

Network characteristics: modelling, measurements and admission control

Dinan Gunawardena, Peter Key, and Laurent Massoulié

Microsoft Research
7 J.J. Thomson Avenue
CB3 0FB Cambridge, United Kingdom

Abstract. Round trip delays constitute the basic network measure that can be obtained by end systems without any network support. Our aim is to design measurement-based admission control strategies for streaming applications based on such minimal feedback on network state. To this end we discuss simple statistical models of packet round trip delays across either local or wide area networks. We observe that the delay component due to queueing scales like the reciprocal of the spare capacity, at least in a ‘heavy traffic’ regime where spare capacity is scarce. Our models also allow to capture the correlations between consecutive measurements. Based on these results we propose a two-stage strategy for inferring spare capacity along a network path. We show consistency of this estimate, and analyse its asymptotic variance when the number of samples becomes large. We have experimented these strategies in a local network environment. We observe a good match between theory and practice for switched Ethernets. Surprisingly, the match deteriorates only slightly when the network path comprises hubs, although the theoretical models seem to be less applicable to such technology.

1 Introduction

In IP networks, Round Trip Time (RTT) measurements constitute the basic feedback information that end hosts can use to infer the state of the network connection between them¹. These are used in TCP Reno, explicitly to set the retransmission time out, and implicitly to trigger packet transmissions, while in TCP Vegas the variation in RTTs is monitored to stabilise the congestion window at a suitable value. Passive measurement of RTTs is considered in [8]. Many other uses of such RTT measurements have been considered, in particular for inferring the bottleneck capacity between two hosts by using active probing. Packet pair [11, 12] and pathchar [6, 14] are two popular methods for doing this.

The goal of the present work is to use such RTT measurements to infer the rate at which a real time connection between two hosts can be initiated, so that it does not disrupt existing connections. In contrast with previous work [10], [3], we do not rely on either loss occurring, or on the availability of ECN marks. To this end, we need to determine the spare capacity or available capacity, rather than the total capacity along

¹ Losses also constitute an important feedback. However their occurrence is not directly observed, but rather inferred from either triple duplicate acks, or time-outs.

the path between the nodes. Related work on available capacity estimation in the wide area considers TCP as an estimator, as well as probing estimators [7]. Also relevant is the Delphi procedure, introduced in [17] to infer characteristics of cross-traffic. In contrast to most current work on network measurements, we use stochastic models of traffic and queues. A notable exception is [1]. In addition, while prior work has concentrated on the wide area context, we are also interested in small networks.

A networked home is an example of a small network (i.e. with low number of hosts, low number of segments), where a small number of devices such as servers, PCs, PDAs and other network devices may be connected using a variety of Layer 2-technologies. Switches, hubs, wireless access points and wireless bridges may be mixed together. Measuring network characteristics for such small networks is important for applications and services such as streaming multimedia across the network. For example, the available bandwidth determines whether we can safely start a new stream between two nodes in the network, and at what rate, while delay or jitter measurement may restrict the applications that can run with acceptable quality. The size of the network poses inference problems different from those for large networks.

We consider simple models for predicting the evolution of round trip times in such LANs, and observe that the queueing component of round trip times is inversely proportional to spare capacity for a wide range of background traffic. This in turn motivates probing strategies for deciding at what rate a real time streaming application can be started between the two devices. The probing strategies have to gather information without disrupting existing traffic. We consider a strategy for inferring spare capacity motivated by the maximum likelihood estimation technique. We establish consistency of this estimator, and analyse its asymptotic behaviour. We give experimental results for this strategy in switched Ethernet and Wireless LAN scenarios. These suggest that simple, robust, non-disruptive probing strategies can be found for admission control in small LANs, using current switches, hubs and home networking technologies.

Although our initial motivation lies in home environments, we believe that the statistical modeling might prove relevant in WAN situations as well. In particular, we present an analysis of correlations between observed RTTs which we use only to predict when distinct measurements can be considered independent. However such correlations do contain information which could be used to improve other estimators, such as the pathchar method for assessing bottleneck bandwidth.

The paper is organised as follows. In Section 2 we survey classical as well as novel models for the observed RTTs. In Section 3 we discuss estimation of spare capacity from such observations. Section 4 describes our experimental set-up and Section 5 the corresponding results.

2 Statistical models for round trip delays

Consider a sender/receiver pair. Assume that at time T_n , $n \geq 0$, the sender sends a packet of size p_n bytes to the receiver, which reflects it back to the sender, where it is received at time S_n . Note $\tau_n := S_n - T_n$ the corresponding round trip time. We first make simple assumptions on the physics of the transmission path. We then review a number of simple background traffic models, and results on the corresponding delay distributions. These

models all exhibit short-range dependence, an assumption that is seldom met in practice (see [15]). However, these models are nevertheless useful in identifying qualitative relationships between spare capacity and observed delays.

2.1 Physical assumptions on the transmission path

Single bottleneck assumption: We assume that these RTTs are fully determined by some fixed propagation time τ , plus some queueing time at a bottleneck link. We thus neglect the impact of several capacity constraints along the path, and rather assume the most stringent bottleneck alone determines the queueing times. This is appropriate in a LAN environment, where multiple bottlenecks rarely occur.

Bottleneck model: The bottleneck is supposed to be a queue with capacity C bytes/second. In addition to the packets used for measurements, the bottleneck receives background traffic. Denote by $A(s, t)$ the total amount of such background traffic received between times s and t . Let $W(t)$ denote the workload at the bottleneck at time t . We assume FIFO queueing, so that the observed transmission times read:

$$\tau_n = \tau + \frac{W(T_n^-) + p_n}{C}. \quad (1)$$

This representation relies on the assumption that the propagation time τ takes place between the bottleneck and the receiver. The impact of finite buffer capacity is discussed in 2.2 below.

2.2 Background traffic models

Unreactive, Poisson stream of packets: Assume that the background traffic consists in packets with independent, identically distributed (i.i.d.) sizes, that reach the bottleneck link at the instants of a Poisson process. Then the Pollaczek-Khinchin formula (see e.g. [2]) gives the mean value of the stationary buffer content distribution. In particular, the average waiting time of a probe packet reads

$$\bar{D} = \frac{\lambda}{C} \frac{\overline{X^2}}{2(C - m)}, \quad (2)$$

where $\overline{X^2}$ is the mean of the squared packet size, λ is the packet arrival rate per second, and m is the load on the queue, given by $\lambda\bar{X}$ where \bar{X} is the mean packet size. Note the inverse dependence on $C - m$.

Unreactive, white noise background traffic: Another convenient traffic model is as follows. Assume one has

$$A(s, t) = m(t - s) + \sigma(B(t) - B(s)), \quad (3)$$

where m is the mean arrival rate, σ describes the burstiness, and $B(t)$ is a standard Brownian motion (Wiener process). Such a traffic model arises naturally as a heavy

traffic limit (see e.g. Reiman [16]) of a queue with short range dependent traffic. Queues with Brownian input have been extensively studied; see for instance the book [5]. For the model under consideration, the conditional distribution of the workload $W(t+h)$ at time $t+h$ given the workload at some time t is given by

$$\mathbf{P}(W(t+h) \leq x | W(t) = w) = \Phi\left(\frac{(C-m)h + x - w}{\sigma\sqrt{h}}\right) - e^{-2(C-m)x/\sigma^2} \Phi\left(\frac{(C-m)h - x - w}{\sigma\sqrt{h}}\right), \quad x \geq 0, \quad (4)$$

where $x \in \mathbb{R}$ and Φ is the cumulative distribution function of a standard normal random variable, i.e.

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy.$$

In particular, when $C > m$, it can be seen by letting h tend to infinity in the above formula that the workload $W(t)$ admits a limiting distribution given by

$$\mathbf{P}(W(t) \leq x) = \mathbf{1}_{x \geq 0} \left\{ 1 - e^{-2(C-m)x/\sigma^2} \right\}, \quad (5)$$

an exponential distribution with mean $\sigma^2/2(C-m)$. Note again the inverse dependence on the spare capacity, $C-m$. Although Equation (4) does in principle capture the correlations between the observed round trip times, it is not straightforward to derive a suitable notion of correlation time from it. This will be made easier from the analysis to follow.

Unreactive white noise, with finite buffer capacity: We now modify the previous model to incorporate a finite buffer memory. The workload process, W_t , obeys the diffusion equation

$$dW_t = -(C-m) dt + \sigma dB_t,$$

with reflection at 0 and b , where b represents the maximal buffer content. Here, B is a standard Brownian motion, $-(C-m)$ is the drift, and σ is as before a positive parameter describing the background traffic burstiness. The previous model is a limiting case of this one, obtained by letting b go to infinity. The results to follow are derived in the Appendix. Define the functions y_k and the numbers λ_k by letting

$$\begin{aligned} y_0(x) &\equiv 1, \quad y_k(x) = e^{dx/\sigma^2} \left(d \sin\left(\frac{k\pi x}{b}\right) - \frac{k\pi\sigma^2}{b} \cos\left(\frac{k\pi x}{b}\right) \right), \\ \lambda_0 &= 0, \quad \lambda_k = \frac{1}{2\sigma^2} \left(d^2 + \frac{k^2\pi^2\sigma^4}{b^2} \right), \quad k \in \mathbb{Z}_+, \end{aligned} \quad (6)$$

where $d = C - m$. One then has:

Theorem 1. *The probability density of W_t conditionally on W_0 is specified by*

$$\pi_t(x | W_0 = z) = \mathbf{1}_{[0,b]}(x) \sum_{k \geq 0} \frac{1}{\beta_k} y_k(x) e^{-2dx/\sigma^2} y_k(z) e^{-\lambda_k t}, \quad (7)$$

where

$$\beta_0 = \frac{1 - e^{-2db/\sigma^2}}{2d/\sigma^2}, \beta_k = \frac{b}{2} \left[d^2 + \frac{k^2 \pi^2 \sigma^4}{b^2} \right], k > 0. \quad (8)$$

The stationary distribution is obtained by letting t go to infinity; it reads

$$\pi_{stat}(x) = \mathbf{1}_{[0,b]}(x) \frac{2d/\sigma^2}{1 - e^{-2db/\sigma^2}} e^{-2dx/\sigma^2}. \quad (9)$$

One of the insights that we derive from this result is as follows. The exponents λ_k capture the speed at which the influence of the initial value W_0 on the distribution of W_t vanishes. In particular, for times t large compared to the reciprocal of λ_1 , the smallest non-zero λ_k , this influence vanishes. In this sense, that result identifies a correlation time scale for the process W_t as

$$\frac{1}{\lambda_1} = \frac{2\sigma^2}{(C-m)^2 + \frac{\pi^2 \sigma^4}{b^2}}.$$

In the case where the buffer limit b is large, this critical time scale also reads

$$T_{critical} = \frac{2\sigma^2}{(C-m)^2}. \quad (10)$$

We note that this model can be used to predict the occurrence of loss: indeed, a packet of size p_n would be lost according to this model when it reaches the queue and finds there a workload W such that $W + p_n$ is greater than the buffer limit b .

Poisson unreactive + reactive traffic: Now, to directly model TCP-like traffic, we assume a fixed number of packets, N say, queue at the bottleneck link together with unreactive traffic with FIFO service. After service at the bottleneck, each packet undergoes a random delay, with given mean $\bar{\delta}$, and then re-enters the bottleneck queue. Effectively these reactive packets are subject to fixed size window control. One could for instance view them as packets belonging to a number of ongoing TCP connections, assuming that the sum of the congestion windows of these TCP connections is roughly constant, and equal to N . This does not account for losses, and would be appropriate in a situation where the bottleneck buffer is large and the TCP window sizes saturate at the receive window limit.

We assume as above that unreactive packets reach the bottleneck queue at the instants of a Poisson process, with intensity λ .

For tractability, we assume that the service time of packets of both reactive and unreactive traffic at the bottleneck queue is exponentially distributed with mean denoted by \bar{X} . The following result is established in the Appendix.

Proposition 1. *Let $\pi(k, l)$ denote the stationary probability that there are k unreactive, and l reactive packets queued at the bottleneck. For all $k \geq 0$, all $l \in \{0, \dots, N\}$, one has*

$$\pi(k, l) = \frac{1}{Z} \binom{k+l}{l} a^k b^l \frac{1}{(N-l)!}, \quad (11)$$

where

$$a = \frac{\lambda \bar{X}}{C}, \quad b = \frac{\bar{X}}{\delta C}, \quad Z = \frac{1}{1-a} \sum_{l=0}^N \left(\frac{b}{1-a} \right)^l \frac{1}{(N-l)!}. \quad (12)$$

In the heavy traffic limit where $1-a$ goes to 0, the stationary waiting time D satisfies the following weak convergence result:

$$\frac{C(1-a)}{\bar{X}} D \xrightarrow{W} \Gamma(N+2, 1), \quad (13)$$

where $\Gamma(N+2, 1)$ is the Gamma distribution with parameters $(N+2, 1)$, in other words it is distributed as the sum of $N+1$ independent random variables, with exponential distribution and mean 1.

We interpret this result as follows. The impact of reactive traffic is to increase queueing delays in a *multiplicative* way, via the parameter $N+1$. More importantly, its introduction does not change the qualitative way in which spare capacity, captured by $C(1-a)$, affects delays: these are still (at least in the heavy traffic limit) proportional to the reciprocal of the spare capacity. We expect this qualitative relationship to be preserved, at least in the heavy traffic regime, for more general packet service time distributions and short range dependent exogeneous packet arrival processes.

Modelling hubs and wireless access points Hubs behave as Ethernet segments, hence CSMA/CD models can be used to approximate their performance. Informally, the waiting time a packet sees is comprised of queueing time waiting for any packets ‘in system to be cleared’, plus the amount of time waiting for retransmissions. An analysis of CSMA/CD (see e.g. [2] p. 318) gives the waiting time as a deviation from from the Pollaczek-Khinchine formula

$$\bar{D} = \frac{\lambda \bar{X}^2 + \beta(4.62 + 2\lambda)}{2(1 - \lambda(1 + 3.31\beta))}, \quad (14)$$

where we assume that the expected packet size \bar{X} and the capacity C are normalised to 1, and β is the duration of a minislot (the time for a signal to propagate along a segment and be detected). This is small, and ignoring this term gives exactly the Pollaczek-Khinchine formula. Wireless uses CSMA/CA, and again we have approximately that the waiting time is proportional to $1/(C-m)$.

3 Spare capacity estimation

Our aim is to infer at what rate a real time connection can be established between two hosts, while preserving some Quality of Service constraint. If C is the capacity of the bottleneck link (or equivalent link), and the current load on the link is m , then the spare capacity is defined as $C-m$. However, this definition requires some clarification: stationarity, or ‘local’ stationarity is implicit in the definition, as is some assumption of timescale for which this holds. We shall take a more specific definition of available capacity, and define layer- n available capacity to be the maximum n -layer throughput

that can be achieved by *unreactive* traffic, without damaging existing unreactive traffic. We now describe the model for round trip times that we adopt based on the previous section. We then analyse a specific estimator of spare capacity derived by applying the maximum likelihood method, and then discuss alternative estimates and corresponding tests.

3.1 Measurement model

We assume that the packet send times T_n are sufficiently separated so that the sequence of workload contents $W(T_n^-)$ appearing in the expressions of the measured transmission times (1) are independent random variables. Relying on both the white noise traffic model, and the heavy traffic limit derived in Proposition 1, we assume that the observed round trip times τ_n read

$$\tau_n = \tau + \frac{1}{C - m} Z_n, \quad (15)$$

where the random variables Z_n are i.i.d., and follow a Gamma distribution with unknown parameters, that do not depend on the background load, m . In addition to these measurements, we shall take another series of round trip time measurements, τ'_n , while sending some additional unreactive traffic. This extra load, noted δm , is known to us. Noting m' the new offered load, i.e. $m' = m + \delta m$, we then have

$$\tau'_n = \tau + \frac{1}{C - m'} Z'_n,$$

where the Z'_n are i.i.d., and drawn from the same Gamma distribution as the Z_n . The parameters of this Gamma distribution, which we shall denote by $N + 1$ and B , as well as τ and C are a priori unknown. In the remainder of this section, we assume that there are k measurements in each sample set. We now consider the estimation of $C - m$ from these samples.

3.2 Maximum-Likelihood estimation

Consider the Log-likelihood L of the samples $\tau_1, \dots, \tau_k, \tau'_1, \dots, \tau'_k$. This reads

$$L = \sum_{n=1}^k -2 \log(\Gamma(N + 1)) + N \log((\tau_n - \tau)(\tau'_n - \tau)) + (N + 1) \log(B^2(C - m)(C - m')) - B((C - m)(\tau_n - \tau) + (C - m')(\tau'_n - \tau)).$$

By differentiating, it is easy to see that the maximum likelihood estimate of the ratio $(C - m)/(C - m')$ is given by $(\bar{\tau}' - \hat{\tau})/(\bar{\tau} - \hat{\tau})$, where $\bar{\tau}$, $\bar{\tau}'$ are the sample means $k^{-1} \sum_{n=1}^k \tau_n$, $k^{-1} \sum_{n=1}^k \tau'_n$ respectively, and $\hat{\tau}$ is the maximum likelihood estimate of the shift parameter τ . Consequently, the maximum likelihood estimate of $C - m$ is given by

$$\delta m \frac{\bar{\tau}' - \hat{\tau}}{\bar{\tau}' - \bar{\tau}}.$$

As maximum likelihood estimates have desirable properties (in particular, asymptotic efficiency; see [4]) we shall rely on this specific form to infer the spare capacity. Unfortunately, no analytical expression for the maximum likelihood estimate $\hat{\tau}$ is available. We shall therefore take $\hat{\tau} = \min_{1 \leq n \leq k} (\tau_n)$ instead².

Introduce the relative estimation error e by letting:

$$\widehat{C - m} = \delta m \frac{\bar{\tau}' - \hat{\tau}}{\bar{\tau}' - \bar{\tau}} =: (C - m)(1 - e). \quad (16)$$

We then have the following, which is established in the appendix:

Proposition 2. *The estimator (16) of $C - m$ is asymptotically consistent, i.e. the error e goes to zero in probability as $k \rightarrow \infty$. Assume that the random variables Z_n follow a Gamma distribution with parameters $(N + 1, B)$, where $N + 1, B$ are two positive parameters. If N is larger than 1, as $k \rightarrow \infty$, the following weak convergence holds:*

$$k^{1/(N+1)} e \xrightarrow{W} \frac{C - m'}{C - m} X, \quad (17)$$

where X is a random variable with a Weibull distribution, specified by the complementary distribution function $\mathbf{P}(X > t) = \exp(-t^{N+1}/\Gamma(N+2))$.

If N is smaller than 1, as $k \rightarrow \infty$, the following weak convergence holds:

$$\sqrt{k} e \xrightarrow{W} \frac{C - m'}{\delta m} 2(N+1)N(0, 1), \quad (18)$$

where $N(0, 1)$ denotes the standard Gaussian distribution.

This shows, on the one hand, that direct estimation of the spare capacity $C - m$ is possible without explicitly inferring the other model parameters. On the other hand, when N is larger than 1, the mean square error goes to zero as $k^{-2/(N+1)}$. This is slower than the classical scaling k^{-1} (which occurs for $N < 1$), which suggests that better estimation techniques might exist.

3.3 A general class of estimators

The previous section has motivated specific estimators of spare capacity. In general, we may define an estimator $S(\tau)$ of the vector $\tau = (\tau_1, \dots, \tau_k)$ to be such that $(C - m)S$ is consistent for either the scale, or the location of Z . In the example above, $S(\tau) = \bar{\tau} - \tau_{\min}$ is a consistent estimator for the location of Z rescaled by $1/(C - m)$, and the standard deviation of τ is consistent for the scale. More generally we may use robust estimates of location or scale instead of just the mean or standard deviation, such as the trimmed mean or median, or other more refined M-estimates or R-estimates for the location, and similarly for the scale. We then have

$$R_{\tau, \tau'} = \frac{S(\tau)}{S(\tau')} \rightarrow \frac{C - m'}{C - m}, \quad (19)$$

² Taking the minimum over both sets of samples should in principle yield a better estimate. We use the minimum over one sample instead because it simplifies the analysis to follow, and does not change its conclusions significantly.

and hence

$$\widehat{C - m} = \frac{\delta m}{1 - R_{\tau, \tau'}}, \quad (20)$$

and

$$\widehat{C - m'} = \delta m \frac{R_{\tau, \tau'}}{1 - R_{\tau, \tau'}} = \delta m \frac{S(\tau)}{S(\tau') - S(\tau)}. \quad (21)$$

This provides a general method for obtaining estimates of spare capacity under the model (15), which does not rely on the fact that the random variables Z_n, Z'_n follow a Gamma distribution, but applies to any arbitrary distribution instead.

These estimates lead naturally to equivalent tests. For example testing whether $C - m' > r \delta m$ can be based on comparing $R_{\tau, \tau'} / (1 - R_{\tau, \tau'})$ to r .

4 Experimental Framework and Implementation

Probing traffic can be generated by different protocol layers. For example, UDP packets could be used to generate the probing traffic. This is the protocol layer of the traffic of interest, however an implementation would then require software alteration to both the sender and receiver (a symmetric solution), and also confounds network delays with delays caused by the operating system. Using ICMP messages (which are handled at a very low level in most IP stacks) allows us to have an asymmetric solution, with measurement and probing software only required on the probing machine. The asymmetric solution is attractive for typical legacy home networks, since the probing and admission control only needs to be added to devices which act as servers or senders. To be able to reduce the effect of measurement errors and timer granularity, for small networks where the round trip time is small, we need better resolution than provided by a standard user-mode ICMP ping, and hence needed to use a kernel mode driver for accurate timestamping. Because we are looking at LANs, we avoid the problems that ICMP has in the wide area.

Our probing strategies used ICMP ping packets to generate the probing traffic and measure the network, with a kernel mode driver handling the basic probing functions. For the ping probes, echo request packet sequences were created and time-stamped by the processor cycle counter before transmission. Incoming echo replies are captured and time-stamped, giving the RTT to microsecond accuracy.

An example test configuration is shown in Figure 1, which shows a two switch (or two hub) network, with potential contention caused by traffic from the background source and the probing traffic sharing a common link. In this example, the switches (or hubs) create a single bottleneck, where the bottleneck could be between the two switches, or, if the interconnects are hubs, between the PCs and the hub. Background traffic was UDP traffic, generated at a specific nominal rate (bursts at 20ms intervals), and TCP traffic. A wireless scenario is shown in the bottom half of Figure 1, where a 802.11b Wireless Access Point is connected to wireless bridges. The switches were configured to 10Mb/s or 100Mb/s. We only report on the 10Mb/s results here. Standard switches were used from different manufacturers. The amount of buffering within each switch is typically about 128kB, giving a maximum queuing delay of about 100ms at 10Mb/s.

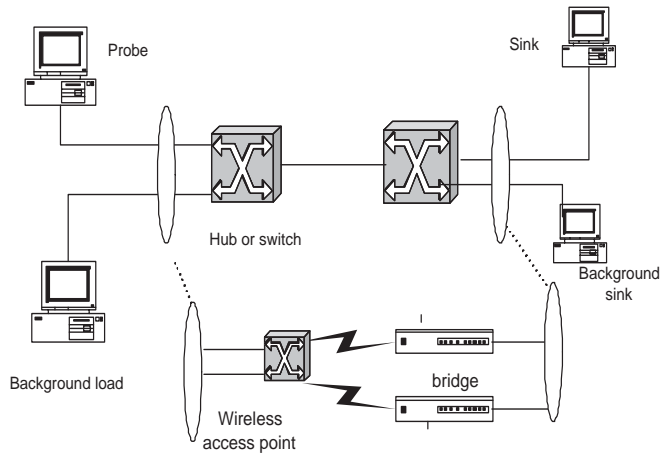


Fig. 1. Network Configuration for Experiments, showing switch/hub and wireless scenarios.

Experiments were run with different nominal background load (the bit-rate of the UDP background), with and without TCP traffic added, the background traffic being increased from 1Mb/s to 9 or 10Mb/s in 1Mb/s steps. For most of the experiments, the time between ICMPs was an exponentially distributed random variable with mean 10ms and 100ms. The two different values were used to assess the presence of correlations in the measurements.

5 Results

Consider first the switch scenario, where there are two 10Mb/s switches connecting the bottleneck link. With just UDP background load, the delay increases slightly before saturating the capacity at between 9 and 10 Mb/s. Measurements from four sets of probing, each comprising ten probes 100ms apart are summarised in the box plot in Figure 2. This plot is truncated to 200ms. In the critically loaded case (UDP load of 10Mb/s) delays as large as 500ms were observed, and 25% of the observed delays were above 350ms.

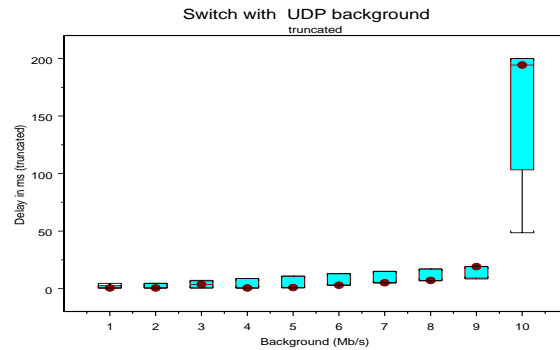


Fig. 2. Box plot of delay measurements for a switch with UDP background traffic.

With UDP background traffic alone, the transition from underload to overload is fairly sudden. In the extreme case of periodic background traffic, the delay is negligible until the switch starts to become overloaded. As the background traffic becomes more bursty, its variance increases, which increases the variability in the delay measurements (see, for example Section 2). When TCP is mixed in with UDP traffic, the delay at the ‘nominal’ background load (the load of the UDP traffic) increases, as does the variance, caused by the smearing effect of the TCP traffic, described analytically in Section 2. The corresponding box plot is shown in Figure 3. For a UDP load of 9 Mb/s, most measurements were below 200ms, while for a UDP load of 10Mb/s, delays as large as 1000ms were observed and 50% of the measurements were above 300ms, with some time-outs observed.

With probing at 100ms intervals, there is little correlation structure in the measurements: the autocorrelation function is plotted in Figure 4, corresponding to probing when there is a background load of 1 Mb/s (UDP) with TCP traffic also present, showing the only (just) significant correlation at lag 1. If we decrease the probing interval to 10ms, there is still little correlation structure evident. We do not present the corresponding picture, which is very similar to Figure 4.

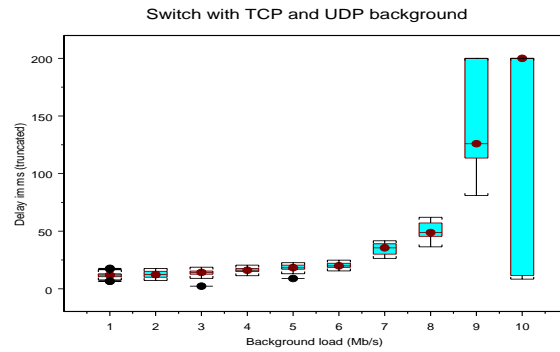


Fig. 3. Box plot of delay measurements for a switch with UDP and TCP background traffic.

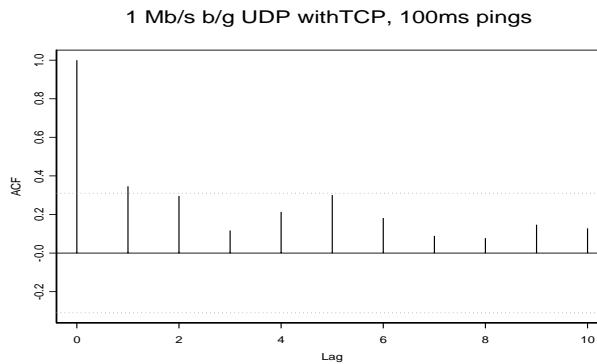


Fig. 4. Autocorrelation function of delay measurements for a switch with UDP and TCP background traffic, probes sent 100ms apart.

A kernel density plot of the RTT delay distributions for 1, 5 and 8Mb/s background UDP mixed with TCP is given in Figure 5, suggesting that the different loadings can be inferred from the delay measurements, provided the loads are not too close. Note how both the scale and location of the densities increase with increasing load, an inverse relation with spare capacity consistent with the models of Section 2. The density shapes are not inconsistent with Gamma distributions of Proposition 1 (note that the curves shown here are based on only 40 samples, and influenced by the kernel estimator – a smoother kernel makes all three density estimates unimodal).

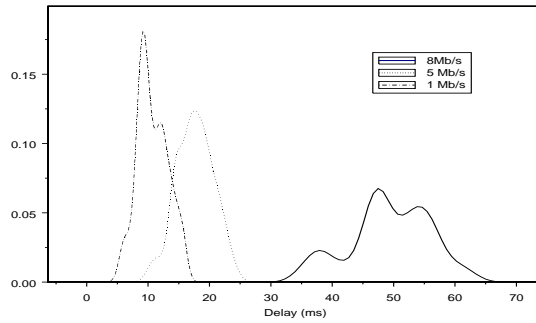


Fig. 5. Density plot for delay measurements for a switch with UDP and TCP background traffic, for 1, 5 and 8 Mb/s UDP load.

We now focus on the estimation of available capacity. The theoretical framework has motivated models and estimates where statistics of the round trip times are inversely related to the spare capacity. To see if this is borne out in practice for the adjusted means, we plot the logarithms of the adjusted means against the logarithm of the notional spare capacity, which should show a linear relationship with gradient -1. This is shown in Figure 6, which does indeed suggest a linear relationship with the correct slope. The accuracy of the estimator (16) is shown in Figure 7, which plots the (post-probe) spare capacity estimator for different loadings (shown for the load after the probe has been admitted), where the estimate is based on 4 runs each using just 10 samples. Two different values of δm are shown: for δm equal to 1Mb/s, the estimates of spare capacity are rather noisy when there is plenty of spare capacity (which one could expect in view of Figure 3), but become more accurate the more heavily loaded the network. For δm equal to 3 Mb/s, the estimates show a good match to the spare capacity.

The set of results for a hub paint a very similar picture. Figure 8 shows a box plot of the delays. Not surprisingly, the delays increase more rapidly than for the switch, but the picture for estimation purposes is similar. The log-log plot for the spare capacity and adjusted mean is shown in Figure 9, and the capacity estimate using probing with the estimator (16) is shown in Figure 10.

For a wireless access point, the delays increase much sooner (since the effective data capacity of 802.11b is about 6 Mb/s), as shown in Figure 11. The log-log plot of adjusted mean against spare capacity in Figure 12 is again roughly linear. However, the slope is about -1.55, which constitutes a significant departure from the inverse dependence on

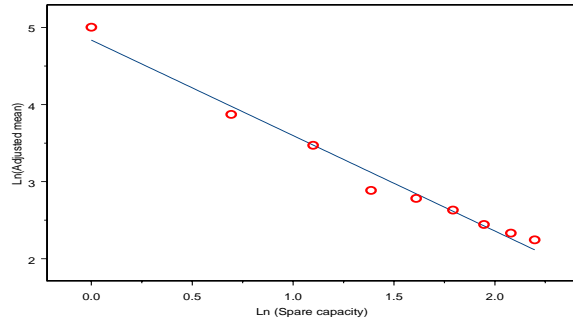


Fig. 6. Log-log plot of spare capacity against adjusted means, for 10Mb/s Switch with UDP and TCP background traffic.

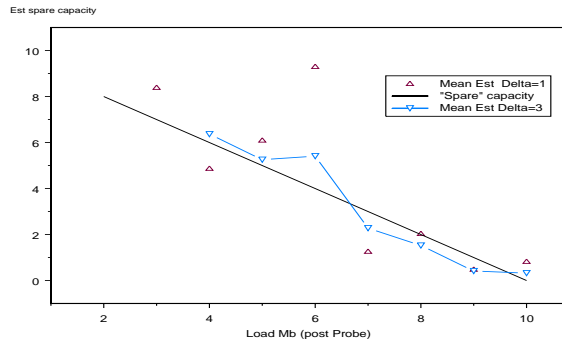


Fig. 7. Estimated (post probe) spare capacity against load (post probe), for 10Mb/s Switch with UDP and TCP background traffic.

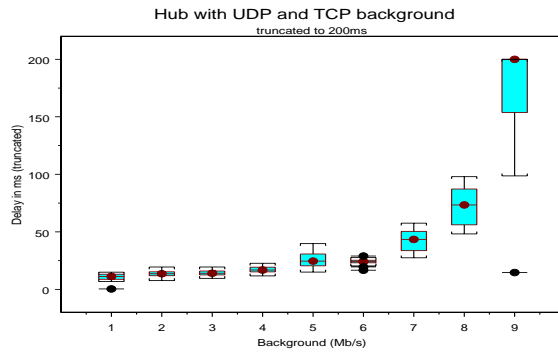


Fig. 8. Box plot of delay measurements for a hub with TCP and UDP background traffic.

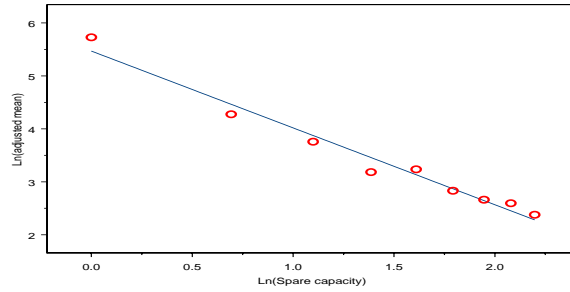


Fig. 9. Log-log plot of spare capacity against adjusted means, for a 10Mb/s hub with UDP and TCP background traffic.

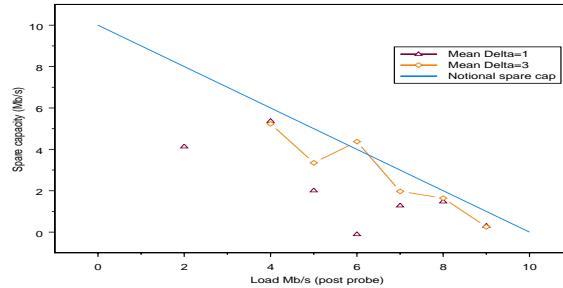


Fig. 10. Estimated (post probe) spare capacity against load (post probe), for a 10Mb/s Hub with UDP and TCP background traffic.

which our estimator relies. Some more work is needed to assess whether this is due to noisy observations, or whether this requires a refinement of our model.

6 Conclusion

We have considered an estimate of spare capacity in a network which makes use of RTT measurements. Heavy traffic analysis of stochastic queuing models, handling a mixture of reactive and unreactive traffic, suggests an inverse relationship between spare capacity and statistics of the RTTs. This allows us to construct estimates of the spare capacity based on ratios of such statistics. The analysis is applicable to both wide area and local area networks; indeed, we would expect the results to be more accurate for larger networks. In particular, the analysis of RTT correlations presented in Theorem 1 might be used to refine existing techniques for bottleneck capacity estimation such as pathchar.

We have given some experimental results for small networks, such as those found in a networked home, and where spare capacity estimation is an essential part of any admission control scheme. Preliminary results are encouraging, showing a good match between theory and practice, and enabling us to design simple and robust admission control schemes. More work is required to tackle the case of wireless networks. We are currently testing a prototype admission control service for multimedia applications,

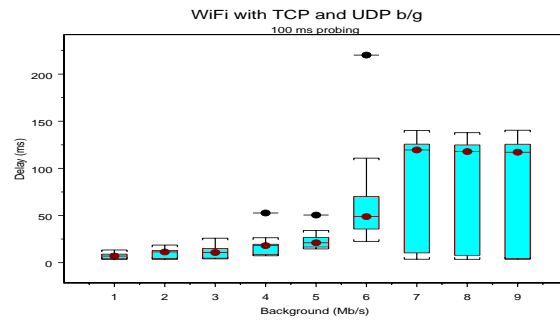


Fig. 11. Box plot of delay measurements for a wireless access point with TCP and UDP background traffic.

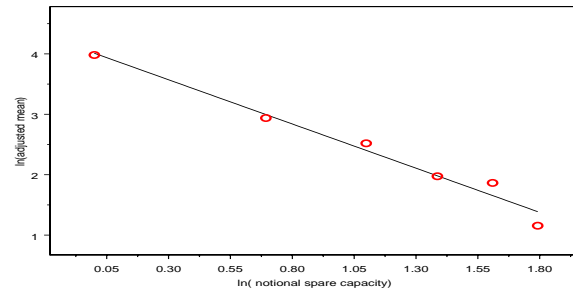


Fig. 12. Log-log plot of spare capacity against adjusted means for a wireless access point with UDP and TCP background traffic.

based on the results of this paper, which is able to make a decision whether to admit or reject a connection within the order of 100 ms.

References

1. S. Alouf, P. Nain and D. Towsley, Inferring Network Characteristics via Moment-Based Estimators, Proceedings of IEEE Infocom, 2001.
2. D. Bertsekas and R. Gallager, *Data Networks*. Prentice Hall 1992.
3. L. Breslau, E. Knightly and S. Shenker, Endpoint Admission Control: Architectural Issues and Performance, Proceedings of ACM Sigcomm, 2000.
4. R. Davidson and J.G. MacKinnon, *Estimation and Inference in Econometrics*, Oxford University Press, 1993.
5. M. Harrison, *Brownian Motion and Stochastic Flow Systems*, John Wiley and Sons, New York, 1985.
6. V. Jacobson, "pathchar". Available at <ftp://ftp.ee.lbl.gov/pathchar/>.
7. M. Jain and C. Dovrolis, End-to-End Available Bandwidth: Measurement, Dynamics and Relation with TCP Throughput, Proceedings of ACM Sigcomm, 2002.
8. H. Jiang and C. Dovrolis, Passive Estimation of TCP Round-Trip Times, Sigcomm Computer Communication Review, Vol. 32, No. 1, 2002.
9. F. Kelly, *Reversibility and Stochastic Networks*, Wiley, Chichester, 1979.
10. F. Kelly, P. Key and S. Zachary, Distributed Admission Control, IEEE Journal on Selected Areas in Communications, Vol. 18, No. 12, 2000.

11. S. Keshav, Packet-pair flow control. Available at <http://www.cs.cornell.edu/skeshav/doc/94/2-17.ps>, 1994.
12. K. Lai and M. Baker, Nettimer: a tool for measuring bottleneck link bandwidth. Available at <http://mosquitonet.stanford.edu/~laik/projects/nettimer>.
13. T. Liggett, *Interacting Particle Systems*, Springer, Berlin, 1985.
14. K. Matoba, S. Ata and M. Murata, Improving Accuracy of Bandwidth Estimation for Internet Links by Statistical Methods, IEICE Transactions on Communications, Vol. E00-B, No. 6 June 2001.
15. K. Park and W. Willinger (Eds), *Self-Similar Network Traffic and Performance Evaluation*, Wiley, 2000.
16. M. Reiman, Open queueing networks in heavy traffic, *Mathematics of Operations Research*, Vol. 9, No. 3, pp. 441-458, 1984.
17. V. Ribeiro, M. Coates, R. Riedi, S. Sarvotham, B. Hendricks and R. Baraniuk, Multifractal Cross-Traffic Estimation, In Proceedings ITC Specialists Seminar, Monterey 2000.

7 Appendix

7.1 Proof of Theorem 1

An alternative description of the process W is via its infinitesimal generator U and the corresponding domain, $D(U)$:

$$Uf(x) = \frac{\sigma^2}{2} f''(x) - df'(x),$$

and

$$D(U) = \{f \in C([0, b]) : f'' \in C([0, b]), f'(0) = f'(b) = 0\}$$

(see e.g. Liggett [13], examples 2.3 and 2.10, pages 13 and 16 respectively). Solving for the differential equation

$$Uf = -\lambda f, f \in D(U),$$

one finds that the eigenfunctions y_k and the corresponding eigenvalues $-\lambda_k$ of the generator on its domain are given by (6). We omit the details here.

This thus yields the expansion

$$\mathbf{E}[f(W_t)|W_0 = z] = \sum_{k \geq 0} a_k y_k(z) e^{-\lambda_k t}, \quad (22)$$

for some coefficients a_k depending on the function f . Also, considering the scalar product $\langle f, g \rangle = \int_0^b f(x)g(x)dx$ on $C([0, b])$, it is readily seen that the eigenfunctions y_k are such that $\langle \tilde{y}_k, \tilde{y}_l \rangle = 0$ when $k \neq l$, where $\tilde{y}_k(x) := e^{-dx/\sigma^2} y_k(x)$. Setting $t = 0$ in (22), the left-hand side is $f(z)$; equating the scalar product of each side with $y_k(z) e^{-2dz/\sigma^2}$ yields

$$a_k = \frac{\int_0^b f(z) y_k(z) e^{-2dz/\sigma^2} dz}{\int_0^b y_k^2(z) e^{-2dz/\sigma^2} dz}.$$

We finally obtain

$$\mathbf{E}[f(W_t)|W_0 = z] = \sum_{k \geq 0} \frac{1}{\beta_k} \left[\int_0^b y_k(x) f(x) e^{-2dx/\sigma^2} dx \right] y_k(z) e^{-\lambda_k t}, \quad (23)$$

where the coefficients β_k are defined in (8). From Equation (23) we can read off the conditional density of W_t given $W_0 = z$, and the results of Theorem 1 follow.

7.2 Proof of Proposition 1

The formula (11) for the stationary distribution π follows from the general results in [9], Chapter 3. The normalising constant $Z(a, b)$ reads

$$Z(a, b) = \sum_{l=0}^N \frac{b^l}{(N-l)!} \sum_{k \geq 0} \binom{k+l}{l} a^k,$$

and the expression (12) thus follows from the negative binomial formula, that is

$$\sum_{k \geq 0} \binom{k+l}{l} a^k = \left(\frac{1}{1-a} \right)^{l+1}.$$

Denote by D the queuing time and by M the number of packets queued at the bottleneck. It holds that

$$D = \frac{1}{C} \sum_{m=1}^M \sigma_m,$$

where the σ_m are the sizes of the enqueued packets, and are by assumption i.i.d., exponentially distributed with mean \bar{X} . For any fixed real number u , one thus has

$$\mathbf{E} \left(e^{iu(1-a)\bar{X}^{-1}CD} \right) = \mathbf{E} \left(\left(\mathbf{E} e^{iu(1-a)\bar{X}^{-1}\sigma_1} \right)^M \right) = \frac{Z(a/(1-iu(1-a)), b/(1-iu(1-a)))}{Z(a, b)}.$$

Using the formula (12) for Z in the previous expression, one readily obtains that

$$\lim_{a \rightarrow 1} \mathbf{E} \left(e^{iu(1-a)\bar{X}^{-1}CD} \right) = \left(\frac{1}{1-iu} \right)^{N+1},$$

which is the characteristic function of the $\Gamma(N+2, 1)$ distribution, thus establishing the proposition.

7.3 Proof of Proposition 2

Introduce the notations

$$\varepsilon_0 = \hat{\tau} - \tau, \quad \varepsilon = \frac{1}{k} \sum_{n=1}^k Z_n - \mathbf{E}(Z), \quad \varepsilon' = \frac{1}{k} \sum_{n=1}^k Z'_n - \mathbf{E}(Z).$$

The identity $\bar{\tau} = \tau + \mathbf{E}(Z)/(C - m) + \varepsilon/(C - m)$ and its analogue for $\bar{\tau}'$, used in conjunction with (16), can be used to establish that, to the first order in ε and ε' , the relative error e verifies

$$e = \frac{C - m'}{\delta m \mathbf{E}(Z)} (\delta m \varepsilon_0 - \varepsilon' + \varepsilon). \quad (24)$$

By definition, ε_0 reads $Z_{(1)}/(C - m)$, where $Z_{(1)}$ is the minimum of the Z_n . We now evaluate the asymptotic distribution of the minimal value $Z_{(1)}$. We need only consider the case where the second parameter B in the Gamma distribution of the Z_n is 1, as this disappears due to the division by $\mathbf{E}(Z)$ in the expression of e . The density of the Z_n is then given by $f(z) = z^N e^{-z}/\Gamma(N + 1)$. It holds that

$$\mathbf{P}(Z_{(1)} > t) = (\mathbf{P}(Z_1 > t))^k.$$

Solving for $\mathbf{P}(Z_1 > t_k) = 1 - u/k$, for some fixed positive u , one finds

$$t_k \sim \left(\Gamma(N + 2) \frac{u}{k} \right)^{1/(N+1)}.$$

One thus gets that

$$\mathbf{P} \left(Z_{(1)} > \left(\frac{u \Gamma(N + 2)}{k} \right)^{1/(N+1)} \right) \sim e^{-u}.$$

This yields that $k^{1/(N+1)} Z_{(1)}$ follows asymptotically a Weibull distribution specified by the complementary distribution function $\exp(-t^{N+1}/\Gamma(N + 2))$. When N is larger than 1, ε_0 will then constitute the principal component in (24), as the other error terms are of order $1/\sqrt{k}$ by the central limit theorem. Thus, for $N > 1$, we have as $k \rightarrow \infty$ the weak convergence

$$k^{1/(N+1)} e \xrightarrow{W} \frac{C - m'}{C - m} X,$$

where X denotes a random variable with the aforementioned Weibull distribution. This establishes the first part of the proposition.

In the case where $N < 1$, the main contributions in (24) are from ε and ε' . By the central limit theorem we then obtain, as $k \rightarrow \infty$, that

$$\sqrt{k} e \xrightarrow{W} \frac{C - m'}{\delta m} 2(N + 1) N(0, 1).$$