Network Tomography

Stephan Bohacek
bohacek@math.usc.edu
and
Boris Rozovskii
rozovski@math.usc.edu

Department of Mathematics,
University of Southern California
Denney Research Building, DRB 155
Los Angeles, California 90089-1113

Abstract— **A new approach to network tomography is presented. This technique is not only useful for assessing the "health" of the network, but also useful for developing advanced congestion control algorithms. The approach relies on sending data packets and receiving acknowledgments. Based on the acknowledgments, the state of the network along the path between the source and destination is determined. This work presents a framework for assessing the state, while it is expected that future work will focus on issues such as determining what the normal state of the network is and combining information about the state along several paths into a picture of the entire network.**

I. Introduction

A model based approach for estimating the "state" of the network via measurements made by hosts is presented. Both network models and methods to estimate the model parameters are developed. These parameters then allow a detailed characterization of the state network. The measurements are made by sending packets between two hosts and observing drops and latency. The state consists of the state of each router along the path between the two hosts and parameters of a drop probability model. The drop models is useful for determining if the drops experienced are due to heavy congestion or an attack. When used in conjunction with distributed data collection from other host pairs, the round-trip time can be used for pinpointing heavily congested routers. This type of monitoring is useful in detecting network attacks. For example, a carefully designed distributed denial of service attack could cause congestion on one critical link without causing congestion at any particular server. Hence, the attack would be difficult to detect directly at the server. The methods presented here allows a detailed state of the network to be developed and, hence, any sudden increase in congestion on a particular link could be easily detected.

The models of the round-trip latency and drop probability are developed independently. The round-trip time is the aggregate of the fixed transmission delay and delays experienced at each router along between the hosts. The state of the router consists of the queue size, the sending rate and average arrival rate. The packet arrivals are modeled as a Poisson processes whereas the sending rate is assumed to be fixed but unknown. The objective is to determine the probability distribution of the state of the router given observed round-trip times. The drop model is a parametric model that depends on the host sending rate, round-trip time and an internal abstract variable that models the level of congestion. Here the objective is to determine the probability distribution of the level of congestion based on the observation of drops.

The measurements are performed at a single sending host. It is assume that packets are sent from the sending host to a receiving host and the receiver sends acknowledgments upon receiving a packet from the sender. This test flow from sender to receiver is referred to as the controlled flow or simply CF. Note that the CF may be pings or actual data, e.g. an http connection. In the case when the CF is an actual data flow, this monitoring technique places no added burden on the network.

The objective of determining the state of the Internet has been address by others. For example, in [8] the drop probability in the multicast network is determined. There are many tools for monitoring traffic on the Internet [2], [3], [5], [1], [4]. Typically these tools ping various routers at low frequency and provide a low resolution description latency and drop probability over a large part of the network. The work presented here takes a different approach and provides a detailed, high frequency description of a small segment of the network. However, this work could be extended to provide such a description of a larger part of the network.

The paper proceeds as follows: Section II-A develops the round-trip time model. Section II-B developed the parametric drop model. Section III-A shows how the observation can be used to determine the state of the routers between the hosts. Section III-B shows how drop observation can be used to estimate the level of congestion.

II. Models

In the next two subsections models for the round-trip time and probability of a packet drop are developed.

A. Round-trip Time Model

We define a model of RTT_t the round-trip time experienced by a packet send at time t. The round-trip time experienced by a packet sent at time t is the following sum

$$RTT_t = T_{fixed} + T_{proc} + D_t, \tag{1}$$

where T_{fixed} is the sum of transmission delays and propagation delays and is assumed to be constant[1] and known,

[1] In particular, it is assumed that the routing policy at the routers between the source and destination is fixed. We neglect stochastic

T_{proc} accounts for processing delay at both the source and destination as well as random delays such as the time to resolve access contention on Ethernet links and D_t is the queuing delay experienced by a packet sent at time t.

We will assume that $T_{proc} \geq 0$ and that the probability distribution of this random variable is known. While there are very elaborate and good models of processing delay, we have found that a simple triangle density works well, i.e.,

$$p\left(T_{proc} = t\right) = \begin{cases} \frac{r+t}{r^2} \text{ for } -r \leq t \leq 0 \\ \frac{r-t}{r^2} \text{ for } 0 < t \leq r \\ 0 \text{ otherwise} \end{cases}, \tag{2}$$

where r is some known number.

The queueing delay, D_t, is responsible for large variations in the round-trip time. We assume that the queueing delay at a router is the queue occupancy of the router multiplied by the sending rate of the router. Note that we assume that the router can process packets at a fixed rate and neglect such effects as the variation in the time required for destination address lookup. We define the occupancy of the i^{th} queue at time t to be q_t^i and the maximum queue size to be $q_{\max}^i$. Furthermore, let μ^i be the time it takes the queue to service a packet of size one. Thus, the queuing delay is

$$D_t = \sum_{i=1}^{n} \mu^i q_t^i, \tag{3}$$

where we assume that there are n routers between source, destination and back to source. We model the arrival of a new packet to router i as Poisson process L_t^i with unknown intensity λ^i. This queue model is greatly simplified. For example, this model neglects the known fact that packet arrivals are correlated and give rise to self-similar queue occupancy [13]. It also neglects the effects of congestion control. For example, if the queue overflows, then a short time later this overflow will be detected by a sender and the sender will decrease its sending rate. Hence, the arrivals are not independent of the past queue size. Similarly, when the queue fills, the round-trip times increase, hence TCP flows reduce there sending rate. Furthermore, due to new connections becoming active and others completing their transfer, the arrival rates should be time varying.

We model the size of the new packet as a random number with known distribution ρ. Estimates of the distribution of packet sizes can found by examining publicly available traces. For example, the traces captured by the NLANR project [11] produced the packet size cumulative distribution shown in Figure 1. This distribution is well a modeled by

$$P\left(z\right) = \begin{cases} 0.91 + 0.02 \ln\left(z\right) & \text{for } 0.0106 \leq z < 1 \\ 1 & \text{for } z \geq 1 \\ 0 & \text{otherwise} \end{cases}.$$

Note that packet size is normalized with a packet size equal to one refers to a packet with 1500 bytes. Hence, the dynamics of the queue are given by

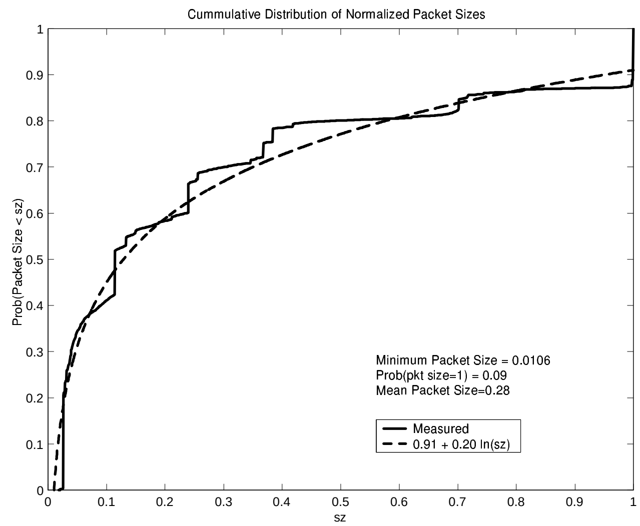


Fig. 1. Cummulative distribution of packet size. A packet of size one has 1500 bytes. The data was collected by the NLANR project.

$$dq_t^i = -\mu^i 1_{\{q_t^i \geq 0\}} dt + 1_{\{q_t^i < q_{\max}^i\}} \gamma dL_t^i, \tag{4}$$

where γ is a random variable representing the size of the packet that arrives at time t, and $1_{\{q_t^i < q_{\max}^i\}}$ is the indicator function and forces $q_t^i \leq q_{\max}^i$ and similarly, the indicator $1_{\{q_t^i \geq 0\}}$ forces $q_t^i \geq 0$.

While in general a packet and acknowledgment pass through many routers and queues, for ease of presentation we will make the simplifying assumption that there is only one queue. This queue models the aggregate of all the queues along the path between source to destination and back. Then (3) is replaced with

$$D_t = \mu q_t \tag{5}$$

and (4) is replaced with

$$dq_t = -\mu 1_{\{q_t \geq 0\}} dt + 1_{\{q_t < q_{\max}\}} \gamma dL_t. \tag{6}$$

In this aggregate model the sending rate is the sum of the sending rates, i.e. $\mu = \sum_{i=1}^{n} \mu^i$. However, it is not true that the arrival rates are the sum of the arrival rates, i.e. in general $\lambda \neq \sum \lambda^i$. Indeed, because a packet may pass through multiple queue, L_t is not even Poisson. In order to determine the distribution of the time between packet arrivals in the aggregate queue, the we must know the probability that a packet in one queue will travel to another queue. While it is possible that estimates of this distribution could be determined, we make the assumption that the arrivals at the aggregate queue are Poisson with intensity λ. It is straight forward to extend the results presented here to the case of many intermediate routers.

B. Drop Event Model

B.1 Model

The is no consensus on the distribution of packet drops in large networks. However, there has been extensive work

routing policies [12] can also lead to variations in propagation delay.

on characterizing the drops experienced by a TCP flow, the predominate congestion control mechanism of Internet traffic. In [14] a small network is considered and a deterministic model for packet drops is developed. In [16] drops are assumed to be highly correlated over short time scales and independent over longer time scales. In [6] drops are assumed to be bursty. Furthermore, [6] makes the distinction between drop events (moments when the congestion window is halved) and packet drops. In the sequel we refer to drop events and as either drops or drop events. In [7] drops events are modeled as a renewal process with various distributions; deterministic, Poisson, i.i.d and Markovian. A specific example of the model in [7] is developed in [15], where drop events are modeled by a Poisson process and, hence, the time between drop events are exponentially distributed. In [17], this approach is generalized and drops events are modeled as a Poisson process where the intensity depends on the window size of the TCP protocol. We further generalize this approach so that the drops events experienced by a TCP flow are modeled as Poisson processes that depend on the round-trip time, the sending rate and an abstract state variable. While our initial experimental results are presented that validate this model, we should stress that extensive work remains before any definitive conclusions can be made. However, the work presented in this paper remains valid for other drop event models.

We model drop events a flow as a doubly stochastic Poisson process $N_t = (N_t)_{t \geq 0}$ with random intensity $n(V^t, \theta^t, RTT^t)$, where V_t is the known data sending rate of the controlled flow (CF), θ is an abstract random variable representing the network congestion and RTT is the round-trip time. In a network, a drop occurs when both the queues are full and packets enter the queue faster than they leave the queue. Since full queues lead to a long round-trip time, it is reasonable to expect some correlation between round-trip time and drops. Indeed our experimental and simulation results indicate this to be the case. The parameter θ models the level of congestion in the network. For example, if n TCP flows are active, the aggregate window size increases at a rate of n/RTT. Thus, when more flows are active we expect the aggregate sending rate of the competing flows to vary rapidly. Hence, we expect that when many flows are active, drops are more likely. Thus, θ can be used to model the variability in the competing traffic. However, we do not specifically define θ to be the number of competing flows, rather θ is an abstract variable that is related to the level of congestion and, therefore, related to the probability of getting a drop. The rational is that we are not interested in the exact number of competing flows, but the probability of drops. Thus, we assume that the congestion level θ is a homogeneous Markov jump process, taking values in the finite alphabet $A = \{a_1, \ldots, a_M\}$, with the intensity matrix $\Lambda = ||\lambda(a_i, a_j)||$ and the known initial distribution $p_q = P(\theta_0 = a_q)$, $q = 1, \ldots, M$. Hence, at random times θ makes jumps where if $\theta_t = a_i$, the time to the next jump is exponentially distributed with parameter $\sum_{j \neq i} \lambda(a_i, a_j)$. In practice these transition probabilities are difficult to estimate. Furthermore, it seems plausible that these transitions will depend on the number of hops, the time of day and need to account for self-similar characteristic of network traffic. A deeper investigation into these probabilities will be left for future work.

One way to view drops is that at any particular time a packet sent has a probability of being dropped. In this case the rate of drops is the product of the rate that packets are sent multiplied by the probability that a packet is dropped, i.e.

$$n\left(\theta^t, RTT^t, V^t\right) = V_t \ g\left(\theta^t, RTT^t, V^t\right),$$

where V_t is the sending rate when the packet was sent and

$$g\left(\theta^t, RTT^t, V^t\right) = P\left(\text{drop} \middle| \, \theta^t, RTT^t, V^t\right)$$

is the probability that a packet is dropped. We will assume that drops are independent. In full generality we would allow g to depend on the not only the states θ_t, RTT_t, V_t but also the history θ^t, RTT^t, V^t. For example, the competing TCP control mechanisms are dynamical systems with memory. Hence, when a queue overflows, the competing flows will be subjected to drops. After a drop, TCP flows go through the usually cut in sending rate and the slow increase. Thus, the competing flows' sending rates are not only a function of the current state of the network, but the past state of the network. Since the sending rates of the competing flows' is closely related to the probability of getting a drop, n depends on the past state of the network. However, we will focus on the memoryless case where

$$g\left(\theta_t, RTT_t, V_t\right) = P\left(\text{drop} \middle| \, \theta^t, RTT^t, V^t\right).$$

One drawback of this model is that it assumes that the time between drops is independent. This is clearly not true for synchronized flows [20]. However, for multi-hop connections, perhaps with routers implementing active queue management such as RED [10], synchronization is often assumed to be rare. Furthermore, while our experiments indicate that drops are not bursty, others have reported experimental results indicating that drops do occur in bursts. Of course, busrtiness over long time scales can be modeled with θ, however, it is not clear if bustiness exists at short time scales. A study of burstiness over short time scales can be found in [9].

One important drawback of the model is that there is no direct link between the round-trip time model and the drop model. In particular it seems plausible that λ, arrival rate of packets to the queue, should be related to the level of network congestion θ. However, for simplicity this connection is neglected.

B.2 Simulations and Experiments

In an attempt to characterize the form of the drop probability g, we have performed simulation and experiments. Based on simulation results we have found a useful form of g is given by

$$\begin{aligned} & g\left(\theta_t, RTT_t, V_t\right) \\ & = a_0\left(\theta_t\right) + a_1\left(\theta_t\right) V_t^{-1} + a_2\left(\theta_t\right) RTT_t + a_3\left(\theta_t\right) RTT_t V_t^{-1}. \end{aligned}$$

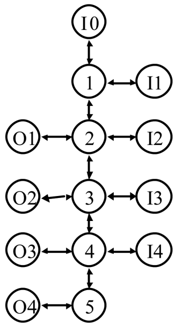


Fig. 2. Topology for verifying form of drop probabilities. The controlled flow has source I0 and destination O4. The competing TCP flows are have source-destination pairs as follows: (I1, O1), (I1, O2), (I1, O3), (I1, O4), (I2, O2), (I2, O3), (I2, O4), (I3, O3), (I3, O4), (I4, O4). The number of TCP sharing each of the above source destination pairs is allowed to vary. Specifically, each pair had one flow or each pair had two flows, etc.

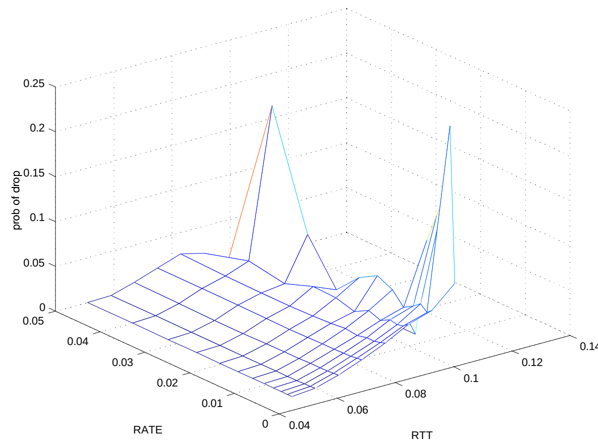


Fig. 3. Typical experimentally calculated drop probability given RTT and sending rate. Here each source-destination pair has 3 TCP flows. The RTT is measured with respect to the CF. Note that the histogram appears affine for small enough RTT. Large RTT's occur very rarely, so the variance for large values of RTT is very high. For example, some of the large RTT values occurred only a few times, were as moderate values of RTT occurred thousands of times.

That is, the probability of a packet being dropped is an affine function of the reciprocal of sending rate, V^{-1}, the round-trip time RTT, and the product of the rate and round-trip time $RTT\ V^{-1}$.

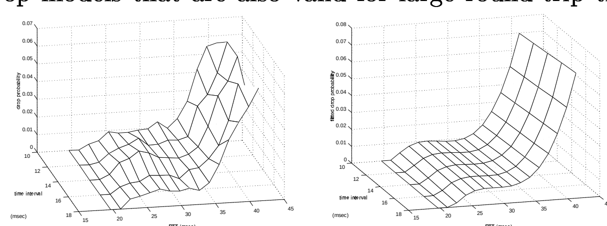
We have verified such an drop model for some topologies via ns-2 simulation. In particular, we simulated multiple TCP flows over the topology shown in Figure 2. The plot of a typical the calculated drop probability given RTT and sending rate is shown in Figure 3. Notice that for small enough RTT, the probability appears affine. While for large RTT, the calculated probability appears highly nonlinear. However, RTT was large a relatively few number of times, hence these calculations are subject to large error. For the most significant values of round-trip time, the model fits the data very well. The coefficients for different number of competing TCP flow are given in the following table.

Number of TCP flows per source-destination pair	a_0	a_1	a_2	a_3
1	−0.06	0.303	1.056	−5.417
2	−0.105	0.501	1.820	−8.824
3	−0.144	0.903	2.477	−14.624
4	−0.179	1.586	3.080	−24.254
5	−0.216	2.256	3.694	−38.142

We performed several experiments. In this case, the g we choose was

$$g(\theta_t, RTT_t, V_t) = a_0(\theta_t) + a_1(\theta_t) V_t^{-1} + a_2(\theta_t) RTT_t + a_3(\theta_t) RTT_t V_t^{-1} + a_4 RTT_t^2 + a_5 RTT_t^3. \tag{7}$$

We analyzed a 14 hop connection with source and destination in the Los Angeles area. A new model was calculated every hour for an entire day. The first set of figures below show the drop probability of the form (7) calculated at 2PM. Figure 4 shows two measure of goodness of fit, the variance and R versus time. Note that in the middle of the day, the model perform quite well. However, in the evening performance is degraded. One factor causing this degradation is that there was less congestion in the evening and hence few drops. With few drops occurring, the drop probability is more difficult to determine. The plots do not show the drop probability for large RTT. While large round-trip times did occasionally occur, they were infrequent enough that calculating the drop probability was extremely noisy. On going work is currently focusing on collecting more data in order to determine drop models that are also valid for large round-trip time.



The lefthand plot show the calculated drop probability versus round-trip time and the reciprocal of the sending rate. The righthand plot shows the function g which approximates the drop probability.

III. Distributions

Next we develop techniques to estimate the parameters of the round-trip time model and the drop model. It is assumed that the acknowledgments arrive at discrete times τ_k. An acknowledgment that arrives at time τ_k provides round-trip time of a packet sent at time $t < \tau_k$. Furthermore, this acknowledgment may indicate that packets have been dropped. We assume that packets are not reordered. Hence, if at time τ_k an acknowledgment arrives for a packet sent at time t, and there is an unacknowledged packet which was sent at time $t_0 < t$, then it is assumed that the un-

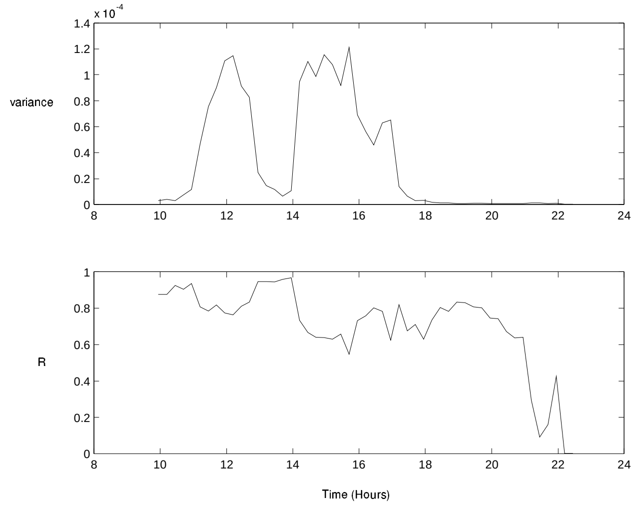


Fig. 4. Goodness of fit of the drop probability versus the time of day. In the middle of the day, the model works well, whereas in the evening, a decrease in the goodness of fit measure R indicates that the model does not fit the data well.

acknowledged packet has been dropped. In the case that the acknowledgment does not indicate a drop, we denote $t(\tau_k, 0)$ as the time in which the packet was sent. Hence, $RTT_{t(\tau_k,0)} = \tau_k - t(\tau_k, 0)$. In the case that the acknowledgment indicates that j packets have been dropped, we denote the time observed round-trip time as $RTT_{\tau(\tau_k,j)}$, and $t(\tau_k, i)$ for $i < j$ as the times when the dropped packets were sent.

A. Round-Trip Time Distribution

Denote the round trip time experienced by a packet sent at time t by RTT_t. At times τ_k, it is observed that a packet sent at time t experiences a round-trip latency RTT_t. According the model used, the dropped packets give no information regarding the round-trip time (see Remark II-B.1). Hence, for notational convince we restrict our attention to the case where a drop has not occurred and the observations are $RTT_{\tau(t_k,0)}$. We denote the aggregate observations as Y^{τ_k}. The objective is to determine the distribution of RTT_t given Y^{τ_k}. This is accomplished by modeling the queues between the source and destination as done in Section II-A.

As discussed in Section II-A, RTT is made up of three components;

$$RTT_t = T_{fixed} + T_{proc} + D_t,$$

where T_{fixed} accounts for fixed delays such as propagation time, transmission time, etc., T_{proc} is a positive random time, with density function u, that models the amount of time the end hosts take to respond to the packet arrival and D_t is the delay incurred at all the routers and is model as one queue. Packets are serviced by this queue at an unknown rate μ, where a lower bound, $\mu_{\min}$, and an upper bound, $\mu_{\max}$, on the service time are known[2]. The arrival of packets onto the queue is modeled as a Poisson process L_t with constant but unknown intensity λ. Furthermore, $\lambda_{\max}$, an upper bound and lower bound, $\lambda_{\min}$, on the intensity of packet arrivals are assumed to be known. The size of the packets are distributed according to a known distribution function ρ. The maximum queue size is $q_{\max}$ which is assumed to be known[3]. The queueing delay at time t is given by (5) and the queue varies according to (6).

1. First, we focus on the probability density of the queue occupancy. Define the probability density $p\left(q_t, \lambda, \mu \middle| Y^{\tau_k}\right)$ to be the probability density that the queue occupancy at time t is q_t, incoming Poisson intensity is λ and the sending rate is μ. We initially assume a prior distribution for $p(q_0, \lambda, \mu)$ and then iteratively determine $p\left(q_{t(\tau_k,0)}, \lambda, \mu \middle| Y^{\tau_k}\right)$ based on observations y_{τ_k}. Note that in general there are two types of observations, one where an acknowledgment arrives and the round-trip time is determined and one where an acknowledgment does not arrive. If an acknowledgment does not arrive it may be due to an increase in round-trip time and could be used to update the probability distribution of the round-trip time. For example, during a sudden burst of congestion, many packets may be dropped. In order to accurately determine the state of the network during this burst, estimates of the round-trip time must be made in between acknowledgment arrivals. However, it is assumed that the congestion is not so severe that all packets are dropped for a long time, and occasionally a packet does not get dropped and provides an accurate measurement of the round-trip time. Thus we ignore the case when the observation is the lack of the arrival of an acknowledgment and assume that y_{τ_k} is the round-trip time experienced by a packet.

Bayes' theorem implies that

$$\begin{aligned} &p\left(q_{t(\tau_k,0)}, \lambda, \mu \middle| Y^{\tau_k}\right) \\ &\propto p\left(q_{t(\tau_k,0)}, \lambda, \mu \middle| Y^{\tau_{k-1}}\right) P\left(y_{\tau_k} \middle| q_{t(\tau_k,0)}, \lambda, \mu, Y^{\tau_{k-1}}\right). \end{aligned} \tag{8}$$

The observation y_{τ_k} determines that round-trip time, i.e. $RTT_{t(\tau_k,0)} = \tau_k - t(\tau_k, 0)$. Since $T_{proc} = RTT_{t(\tau_k,0)} - T_{fixed} - \mu q_{t(\tau_k,0)}$, we have that y_{τ_k}indicates that $T_{proc} = \tau_k - t(\tau_k, 0) - T_{fixed} - \mu q_{t(\tau_k,0)}$, hence

$$\begin{aligned} &P\left(y_{\tau_k} \middle| q_{t(\tau_k,0)}, \lambda, \mu, Y^{\tau_{k-1}}\right) \\ &= u\left(\tau_k - t(\tau_k, 0) - \mu q_{t(\tau_k,0)} - T_{fixed}\right), \end{aligned}$$

where u is the distribution of T_{proc}, for example see (2). Now, define $p\left(q, \lambda, \mu, t \middle| Y^{\tau_{k-1}}\right) := p\left(q_t, \lambda, \mu \middle| Y^{\tau_{k-1}}\right)$, then, by (6),

$$\begin{aligned} &\frac{\partial}{\partial t} p\left(q, \lambda, \mu, t \middle| Y^{\tau_{k-1}}\right) = \mu \frac{\partial}{\partial q} p\left(q, \lambda, \mu, t \middle| Y^{\tau_{k-1}}\right) \\ &+ \lambda \int_0^1 p\left(q - z, \lambda, \mu, t \middle| Y^{\tau_{k-1}}\right) \rho(dz)\, 1_{\{q>z\}} \\ &- \lambda p\left(q, \lambda, \mu, t \middle| Y^{\tau_{k-1}}\right) \end{aligned} \tag{9}$$

[2]Note that $\mu_{\max}$ can easily be found. For example, if the sender sends a packet with size one and suffers delay RTT, then $\mu < RTT$. Hence, we taken $\mu_{\max} = \min(RTT)$.

[3]The actual value of $q_{\max}$ does not appear to greatly effect the results.

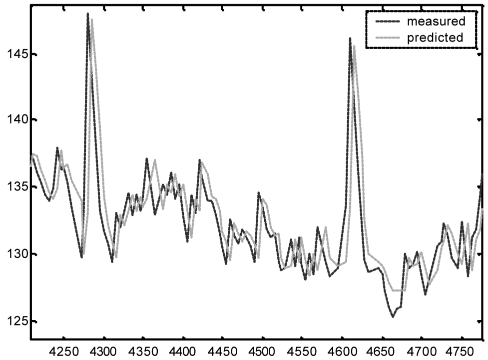


Fig. 5. The above plot shows the actual round-trip time and the one step ahead predicted value of the round-trip time.

with $p\left(q,\lambda,\mu,0\middle|Y^{\tau_{k-1}}\right) = p\left(q_{t(\tau_k,0)},\lambda,\mu\middle|Y^{\tau_{k-1}}\right)$. Note that (9) is closely related the Takács integrodifferential equation [18].

With joint distribution of the queue occupancy, sending and arrival rates known, the distribution of the queuing delay can be computed via

$$P\left(D_t \in A\middle|Y^{\tau_k}\right) = \int_{\lambda_{\min}}^{\lambda_{\max}} \int_{\mu_{\min}}^{\mu_{\max}} P\left(\mu q \in A, d\lambda, d\mu\middle|Y^{\tau_k}\right). \tag{10}$$

Finally, the density of the round-trip time is

$$\begin{aligned} &p\left(RTT_t\middle|Y^{\tau_k}\right) \\ &= \int_0^{\mu_{\max}q_{\max}} u\left(RTT_t - D_t - T_{fixed}\right) p\left(D_t\middle|Y^{\tau_k}\right) dD_t. \end{aligned} \tag{11}$$

The above technique was applied to experiments. The data from 16 hop connection between Los Angeles and Washington DC was analyzed. Figure 5 shows both the actual round-trip time and the predicted round-trip time. While the prediction is good, the mean predicted value is nearly that same as the last value of the round-trip time. However, it should be emphasized that the techniques described above provides not just the mean prediction, but the entire distribution of the predicted round-trip time. Conceivably further improvements can be had by using more sophisticated models for λ. In particular, allowing λ to vary according to a Markov chain allows the model to account for bursts of packet arrivals. Such approaches will be explored in future papers.

B. Drop Event Probability

As discussed in Section II-B, it is assumed that drop events can be model as a doubly stochastic Poisson process N with drops occurring with intensity $n\left(V_t,\theta_t,RTT_t\right)$, where V_t is the controlled flow (CF) sending rate, θ_t is an abstract variable representing the level of congestion and RTT_t is the round-trip time. It is further assumed that θ is a homogeneous Markov jump process with known the intensity matrix Λ and with known initial distribution. The objective in this section is to determine the drop probability given a sequence of observations of acknowledgments arriving at discrete times τ_k. We denote the observation at time τ_k as y_{τ_k} and the aggregate of all observation up to time τ_k as Y^{τ_k}. We initially assume a prior distribution for the congestion level $p\left(\theta_0\right)$. Then, based on observations, we determine the distribution $P\left(\theta_{t(\tau_k,0)}\middle|Y^{\tau_k}\right)$. Note that the observations take the form of either acknowledgment or non-acknowledgments. As in the previous section, we only examine the positive observation case where the observation at time τ_k indicates an acknowledgment has arrived. Furthermore, it is assumed that V_t is known for all times t.

We first consider the case where the observation y_{τ_k} indicates that no drop has occurred. The observation gives complete information about the round-trip time at time $RTT_{t(\tau_k,0)}$. Thus

$$\begin{aligned} &P\left(\theta_{t(\tau_k,0)}\middle|V_{t(\tau_k,0)},Y^{\tau_k}\right) \\ &= P\left(\theta_{t(\tau_k,0)}\middle|RTT_{t(\tau_k,0)},V_{t(\tau_k,0)},Y^{\tau_k}\right). \end{aligned}$$

Furthermore, drops are assumed to be independent and given the round-trip time, CF sending rate, $V_{t(\tau_k,0)}$, and $\theta_{t(\tau_k,0)}$, the probability of a drop is given by $g\left(\theta_{t(\tau_k,0)},RTT_{t(\tau_k,0)},V_{t(\tau_k,0)}\right)$ (See Section II-B for details). In particular, since the observation y_{τ_k} indicates a drop has not occurred

$$\begin{aligned} &P\left(y_{\tau_k}\middle|\theta_{t(\tau_k,0)},RTT_{t(\tau_k,0)},V_{t(\tau_k,0)},Y^{t_{k-1}}\right) \\ &= 1 - g\left(\theta_{t(\tau_k,0)},RTT_{t(\tau_k,0)},V_{t(\tau_k,0)}\right). \end{aligned}$$

Furthermore, we assume that given the history $Y^{\tau_{k-1}}$, the future values of RTT and θ are independent, hence $P\left(\theta_{t(\tau_k,0)} = i\middle|RTT_{t(\tau_k,0)},Y^{\tau_{k-1}}\right) = P\left(\theta_{t(\tau_k,0)} = i\middle|Y^{\tau_{k-1}}\right)$. Thus

$$\begin{aligned} &P\left(\theta_{t(\tau_k,0)} = i\middle|V_{t(\tau_k,0)},Y^{\tau_k}\right) \\ &= P\left(\theta_{t(\tau_k,0)} = i\middle|RTT_{t(\tau_k,0)},V_{t(\tau_k,0)},Y^{\tau_k}\right) \\ &\propto P\left(y_{\tau_k}\middle|\theta_{t(\tau_k,0)} = i,RTT_{t(\tau_k,0)},V_{t(\tau_k,0)},Y^{\tau_{k-1}}\right) \\ &\times P\left(\theta_{t(\tau_k,0)} = i\middle|RTT_{t(\tau_k,0)},V_{t(\tau_k,0)},Y^{\tau_{k-1}}\right) \\ &= \left(1 - g\left(i,RTT_{t(\tau_k,0)},V_{t(\tau_k,0)}\right)\right) \\ &\times P\left(\theta_{t(\tau_k,0)} = i\middle|RTT_{t(\tau_k,0)},V_{t(\tau_k,0)},Y^{\tau_{k-1}}\right) \\ &= \left(1 - g\left(i,RTT_{t(\tau_k,0)},V_{t(\tau_k,0)}\right)\right) \\ &\times P\left(\theta_{t(\tau_k,0)} = i\middle|V_{t(\tau_k,0)},Y^{\tau_{k-1}}\right). \end{aligned}$$

We assume that θ is independent of the CF sending rate, hence

$$\begin{aligned} &P\left(\theta_{t(\tau_k,0)} = i\middle|V_{t(\tau_k,0)},Y^{\tau_{k-1}}\right) \\ &= P\left(\theta_{t(\tau_k,0)} = i\middle|Y^{\tau_{k-1}}\right). \end{aligned}$$

Thus

$$P\left(\theta_{t(\tau_k,0)}=i\middle|\,V_{t(\tau_k,0)},Y^{\tau_{k-1}}\right)= \tag{12}$$

$$\sum_j\left(P\left(\theta_{t(\tau_k,0)}=i\middle|\,\theta_{t(\tau_{k-1},*)}=j\right)\right. \tag{13}$$

$$\left.\times P\left(\theta_{t(\tau_{k-1},*)}=j\middle|\,Y_{\tau_{k-1}}\right)\right)= \tag{14}$$

$$\sum_j\left(P\left(\theta_{t(\tau_k,0)}=i\middle|\,\theta_{t(\tau_{k-1},*)}=j\right)\right.$$

$$\left.P\left(\theta_{t(\tau_{k-1},*)}=j\middle|\,V_{t(\tau_{k-1},0)},Y_{\tau_{k-1}}\right)\right) \tag{15}$$

where $P\left(\theta_{t(\tau_k,0)}=i\middle|\,\theta_{t(\tau_{k-1},*)}=j\right)$ is the (i,j)element of the matrix $\exp\left(\left(t\left(\tau_k,0\right)-t\left(\tau_{k-1},*\right)\right)\Lambda\right)$.

Now, suppose that y_{τ_k} indicates that the packet that was sent at $t\left(\tau_k,0\right)$ was dropped. In this case the round-trip time experienced by the drop packet is not known, that is, the queueing delay at time $t\left(\tau_k,0\right)$ and processing delay are not known. Thus

$$P\left(\theta_{t(\tau_k,0)}=i\middle|\,V_{t(\tau_k,0)},Y^{\tau_k}\right) \tag{16}$$

$$=\int P\left(\theta_{t(\tau_k,0)}=i\middle|\,RTT_{t(\tau_k,0)},V_{t(\tau_k,0)},Y^{\tau_k}\right) \tag{17}$$

$$\times\, dP\left(RTT_{t(\tau_k,0)}\middle|\,V_{t(\tau_k,0)},Y^{\tau_k}\right), \tag{18}$$

where

$$\begin{aligned}
&P\left(\theta_{t(\tau_k,0)}=i\middle|\,RTT_{t(\tau_k,0)},V_{t(\tau_k,0)},Y^{\tau_k}\right)\\
&\propto P\left(y_{\tau_k}\middle|\,\theta_{t(\tau_k,0)}=i,RTT_{t(\tau_k,0)},V_{t(\tau_k,0)},Y^{t_{k-1}}\right)\\
&\times P\left(\theta_{t(\tau_k,0)}=i\middle|\,RTT_{t(\tau_k,0)},V_{t(\tau_k,0)},Y^{t_{k-1}}\right)\\
&=g\left(i,RTT_{t(\tau_k,0)},V_{t(\tau_k,0)}\right)\\
&\times P\left(\theta_{t(\tau_k,0)}=i\middle|\,V_{t(\tau_k,0)},Y^{t_{k-1}}\right).
\end{aligned}$$

Based on the observations Y^{τ_k}, the distribution, $P\left(RTT_{t(\tau_k,0)}\middle|\,Y^{\tau_k}\right)$, can be found as in Section III-A. Furthermore since the CF sending rate is independent of the round-trip time

$$P\left(RTT_{t(\tau_k,0)}\middle|\,V_{t(\tau_k,0)},Y^{\tau_k}\right)=P\left(RTT_{t(\tau_k,0)}\middle|\,Y^{\tau_k}\right).$$

Now suppose, that the observation y_{τ_k} indicates that the packet sent at $t\left(\tau_k,1\right)$ was also dropped. In this case y_{τ_k} holds information about two drops. We process information about each drop sequentially,

$$\begin{aligned}
&P\left(\theta_{t(\tau_k,1)}=i\middle|\,\begin{array}{c}V_{t(\tau_k,0)},Y^{\tau_k},\text{packet sent}\\ \text{at } t\left(\tau_k,0\right)\text{ was dropped}\end{array}\right)=\\
&\int P\left(\theta_{t(\tau_k,1)}=i\middle|\,\begin{array}{c}RTT_{t(\tau_k,1)},V_{t(\tau_k,0)},Y^{\tau_k},\text{packet}\\ \text{sent at } t\left(\tau_k,0\right)\text{ was dropped}\end{array}\right)\\
&\times\, dP\left(RTT_{t(\tau_k,1)}\middle|\,\begin{array}{c}V_{t(\tau_k,0)},Y^{\tau_k},\text{packet sent}\\ \text{at } t\left(\tau_k,0\right)\text{ was dropped}\end{array}\right),
\end{aligned}$$

Since the round-trip time is independent of drops

$$\begin{aligned}
&P\left(RTT_{t(\tau_k,1)}\middle|\,\begin{array}{c}V_{t(\tau_k,0)},Y^{\tau_k},\text{packet sent}\\ \text{at } t\left(\tau_k,0\right)\text{ was dropped}\end{array}\right)\\
&=P\left(RTT_{t(\tau_k,1)}\middle|\,V_{t(\tau_k,0)},Y^{\tau_k}\right).
\end{aligned}$$

Furthermore,

$$\begin{aligned}
&P\left(\theta_{t(\tau_k,1)}=i\middle|\,\begin{array}{c}RTT_{t(\tau_k,1)},V_{t(\tau_k,0)},Y^{\tau_k},\text{packet}\\ \text{sent at } t\left(\tau_k,0\right)\text{ was dropped}\end{array}\right)\propto\\
&P\left(y_{\tau_k}\middle|\,\begin{array}{c}\theta_{t(\tau_k,1)}=i,RTT_{t(\tau_k,1)},V_{t(\tau_k,0)},Y^{t_{k-1}},\\ \text{packet sent at } t\left(\tau_k,0\right)\text{ was dropped}\end{array}\right)\times\\
&P\left(\theta_{t(\tau_k,1)}=i\middle|\,\begin{array}{c}RTT_{t(\tau_k,1)},V_{t(\tau_k,0)},Y^{t_{k-1}},\\ \text{packet sent at } t\left(\tau_k,0\right)\text{ was dropped}\end{array}\right)\\
&=g\left(i,RTT_{t(\tau_k,1)},V_{t(\tau_k,1)}\right)\times\\
&P\left(\theta_{t(\tau_k,1)}=i\middle|\,\begin{array}{c}V_{t(\tau_k,0)},Y^{t_{k-1}},\text{packet sent}\\ \text{at } t\left(\tau_k,0\right)\text{ was dropped}\end{array}\right).
\end{aligned}$$

where

$$\begin{aligned}
&P\left(\theta_{t(\tau_k,1)}=i\middle|\,\begin{array}{c}V_{t(\tau_k,0)},Y^{t_{k-1}},\text{packet sent}\\ \text{at } t\left(\tau_k,0\right)\text{ was dropped}\end{array}\right)=\\
&\sum_j\left(P\left(\theta_{t(\tau_k,1)}=i\middle|\,\theta_{t(\tau_{k-1},0)}=j\right)\right.\\
&\left.\times\, P\left(\theta_{t(\tau_k,0)}=i\middle|\,\begin{array}{c}V_{t(\tau_k,0)},Y^{t_{k-1}},\text{packet sent}\\ \text{at } t\left(\tau_k,0\right)\text{ was dropped}\end{array}\right)\right)
\end{aligned}$$

and

$$P\left(\theta_{t(\tau_k,0)}=i\middle|\,\begin{array}{c}V_{t(\tau_k,0)},Y^{t_{k-1}},\text{packet sent}\\ \text{at } t\left(\tau_k,0\right)\text{ was dropped}\end{array}\right)$$

is given by (16). Notice that we have used the assumption that drops are independent, hence

$$\begin{aligned}
&P\left(y_{\tau_k}\middle|\,\begin{array}{c}\theta_{t(\tau_k,1)}=i,RTT_{t(\tau_k,1)},V_{t(\tau_k,0)},Y^{t_{k-1}},\\ \text{packet sent at } t\left(\tau_k,0\right)\text{ was dropped}\end{array}\right)=\\
&g\left(i,RTT_{t(\tau_k,1)},V_{t(\tau_k,1)}\right).
\end{aligned}$$

This process is repeated for each drop indicated by y_{τ_k}.

IV. Conclusion

Models and techniques for network tomography have been presented. These techniques provide a mechanism to determine a detailed picture of the state of a segment of the network. These techniques are currently being used in "live" and simulated network experiments. Early results are encouraging, but point out the clearly anticipated computational problems. In particular, the model for round-trip times have many parameters and it is difficult to efficiently process the data. However, as experienced is gained in initial probability distributions and Markov Chain Monte Carlo computational techniques are utilized, these difficulties are likely to be overcome.

Future work will focus on how to best combine the information of multiple host pairs. For example, in the case of a sender sending data to two distinct hosts, the two paths may have some common links. The information from both connections can be utilized to better estimate the router states of the network along these common links and, hence, better estimate the router states along the entire paths. Once experienced is gained on the typical variation of arrival rates and drop probabilities, techniques such as change point detection and sequential analysis [19] can be applied to quickly detect changes in the models and hence quickly detect attacks.

References

[1] Http://www.ihr.daze.net/.

[2] Http://www.InternetTrafficReport.com.

[3] Http://www.mids.org/weather.

[4] Http://www.netstatsys.com.

[5] Http://www.noc.ucla.edu/networking/weather.html.

[6] E. Altman, K. Avrachenkov, and C. Barakat. TCP in the presence of bursty losses. In *ACM SIGMETRICS*, June 2000.

[7] E. Altman, K. Avrachenkov, and C. Barakat. A stochastic model of TCP/IP with stationary random losses. In *ACM Sigcomm*, Sep. 2000.

[8] R. Caceres, N. Duffield, S. B. Moon, and D. Towsley. Inference of internal loss rates in the MBone. In *Proc. IEEE Global Internet '99*, Rio de Janeiro, Brazil, Dec. 1999.

[9] A. Feldman, A. Gilbert, P. Huang, and W. Willinger. Dynamics of IP traffic: A study of the role of variability and the impact on control. In *ACM SIGCOMM*, 1999.

[10] S. Floyd and V. Jacobson. Random early drop detection gateways for congestion avoidance. *IEEE/ACM Transactions on Networking*, 1:397–413, 1993.

[11] N. L. for Applied Network Research (NLANR). http://www.nlanr.net/.

[12] J. Hespanha and S. Bohacek. Preliminary results in routing games. In *American Control Conference*, Arlington, VA, June, 2001. IEEE.

[13] W. E. Leland, M. S. Taqq, W. Willinger, and D. V. Wilson. On the self-similar nature of Ethernet traffic. In *ACM Sigcomm*, pages 183–193, San Francisco, California, 1993.

[14] M. Mathis, J. Semke, J. Mahdavi, and T. Ott. The macroscopic behavior of the TCP congestion avoidance algorithm. *Computer Communication Review*, 27, 1997.

[15] V. Misra, W. Gong, and D. Towsley. Stochastic differential equation modeling and analysis of TCP-windowsize behavior. In *Performance '99*, Istanbul, Turkey, 1999.

[16] J. Padhye, V. Firoiu, D. Towsley, and J. Kurose. Modeling TCP throughput: A simple model and its empirical validation. In *ACM Sigcomm*, Sept. 1998.

[17] S. Savari and E. Telatar. The behavior of certain stochastic processes arising in window protocols. In *IEEE Globecom*, Dec. 1999.

[18] L. Takacs. Investigation to the waiting time problems by reduction to Markov processes. *Acta Math Acad. Sci. Hung*, 6:101–129, 1955.

[19] A. Wald. *Sequential Analysis*. John Wiley and Sons, Inc., 1947.

[20] L. Zhang, S. Shenker, and D. D. Clark. Observations on the dynamics of a congestion control algorithm: The effects of two-way traffic. In *ACM Sigcomm*, Sept. 1991.