, a + \Delta)$ is independent of $A_K(b, b + \delta)$ as $K \rightarrow \infty$. Consider $P_K^S\left(\frac{a}{K}, \frac{a+\Delta}{K}\right)$.

- $P_K^S\left(\frac{a}{K}, \frac{a+\Delta}{K}\right)$ is independent of $P_K^S\left(\frac{b}{K}, \frac{b+\delta}{K}\right)$ for all K since they correspond to disjoint intervals of the Poisson process $P_K^S(t)$.

- As before, let the points in $P_K^S\left(\frac{a}{K}, \frac{a+\Delta}{K}\right)$ occur at times t_1, t_2, \dots, t_{k_S} , $\frac{a}{K} < t_i < \frac{a+\Delta}{K}$ and let D_1, D_2, \dots, D_{k_S} be the corresponding S th packet service times for these points. Then these exactly correspond to k_S points in the process $P_K^{S-1}(t)$ at times $(t_1 - D_1, \dots, t_k - D_{k_S})$. If

$$\frac{a+\Delta}{K} - \frac{1}{R_h} < \frac{b}{K}, \quad (10)$$

then $\left(\frac{b}{K}, \frac{b+\delta}{K}\right)$ and $\left(\frac{a}{K} - \frac{1}{R_l}, \frac{a+\Delta}{K} - \frac{1}{R_h}\right)$ are disjoint intervals. (10) is equivalent to $R_h < \frac{K}{a+\Delta-b}$, which holds for any finite Δ, a and b ($a > b$) for large enough K . Thus $P_K^S\left(\frac{a}{K}, \frac{a+\Delta}{K}\right)$ is independent of $P_K^{S-1}\left(\frac{b}{K}, \frac{b+\delta}{K}\right)$ as $K \rightarrow \infty$.

Similarly, we can show that each term in the expansion (9) of $A_K(a, a + \Delta)$ is independent of each term of $A_K(b, b + \delta)$ as $K \rightarrow \infty$. This completes the proof that $A_K(a, a + \Delta)$ is independent of $A_K(b, b + \delta)$ for all sufficiently large K .

Random Flow-sizes

Let p_i denote the probability that a flow has size i , so that $E[S] = \sum_i i p_i$. Denote by $P_{Ki}(t)$ the packet-departure process corresponding to tagged flows with size i . Then the packet departure process $P_K(t)$ can be written as

$$P_K(t) = \sum_i P_{Ki}(t) \quad (11)$$

and hence

$$A_K(t) = \sum_i A_{Ki}(t) = \sum_i P_{Ki}(t/K). \quad (12)$$

The processes $\{P_{Ki}(\cdot)\}_i$ are mutually independent because the assumption of iid drops ensures that the (time-varying) access rate of each tagged flow is independent of all other tagged flows.

The arrival process of tagged flows of size i is a Poisson process of rate $K p_i \lambda$ (thinning a Poisson process of rate $K \lambda$ by a factor of p_i). From the result proved above for the deterministic case, $A_{Ki}(t)$ converges to a Poisson process of rate $i p_i \lambda$. Thus in the limit $K \rightarrow \infty$, (12) is a sum of independent Poisson processes with the i th process in the sum have rate $i p_i \lambda$. Thus, $\lim_{K \rightarrow \infty} A_K(t)$ is a Poisson process with rate $\sum_i i p_i \lambda = E[S] \lambda$.