A Nonstationary Traffic Train Model for Fine Scale Inference from Coarse Scale Counts

Chuanhai Liu Scott Vander Wiel and Jiahai Yang

Chuanhai Liu and Scott Vander Wiel are with Bell Laboratories, Lucent Technologies. E-mail: {liu, scottyv}@research.bell-labs.com.

Jiahai Yang is with Tsinghua University and CERNET. E-mail: yang@cernet.edu.cn.

Abstract

The self-similarity of network traffic has been convincingly established based on detailed packet traces. This fundamental result promises the possibility of solving on-line and off-line traffic engineering problems using easily-collectible coarse time-scale data, such as SNMP measurements. This paper proposes a statistical model that supports predicting fine time-scale behavior of network traffic from coarse time-scale aggregate measurements. The model generalizes the commonly used fractional Gaussian noise process in two important ways: (1) it accommodates the recurring daily load patterns commonly observed on backbone links; and (2) features of long range dependence and self-similarity are modeled only at fine time scales and are progressively damped as the time period increases. Using the data we collected on the Chinese Education and Research Network, we demonstrate that the proposed model fits five-minute data and generates ten-second aggregates that are similar to actual ten-second data.

Index Terms

Daily pattern, Fractional Gaussian Noise, Poisson process.

I. INTRODUCTION

Network traffic engineering problems are often addressed with readily available aggregate statistics such as packet counts in fixed intervals. For example, engineers at Global Crossing have reported using aggregate traffic statistics on label switched paths (LSPs) to resize and reroute the LSPs in both core and regional networks [15]. This approach is practical and has proven effective, even if it may not be as refined as a method that takes into account detailed traffic descriptors or measurements of packet-level timing and queuing behavior. Routine traffic monitoring, wavelength assignments, route optimization, admission control policies for new data services and peering policies are also amenable to practical solutions based on aggregate traffic counts. Elegant optimization techniques have been proposed for many of these engineering problems but they often require detailed traffic descriptors, such as effective bandwidths of traffic on individual LSPs or between source-destination pairs. Typically these specifications are not precisely known, they vary by time of day, and they change as Internet Service Providers (ISPs) continue to add customers, provision additional services, and upgrade network elements.

The simplest and most common type of traffic measurements are packet counts and byte counts taken at, say, five minute intervals. Five-minute packet counts are easy to collect and store for a large number of network elements over long periods of time, but the data are not as

rich as traces of time-stamped packet headers that identify the sizes and spacings of individual packets as they traverse the network. Such packet traces, however, are not typically available with the comprehensive coverage over both time and topology that is needed to address many ongoing network-wide traffic engineering problems. Even if packet traces are available, most or all of the relevant information for many engineering problems might be obtained from coarse time-scale measurements. This paper explores ways to extract information relevant to fine time-scale dynamics using aggregate statistics polled regularly from router and switch interfaces or from LSPs in Multiprotocol Label Switching (MPLS) networks.

The vast literature on characteristics of network traffic based on detailed packet trace data provides strong evidence that different types of traffic on high-speed data networks are self-similar, so the Hurst parameter [11] which describes long range dependence, remains the same for different levels of aggregation. These include studies of Local Area Network traces [9], Wide Area Network traffic [12], and variable bit rate traffic [1]. Crovella and Bestavros [2] provided an example in which the estimated Hurst parameter remains constant as the aggregation time scale increases from 1 second to 3 minutes. The self-similarity of network traffic at time scales up to several minutes suggests that coarse scale five-minute measurements can be used for traffic engineering that depends on finer time-scale characteristics. But this is challenging because the previous self-similar models are stationary so they do not allow daily and hourly patterns in traffic load [13]. On the other hand, sequences of five-minute counts long enough to usefully estimate self-similar properties are typically too long to regard as stationary: 24 five-minute data values cover a period of two hours and the average load on a link can easily double in that time.

This paper provides a new *Nonstationary Traffic Train (NTT)* model that supports predicting fine time-scale behavior of network traffic from coarse time-scale aggregate measurements. The NTT model generalizes the commonly used fractional Gaussian noise (fGn) [9], [14], [2] process in two important ways. (1) The model accommodates the recurring daily load patterns commonly observed on backbone links and can allows for day of week effects or trends in the traffic rate over longer periods of time. (2) Long range dependence and self-similarity are modeled only at fine time scales and are progressively damped as the time scale increases. The result is a traffic model that is more realistic for describing patterns observed in five-minute aggregate data and is consistent with the fGn model at fine time scales.

Duffield, Massey and Whitt [3] considered the properties of a more general *nonstationary offered-load model* in which a non-homogeneous Poisson process describes customer arrivals and the time-varying bandwidth requirement of each customer follows a general random process. The NTT model is a Gaussian approximation to the special case where traffic train durations are exponentially distributed and within-train packet-counts are fGn processes. We provide detailed formulas for the NTT covariance structure. More important, however, is our empirical demonstration that the NTT stochastic process model provides sound inference on network traffic properties at fine time scales using only coarse scale traffic measurements. Inference at fine scale is made possible by the self-similarity property of the within-train fGn processes. Using seven days of five minute packet count data from the China Education and Research Network, the fitted NTT model is used to estimate upper quantiles of 10 second counts. The estimates track empirical quantiles with good accuracy.

The remainder of the paper is arranged as follows. Section II describes data that we analyze. Section III derives our non-stationary generalization of the fGn model as the superposition of a large number of packet trains that arrive and depart the network in a non-stationary fashion. Section IV describes estimation of model parameters along with a bootstrap method for assessing uncertainty. Section V shows fits of the model to packet count data. Section VI concludes with a few remarks.

II. THE DATA

The China Education and Research Network (CERNET) is the nationwide education and research computer data network in China funded by the Chinese Ministry of Education and operated by Tsinghua and other leading universities. The network serves more than 750 education and research institutions and approximately 4 million users throughout the country. CERNET is organized hierarchically with a nation-wide backbone connecting ten regional networks that in turn serve lower-level provincial and campus networks.

We have collected packet and byte counts at 10-second intervals using SNMP polling on all interfaces of two CERNET routers and two switches running in the CERNET backbone at 100 Mbps. Ten-second polling does not burden the devices or the network but would present storage issues if continued permanently on a large number of devices. Five-minute polling intervals are common for network management systems and typically allow several months of traffic data to

be maintained online. To mimic coarser polling, we aggregate the 10-second data to intervals 300-seconds (five-minutes.)

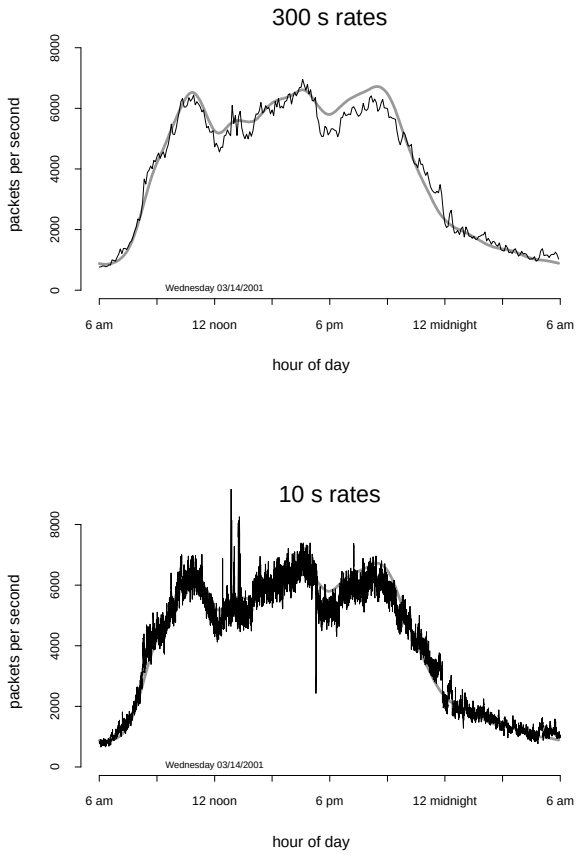

Fig. 1. A sample of packet count data at two levels of aggregation: 300-second (upper plot) and 10-second (lower plot.) The packet rates clearly follow a daily pattern but also wander significantly from it. The gray curves in the background are smoothed daily averages across 12 consecutive weekdays of data.

Packet count data for a single representative day in one direction on a 100 Mbps link are plotted in Figure 1. Approximately 80% of the packets are TCP. The lower plot shows the 10 second packet counts (expressed in packets per second) while the upper plot shows the corresponding 300-second aggregate rates. The plots for all seven days (not shown) demonstrate several characteristics of the packet count timeseries. There is a consistent daily pattern in which the traffic level begins to rise steeply around 7 a.m. and peaks at 10:30 a.m. Additional peaks are seen at 3 p.m. and 9 p.m. After 9 p.m. traffic levels taper into the early morning hours. The late-night decline is more gradual than the 7 a.m. rise, however. These data also tend to wander to one

side of the mean level for periods of time. This is characteristic of the long-range dependence observed in many kinds of network traffic data. Another obvious feature of the packet counts is that the 10-second rates are much more variable than the 300-second aggregates. This scaling of variability with the time resolution is an important feature that the NTT model needs to capture—but it should do so while only using information in the 300-second traces.

The 300-second data is easy to collect within CERNET or any other data network. Many traffic engineering problems, however, need to account for variations and burstiness of traffic over much finer timer scales. Variations at 10 to 100 milliseconds are relevant for many problems but in most cases data at these scales are not widely available. In the next section a model is developed for coarse-scale aggregate packet counts that can be used to infer characteristics of counts at finer time scales even if burstiness is not as evident in the coarse scale counts. We use it in Section V to predict characteristics of the 10-second data using only the 300-second series.

III. The Model

It has been well established that network traffic, such as packet count sequences over short periods of time, can be modeled as a fractional Gaussian noise (fGn) process [9], [14], [2], which is necessarily stationary. To understand fine time-scale behavior of network traffic based on coarse time-scale SNMP measurements, we extend the fGn model to the non-stationary traffic train (NTT) model.

The fGn model for network traffic has been motivated using a *byte-train* formulation with ON/OFF periods following long-tailed distributions [14]. This is depicted in Figure 2 showing a large number of active byte trains each with active ON periods followed by inactive OFF periods. The ON and OFF periods within a train are independent and the trains are independent and have identical statistical properties. The curve at the bottom of the figure is the aggregate of the trains. Its value across each sampling interval is equal to the total ON time of all trains during that interval. As the number of trains increases, the fine time-scale aggregate process approaches fractional Gaussian noise. In this byte-train formulation, long-range dependence of the aggregate comes from the long-tailed ON and OFF distributions. More detailed mathematical descriptions are given in Sections III (A) and (B).

Figure 2 shows a nonstationary byte-train formulation that underlies the NTT model. Instead of a fixed number of byte trains across all time, trains now arrive and depart the system at

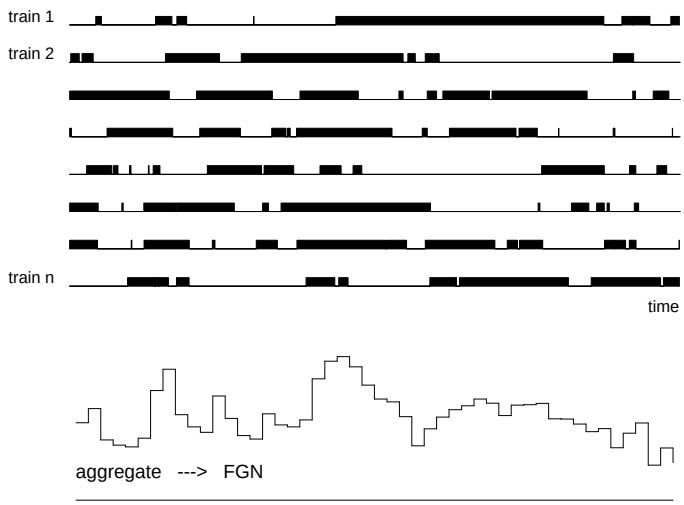

Fig. 2. Superposition of byte trains with ON and OFF periods. Aggregation is over trains and across regularly spaced time intervals. The aggregate process tends toward fractional Gaussian noise which is self-similar and inherits long-range dependence from the individual trains.

random times. Train arrival times form a non-stationary Poisson process and their *lifetimes* are exponentially distributed with a *large* mean. The figure depicts a situation where the arrival rate of new trains suddenly increases at the halfway point after which the number of active trains tends to increase to a new level. With a large number of trains the aggregate process shown at the bottom of the figure represents our new NTT model. At small time scales it inherits properties of self-similarity and long-range dependence from the individual ON/OFF processes but the NTT model is nonstationary if the train arrival rate varies through time. Also correlations at larger time-scales are damped in the NTT model because the trains do not continue forever.

To simplify the derivation of the NTT model, we can also think of an fGn sequence as a superposition of a large number of independently and identically distributed fGn processes, each of which represents an active *traffic train*. Although there is a distinction between byte trains and traffic trains, that is, byte-trains are 0-1 processes whereas traffic trains are fGn processes, in either case the aggregate traffic follows the conventional fGn model so the distinction is not important when modeling aggregates. Using fGn traffic trains simplifies the theoretical NTT model formulation: and NTT model is the superposition of a large of number of active fGn traffic train segments, where train arrivals form a non-homogeneous Poisson process and train durations are exponential.

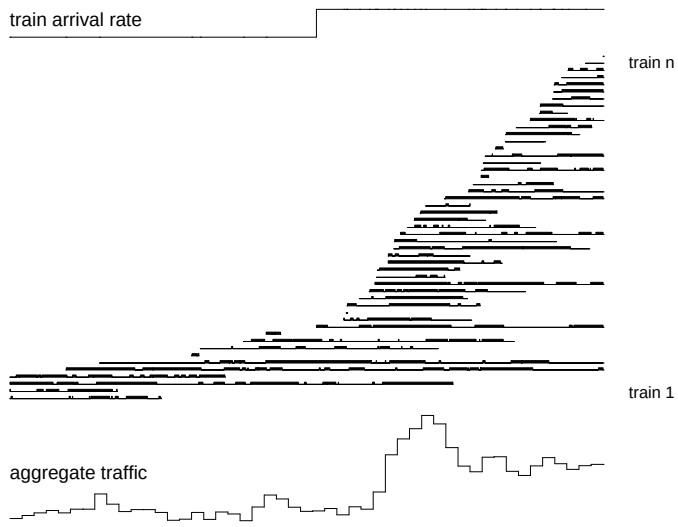

Fig. 3. Superposition of byte trains with random arrival and departure times. The aggregate process tend toward the NTT model. Non-stationarity comes from changes in the train arrival rate over time. Self-similarity and correlation properties at fine time-scales are the same as for fractional Gaussian noise but these are damped at larger time-scales.

A. Network traffic self-similarity and fractional Brownian motion

Network traffic counts aggregated over short time scales have been found to be self-similar [9], [14], [2] over short time periods up to several minutes. The increments of a continuous time process $\{X(t) : -\infty < t < \infty\}$ are said to be self-similar with parameter H $(H \geq 0)$ if for any $s > 0$ and any t_0 and τ

$$\{X(t_0 + \tau) - X(t_0)\} \stackrel{d}{=} \{s^{-H} X(t_0 + s\tau) - X(t_0)\},$$

where the notation $\{X(t)\} \stackrel{d}{=} \{Y(t)\}$ means that the two processes $\{X(t)\}$ and $\{Y(t)\}$ have the same finite joint distributions [11]. The self-similarity parameter H is also called Hurst parameter.

Fractional Brownian motion (fBm) is the unique self-similar Gaussian process so it makes a natural choice for a model of cumulated traffic. Increments of an fBm process are known as fractional Gaussian noise (fGn) which is the model used for network traffic counts. More specifically, the fGn model for network counts with time scale S at discrete times $t = 0, \pm 1, \pm 2, ...$ is written as

$$x_{St} = \alpha S + S^H \sigma G_H(t), \tag{1}$$

where x_{St} denotes the aggregated traffic over the time interval $(St, S(t+1))$, and $G_H(t)$ is a

stationary Gaussian process with mean zero and autocovariance function

$$\begin{aligned} & c(s, t = s + d; H, \sigma^2) \equiv \mathrm{E}\left\{G_H(s)G_H(t)\right\} \\ = \; & \frac{1}{2}\left[|d+1|^{2H} + |d-1|^{2H} - 2|d|^{2H}\right]. \end{aligned} \tag{2}$$

B. A nonstationary traffic train (NTT) model

To develop the NTT model, think of x_{St} as counts from a large number, $m(t) = m$, of independently and identically distributed *traffic trains*, each of which is itself a fGn process with the common parameters α, σ^2, and H. So,

$$x_{St}|\{m_t = m\} = \alpha m S + S^H \sum_{i=1}^{m} \sigma G_H^{(i)}(t). \tag{3}$$

Now α can be interpreted as the traffic rate per train and per unit time.

Next, consider a latent Poisson process n_t with intensity $\lambda(t)$ $(\lambda(t) \geq 0)$ for the cumulative count of traffic train arrivals in the time interval $(0, t)$. Let train *lifetimes* or *durations* be independent exponentials with mean μ and let $A(t)$ denote the set of active trains at time t. Then $m(t) = |A(t)| = n_t -$ number of train deaths in$(0, t)$. Proposition 1 describes the distribution of the $m(t)$.

Proposition 1: Suppose $m(0) \sim$ Poisson $\{\Lambda(0)\}$ with given initial mean $\Lambda(0)$, then

$$m(t) \sim \text{Poisson}\left\{\Lambda(t)\right\}$$

and autocovariance function

$$\text{cov}\left\{m(s), m(t)\right\} = \Lambda(\min\{s,t\})\exp\left\{-\frac{|t-s|}{\mu}\right\}, \tag{4}$$

where $\mathrm{E}\left\{m(t)\right\} = \Lambda(t)$ has the following recursion relationship

$$\begin{aligned} \Lambda(t+\Delta t) &= \Lambda(t)\exp\left\{-\frac{\Delta t}{\mu}\right\} \\ &+ \exp\left\{-\frac{\Delta t}{\mu}\right\}\int_0^{\Delta t} \lambda(t+u)\exp\left\{-\frac{u}{\mu}\right\} du \end{aligned} \tag{5}$$

for any $\Delta t > 0$, or equivalently,

$$\frac{d\Lambda(t)}{dt} + \frac{\Lambda(t)}{\mu} = \lambda(t). \tag{6}$$

■

A simple proof is provided in Appendix (A), although the results are not new. More general results in the context of $M_t/G/\infty/$ queues are given in [5]. (See, for example, Theorem 2 and Corollary 4 of [5] and included references.)

Since $\lambda(t) \geq 0$, Equation (6) imposes the following constraint on μ

$$\frac{\Lambda(t)}{\mu} \leq -\frac{d\Lambda(t)}{dt}, \tag{7}$$

in addition to the positivity constraint $\mu > 0$.

For small or *infinitesimal* time scale S we can write (3) as

$$x_{tS} = \alpha S m(tS) + S^H \sum_{i \in A(tS)} \sigma G_H^{(i)}(t). \tag{8}$$

Note that

$$\mathrm{E}\left\{x_{tS}\right\} = \alpha S \Lambda(tS)$$

and

$$\mathrm{var}\left\{x_{tS}\right\} = \left[\alpha^2 S^2 + S^{2H}\sigma^2\right] \Lambda(tS).$$

Since $\mathrm{E}\left\{m(tS)\right\} = \mathrm{var}\left\{m(tS)\right\} = \Lambda(tS)$ and $\frac{m(tS)}{\Lambda(tS)}$ tends to one in probability as $\Lambda(tS) \to \infty$, then for any t

$$\begin{aligned}
&\lim_{\Lambda(tS)\to\infty} \frac{x_{tS} - \mathrm{E}\left\{x_{tS}\right\}}{\sqrt{\alpha\Lambda(tS)}} \\
&= \alpha^{1/2} S \lim_{\Lambda(tS)\to\infty} \frac{m(tS) - \Lambda(tS)}{\sqrt{\Lambda(tS)}} \\
&\quad + S^H \alpha^{-1/2} \sigma \lim_{\Lambda(tS)\to\infty} \frac{\sum_{i\in A(tS)} G_H^{(i)}(t)}{\sqrt{\Lambda(tS)}} \\
&\stackrel{d}{=} \alpha^{1/2} S z(t) + S^H \alpha^{-1/2} \sigma y(t)
\end{aligned}$$

where $z(t)$ and $y(t)$ are independent standard normal deviates. Making use of (4) leads to

$$\mathrm{cov}\left\{z(s), z(t)\right\} \approx \frac{\Lambda(\min\{s,t\}S)}{\sqrt{\Lambda(sS)\Lambda(tS)}} \exp\left\{-\frac{|t-s|S}{\mu}\right\}.$$

Following the same arguments used to prove Proposition 1, we can show that

$$\begin{aligned}
&\mathrm{cov}\left\{y(s), y(t)\right\} \\
&\approx \frac{\Lambda(\min\{s,t\}S)}{\sqrt{\Lambda(sS)\Lambda(tS)}} \exp\left\{-\frac{|t-s|S}{\mu}\right\} c(s,t;H,\sigma_y^2).
\end{aligned}$$

This leads to the following traffic train model for *infinitesimal* time scale S.

$$\begin{aligned} x_{tS} \quad \stackrel{d}{=} \quad & \alpha S\Lambda(tS) + \alpha S\Lambda^{1/2}(tS)z(tS) \\ & +\sigma_y S^H \Lambda^{1/2}(tS)y(tS) \\ & +\sigma_e S^{1/2}\Lambda^{1/2}(tS)e(tS), \end{aligned} \tag{9}$$

where $\{z(t)\}$, $\{y(t)\}$, and $\{e(t)\}$ are independent, mean zero Gaussian processes with autocovariance functions

$$\text{cov}\left\{z(s), z(t)\right\} = \frac{\Lambda(\min\{s,t\}S)}{\sqrt{\Lambda(sS)\Lambda(tS)}}\exp\left\{-\frac{|t-s|S}{\mu}\right\},$$

$$\text{cov}\left\{y(s), y(t)\right\} = \text{cov}\left\{z(s), z(t)\right\} c(s,t;H),$$

and

$$\text{cov}\left\{e(s), e(t)\right\} = \begin{cases} 1 & \text{for } s = t; \\ 0 & \text{otherwise.} \end{cases}$$

The process $\{e(tS)\}$ is white noise and is included to accommodate possible extra variability, such as measurements noise.

Next we derive the NTT model for a finite time scale T as the aggregated traffic counts when infinitesimal counts follow (9). Let $X(tT)$ denote the cumulative traffic on the link in the interval $(0, tT)$. Then $x(tT) \equiv X(tT+T) - X(tT)$ and the distribution of the increments $x(tT)$ can be obtained by representing them as a sum of increasingly smaller increments and then using (9); that is, $x(tT)$ is

$$\lim_{N\to\infty} \sum_{j=0}^{N-1} \left[X(tT + (j+1)T/N) - X(tT + jT/N)\right],$$

where for large N, the summands approach the infinitesimal process (9). Let $S = T/N$ and $u_j = tT + jS$. Finding the NTT model representation at the time scale T requires determining the following limiting processes:

$$\theta(tT) \equiv \alpha \lim_{N\to\infty} \sum_{j=0}^{N-1} \Lambda(u_j)S,$$

$$z(tT) \equiv \alpha \lim_{N\to\infty} \sum_{j=0}^{N-1} \Lambda^{1/2}(u_j)z(u_j)S,$$

$$y(tT) \equiv \sigma_y \lim_{N\to\infty} \sum_{j=0}^{N-1} \Lambda^{1/2}(u_j) y(u_j) S^H,$$

and

$$e(tT) \equiv \sigma_e \lim_{N\to\infty} \sum_{j=0}^{N-1} \Lambda^{1/2}(u_j) e(u_j) S^{1/2}.$$

Although we have been mathematically somewhat informal, the limiting processes are independent Gaussian with covariances as described in the following formal definition of the NTT model at finite time scale T.

Definition 1—**Nonstationary Traffic Train Model:**
Aggregate traffic counts $x(tT)$ at time-scale T follow an NTT model if they can be represented as

$$x(tT) = \theta(tT) + z(tT) + y(tT) + e(tT), \tag{10}$$

where $\theta(tT)$ is a deterministic mean curve

$$\theta(tT) = \alpha \int_0^T \Lambda(tT+u)du \tag{11}$$

where $\Lambda(\cdot)$ satisfies (6) for some nonnegative intensity function $\lambda(\cdot)$, and where $\{z(tT)\}$, $\{y(tT)\}$, and $\{e(tT)\}$ are independent zero-mean Gaussian processes. Let $p = \exp\{-\mu^{-1}\}$. Variations from traffic train arrivals and departures are represented by $z(tT)$ with autocovariance

$$C_T^{(z)}(t,t) = 2\alpha^2 \int_0^T \int_0^u \Lambda(Tt+z) p^{u-z} dz du$$

and

$$C_T^{(z)}(s,t) = \alpha^2 \int_0^T \int_0^T \Lambda(Ts+u) p^{T(t-s)+z-u} dz du$$

for $t > s$. Within-train traffic variations are represented by the process $y(tT)$ with autocovariance

$$\begin{aligned} C_T^{(y)}(t,t) &= 2H(2H-1)\sigma_y^2 \int_0^T \int_0^u \Lambda(Tt+z) p^{u-z} \\ &\quad (u-z)^{2H-2} dz du \end{aligned}$$

and

$$\begin{aligned} C_T^{(y)}(s,t) &= H(2H-1)\sigma_y^2 p^{T(t-s)} \int_0^T \int_0^T \Lambda(Ts+u) \\ &\quad p^{z-u}(T(t-s)+z-u)^{2H-2} dz du, \end{aligned}$$

for $t > s$. Additional independent noise is represented by the process $e(tT)$ with autocovariance simply

$$C_T^{(e)}(s,t) = \begin{cases} \sigma_e^2 \int_0^T \Lambda(Tt+u)du, & \text{if } s = t; \\ 0, & \text{otherwise.} \end{cases}$$

If $H = 1$, then $\{y(tT)\} \stackrel{d}{=} \{z(tT)\}$ and if $H = 1/2$ then $\{y(tT)\} \stackrel{d}{=} \{e(tT)\}$.

∎

IV. Estimation

In accordance with the hierarchical formulation of the NTT model in terms of a train arrival process and a within-train packet processes, we use a two-step procedure for fitting the NTT model to data. First, we estimate the daily pattern $\Lambda(\cdot)$ with a smooth curve. Second, we find the maximum likelihood estimate of the parameters $\alpha, \sigma_y^2, \sigma_e^2, \mu, H$ given $\Lambda(\cdot) = \hat{\Lambda}(\cdot)$. Third, we describe a bootstrap method [4] for assessing the uncertainty in the estimated parameters.

A. Estimation

Here we estimate $\alpha\Lambda_t$ from T-second aggregate data collected across D days having similar daily traffic pattern. We use $T = 5 \times 60 = 300$ seconds. Let $\bar{x}_{Tt}$ be the average across D days of the observed number of packets during the daily time interval $(Tt, T(t+1))$ for $t = 0, ..., n = 24 \times 60 \times 60/T - 1$. Then

$$\mathrm{E}\left\{\bar{x}_{Tt}\right\} = \alpha \int_0^T \Lambda_{Tt+u} du.$$

$\Lambda(\cdot)$ is expected to vary smoothly over time and, therefore, it is estimated by smoothing or filtering the $\bar{x}_{Tt}$ values to smooth them further. The smoothing uses cyclic time indices so that the end of the day is adjacent to its beginning. We use a filter that corresponds to fitting a local least squares quadratic curve to the data within a three hour window centered at Tt with *tricube* weights; that is, $\bar{x}_{(Tt-k)}$ (for $|k| < B \equiv 3 * 60 * 60/(2T)$) has weight $(1 - |k/B|^3)^3$ in the local regression.

The smoothing filter produces estimates of $\alpha\Lambda_t$ where each t represents a multiple of five-minutes. These discrete points are then joined using quadratic interpolation (with continuous first derivatives) to obtain an estimated daily pattern, $\hat{\alpha}\hat{\Lambda}(Tt)$ as a curve in continuous time.

Given the daily pattern $\Lambda(\cdot)$, the model for the packet counts is simply a variance-component normal model. We apply Quasi-Newton method to find the maximum likelihood estimates of $\alpha, \sigma_y^2, \sigma_e^2, \mu,$ and H, where μ is constrained by (7).

The likelihood function has a simple expression using the vector and matrix notation. Let $\boldsymbol{x}_i$ denote the vector of aggregate traffic measurements x_{Tt} in day i and $\boldsymbol{\theta}_i$ the corresponding vector of means $\theta(Tt)$. We model the $\boldsymbol{x}_i$'s as independent with distribution determined by model (10):

$$\boldsymbol{x}_i - \boldsymbol{\theta}_i \overset{iid}{\sim} \mathrm{N}\left(0, \alpha R_z + \sigma_y^2 R_y + \sigma_e^2 R_e\right) \tag{12}$$

where $R_z = \alpha^{-2} C_T^{(z)}$, $R_y = \sigma_y^{-2} C_T^{(y)}$ and $R_e = \sigma_e^{-2} C_T^{(e)}$. The log-likelihood is

$$\begin{aligned} \mathcal{L}(\alpha, \sigma_y^2, \sigma_e^2, \mu, H) &= -\frac{nm}{2}\ln(2\pi) - \frac{n}{2}\ln\det\Sigma \\ &\quad -\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\theta}_i)'\Sigma^{-1}(\boldsymbol{x}_i - \boldsymbol{\theta}_i) \end{aligned}$$

where $\Sigma = \Sigma(\alpha, \sigma_y^2, \sigma_e^2, p, H) = \alpha R_z(\mu) + \sigma_y^2 R_y(\mu, H) + \sigma_e^2 R_e$. Dependence on time scale T and on $\Lambda(Tt)$ has been suppressed from the notation. Numerical maximization can be performed using derivatives of $\mathcal{L}$ with respect to the parameters.

Appendix (B) gives the needed formulas for calculating the covariance matrices. The first-order derivatives of the likelihood function, not presented here, require evaluation of incomplete gamma functions.

B. Assessment of the uncertainty in parameter estimates

To assess the uncertainty in the estimated parameters, we apply a variant of the bootstrap method [4], which is a simple and widely used resampling technique in statistics for assessing the uncertainty about parameter estimates. Denote by D the number of days. Our bootstrap repeats the following two steps a large number, M, of times.

1) Obtain a bootstrap sample of size D days with replacement and then apply the smoothing procedure of Section IV.A to the bootstrap sample to obtain the estimate of the daily pattern $\Lambda(t)$.
2) Obtain another bootstrap sample of size D days with replacement and then find the maximum likelihood estimate of $(\alpha, \sigma_y^2, \sigma_e^2, \mu, H)$ based on the bootstrap sample and the estimate of $\Lambda(t)$ obtained in Step 1.

V. THE RESULTS

This section illustrates results of fitting the NTT model to CERNET packet count data over at various levels of aggregation.

A. Model estimates and uncertainty

Figure 4 displays seven weekdays of 300-second aggregate counts along with a smooth thick curve that represents the estimate $\hat{\alpha}\hat{\Lambda}(\cdot)$, expressed in packets per second. The smooth curve is obtained by applying a local quadratic filter to the data as described in Section IV. The three peaks in mean rate at 11 a.m., 4 p.m. and 8:30 p.m. are approximately eight times as large as the minimum rate at 6 a.m. At 8 a.m. the load is increasing at a rate of over 90% per hour. Clearly, this strong daily pattern must be accounted for when modeling the packet count series. Even ten-minute segments of fine time-scale data can cover more than a 15% change in the mean packet rate. Ignoring this trend would alter any assessment of stochastic dependence in the data.

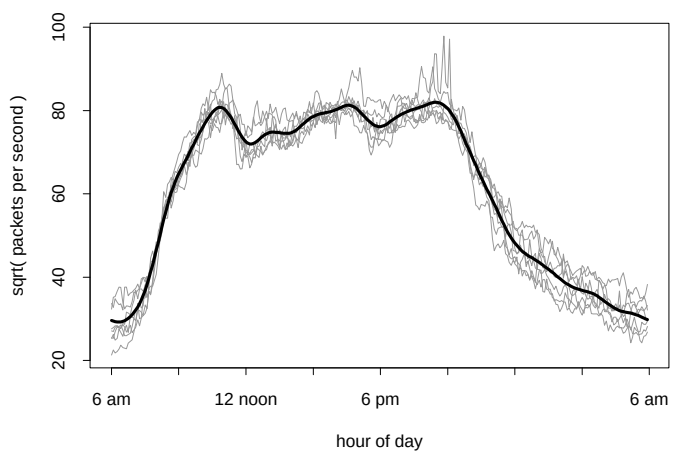

Fig. 4. Estimate of $\alpha\Lambda(t)$ (thick curve) along with seven days of 300-second counts scaled as the square root of packets per second. The daily pattern is strong; for example, there is an eight-fold increase in load between 6 a.m. and 11 a.m.

Given the estimated mean rate, $\hat{\Lambda}(\cdot)$, maximum likelihood estimates and bootstrap 90% confidence limits of the remaining parameters are as follows:

parameter	α	σ_y	μ (hr.)	H
MLE	0.0067	8.6	2.3	0.952
lower conf. limit	0.0051	8.0	1.7	0.916
upper conf. limit	3.4	9.3	2.4	0.972

Both the maximum likelihood estimate (MLE) and *all* bootstrap estimates for σ_e were below 0.001 indicating that the additional noise term $e(t)$ is not necessary. (500 bootstrap replications were used.) The point of steepest relative decrease in $\hat{\Lambda}(t)$ occurs about 5:30 p.m. and determines an upper bound on the MLE of μ through (7). It happens that $\hat{\mu}$ is equal to its upper bound for this data. Nevertheless, the upper confidence limit exceeds $\hat{\mu}$ because each on replication of bootstrap method $\Lambda(t)$ is re-estimated from a different sample of data so the upper bound on μ varies across bootstrap samples. Thus, the bootstrap recognizes uncertainty in μ by properly accounting for uncertainty in $\Lambda(\cdot)$. The small estimate of α implies that a large number of latent packet trains (eg. mean rate/α) are modeled as contributing to the total flow. Since $z(\cdot)$ is proportional to α, relatively little variation in the aggregate counts is modeled as arising from the train arrival and departure process.

B. Inference on fine-scale characteristics

Figures 5 and 6 compare actual packet count series to simulated series at aggregation levels of 300 seconds (Figure 5) and 10 seconds (Figure 6). These displays help assess how well the nonstationary traffic model describes typical fluctuations at different levels of aggregation. In Figure 5 the upper block of panels plots 300-second packet counts for six days of the CERNET data. The lower block of panels plots eight days of simulated data. Each simulated day comes from a random bootstrap fit so as to incorporate model uncertainty into the simulations. The smooth gray curve in each panel is the estimated daily pattern. The simulated counts have the same broad features as the actual counts: a prominent daily pattern, wandering departures from the mean, and variability that increases with level. There are, perhaps, somewhat more abrupt features in some of the actual series than are seen in the simulated ones, but the model is describing 300-second packet counts that, to a large extent, behave like the observed ones.

Figure 6 is laid out in the same manner as Figure 5 but displays actual and simulated 10-second packet counts. Once again the 10-second counts are simulated from bootstrap fits of

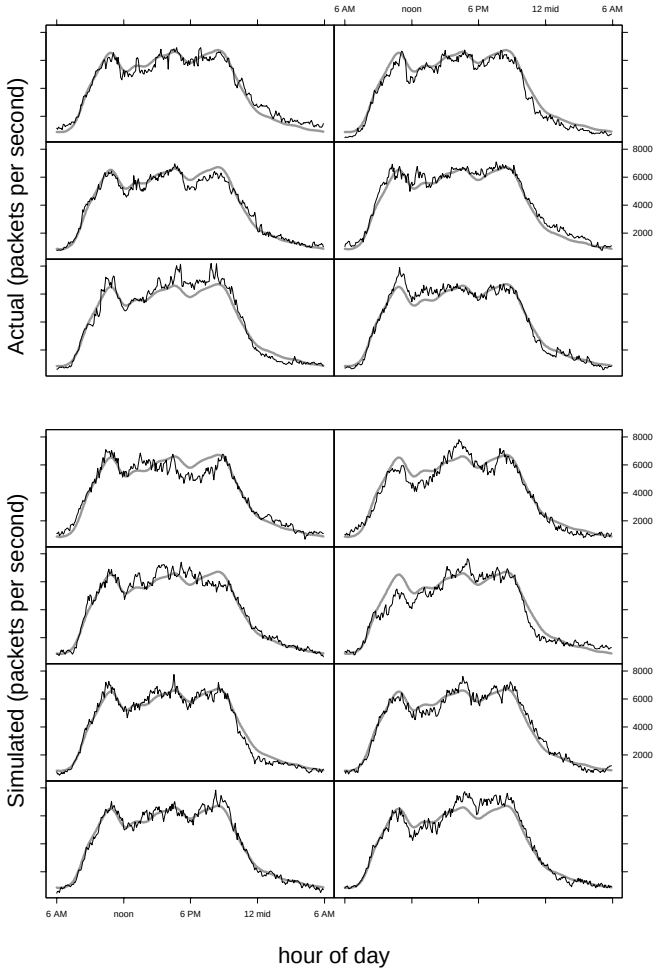

Fig. 5. Actual and simulated 300-second packet counts. Six days of actual counts are shown in the top block. Eight days of simulated counts are shown in the bottom block. The simulations are based on random bootstrap fits to 300-second data. The NTT model clearly captures the daily pattern and various excursions from it.

the 300-second data—so the figure demonstrates that a nonstationary model fit to data at coarse time-scale (300 seconds) can predict features of data on a much finer time-scale (10 seconds). In fact, the 10 second data has the same general trend and roughly the same variance and tendency to wander as the actual 10 second data. These features are well-modeled even though no fine-scale counts were involved in the fitting process.

Figure 7 gives another demonstration of the NTT model's ability to extrapolate to fine time scales by plotting two versions of 0.9, 0.99, and 0.999 quantiles for 10-second packet counts—one estimated from the NTT model and the other calculated from the actual 10-second data.

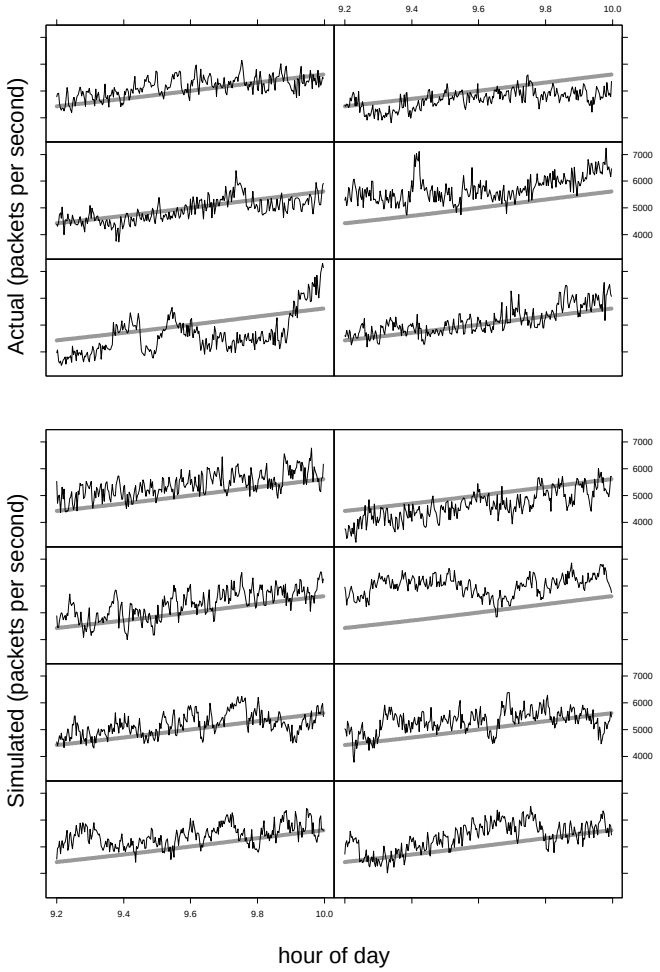

Fig. 6. Actual and simulated 10-second packet counts for a 45 minute period between 9 a.m. and 10 a.m. arranged as in Figure 5. The 10-second simulations are based on random bootstrap fits to 300-second data. The NTT model captures the upward trend, local variability, and wandering nature of the actual 10-second data.

The NTT-based estimates are 95% upper confidence bounds of the quantiles derived from the bootstrap fits to 300-second data. The actual quantiles are computed as follows: (1) subtract the estimated mean $\hat{\Lambda}(Tt)$ from the 10-second data across all days; (2) calculate quantiles from the residuals in one-hour time blocks; (3) add the estimated mean back into the residual quantiles. The plots show that the NTT estimates are in good agreement with the actual quantiles and they could well be used to provide slightly conservative load estimates by time of day for various traffic engineering problems. The actual .99 and .999 quantiles have aberrant values at 8 and 9 p.m. During this time period on one day, the link experienced unusually heavy loads for about

2 hours and this is readily apparent in plots of the 10-second data for that day. If these data are removed from the quantile calculations the agreement is very good.

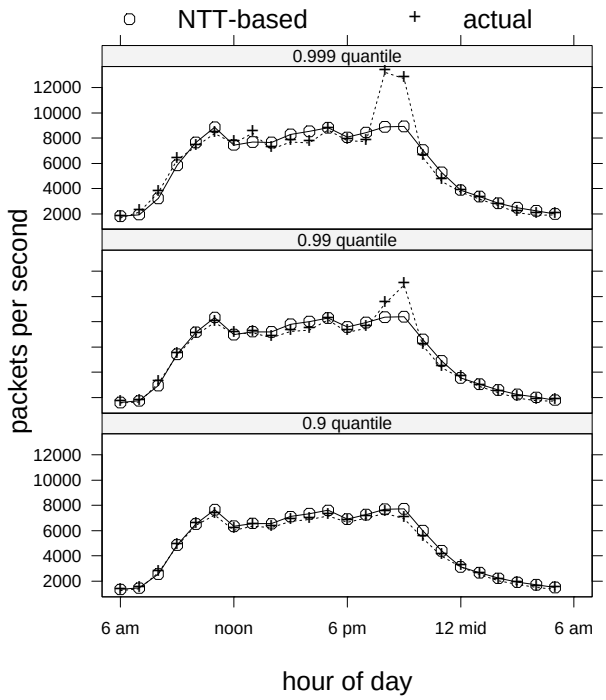

Fig. 7. Comparison of 10 second packet count quantiles. The actual quantiles are based on a 1 hour window of differences between 10 second packed counts and the estimated daily pattern. The NTT-based curves are 95% confidence bands of the quantiles derived from the bootstrap results of the model fit to 300-second packet counts.

Additional checks on the appropriateness of the nonstationary model are obtained by fitting it to disjoint blocks of data at successively finer time-scales starting with 300 s. Figure 8 displays the results of these model fits for the three parameters H, μ, and σ_y. The first column of panels plots the parameter estimates for fit to three blocks of 100-second packet counts. The second column shows estimates from 30 blocks of 10-second counts. Thin horizontal lines indicate the parameter estimates from 300 second counts over all hours of the day. Fitting the model to counts at progressively finer time scales demonstrates to what extent the parameters are independent of the time of day and the time-scale of the data. Each model fit used 288 packet counts values on each of seven days. Estimates of σ_e and α are not shown because they are so small.

The only clear indication of a non-constant parameter is seen in the upper right panel in which

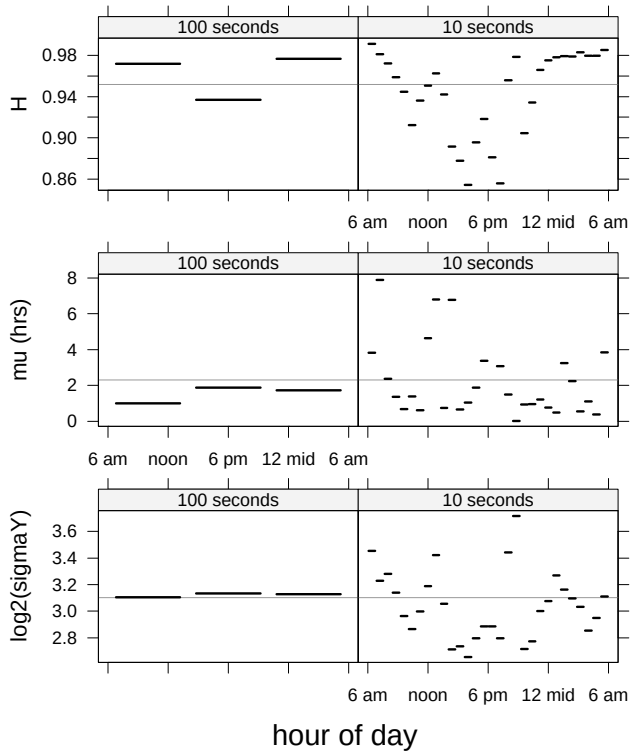

Fig. 8. Parameter estimates based on packet counts from different blocks time and at different aggregation time scales (100 seconds and 10 seconds). The thin gray lines represent parameter estimates from 300 second data across the whole day. The block-wise estimates, for the most part, do not change consistently with either time of day or aggregation scale with the exception that for 10 second data H appears to have lower values in the daytime than at night.

the estimate of H tends to be high in the nighttime hours with lower values during the day. Plots of data for each individual day (not shown here) reveal that the average rate during the hours after midnight tends to increase within each week, beginning with a low on Sunday night. This *day of week* pattern could easily be incorporated into the mean curve $\Lambda(Tt)$ and doing so would likely reduce both the magnitude and variation in fitted values of H. However, good day of week modeling is not possible using only the seven days of data available in this analysis.

VI. Discussion

We have proposed a Non-stationary Traffic Train model for analyzing sequences of packet counts on network links. The NTT model takes into account various sources of variation that have different interpretations and effects. First, a daily pattern is explicitly specified and estimated from observed data crossing multiple days. This makes it possible to distinguish low

frequency components of observed data from a deterministic daily pattern. Second, latent traffic train segments are introduced to provide a simple and realistic way of modeling the non-stationarity through the arrivals and departures of the trains. Third, each traffic train is treated as an fGn process to accommodate bursty traffic. We use fGn's because previous authors have shown them to describing fine time-scale characteristics of aggregate network traffic and they have an appealing derivation under a *byte-train* formulation with ON/OFF periods following long-tail distributions [14]. Finally, an additional uncorrelated noise term is included but found to be negligible compared with the other two random components listed above. The resulting NTT model and likelihood-based estimation promise sound and coherent statistical inference for fine time-scale behavior using coarse-scale traffic counts.

We demonstrated that the model performs well on a collection of CERNET packet count data. Actual 10-second packet count sequences are well represented by 10-second sequences simulated from NTT models that have been fit to 300-second data. Similarly, NTT-based quantiles of 10-second data for different times of the day closely track the sample quantiles of actual 10-second counts. The NTT model supports inference at *any* time-scale so the relevant scale can be used for traffic engineering even when data are only available at coarse scales such as with five-minute packet counts.

Blockwise fitting of the model to 10 second data indicates that the Hurst parameter H changes with time of day. We speculate that this may be caused by a trend, that we did not model, in which the traffic during night hours increases as the week progresses from Monday through Friday. In short, all weekdays are not the same and the mean curve $\Lambda(\cdot)$ should model these day of week effects in cases where enough data is available to establish the pattern. We have also considered modeling random daily effects in cases where little data is available or there is not obvious structure in the day-to-day changes.

The NTT Hurst parameter around .95 is larger than values typically reported in the literature, for example, [2], in the range of .75 to .85. Ignoring day of week effects probably explains some of this difference. Another contributing factor is that the NTT model dampens correlations in counts separated by a time that is large relative to μ. Thus, with stationary sequences, the NTT model will tend to match long-term correlations in the data by estimating a higher H relative to that needed for an fGn process. It would make an interesting study to compare fine time-scale

inference when using both fGn and NTT models in cases where the actual data is from one of the two models. Fine time-scale inference with a fGn model would be degraded by a deterministic trend in the coarse-scale aggregate counts and also by coarse-scale correlations that are not as strong as those at fine time-scales.

VII. Appendix

A. Proof of Proposition 1

Let $m_{t\to t+\Delta t}$ denote the number of trains that are active from time t to time $t+\Delta t$, and let $m_{t:t+\Delta t}$ denote the number of trains that start in the time interval $(t, t+\Delta t)$ and last at least until time $t+\Delta t$. First, we show that the distribution of the number of the trains $m_{t+\Delta t}$ at time $t+\Delta t$ given m_t, the number of the trains at time t, can be written as

$$m_{t+\Delta t}|m_t \stackrel{d}{=} m_{t\to t+\Delta t} + m_{t:t+\Delta t}$$

for $t \geq 0$ and $\Delta t > 0$. Let $p = \exp\{-\mu^{-1}\}$. Then

$$m_{t\to t+\Delta t} \sim \text{Binomial}\left(m_t, p^{\Delta t}\right)$$

because the exponential distribution of train lifetimes implies that each of the m_t trains has a probability $p^{\Delta t}$ of surviving at least an additional time $\Delta t > 0$. To obtain the distribution of $m_{t:t+\Delta t}$, we partition the interval $(t, t+\Delta t)$ into N equal-length intervals and *count* the number of trains starting in each of the N intervals. Let $n_t^{(N)}$ denote the sum of these counts and let $\Delta u = \Delta t/N$. Then $n_t^{(N)}$ follows the Poisson distribution

$$\text{Poisson}\left(\sum_{i=0}^{N-1} \lambda(t+i\Delta u)\Delta u p^{\Delta t - i\Delta u}\right),$$

where we used the fact that *if* $X \sim$ *Poisson* (λ) *and* $Y|X \sim$ *Binomial* (X, p) *then* $Y \sim$ *Poisson* (λp). Letting $N \to \infty$ shows that $m_{t:t+\Delta t}$ is

$$\text{Poisson}\left(p^{\Delta t}\int_0^{\Delta t} \lambda(t+u)\exp\{\mu^{-1}u\}du\right)$$

and the results follow obviously.

B. Autocovariance matrices of NTT model

First note that for $H = 1$ they are proportional—that is, $\sigma_y^{-2} C_T^{(y)}(s,t) = \alpha^{-2} C_T^{(z)}(s,t)$—and for this reason we derive only $C_T^{(y)}(s,t)$ for general H. Define μ such that $p = e^{-1/\mu}$ and define $\rho \equiv T/\mu$. Suppose the rate $\Lambda_{T(t+u)}$ is quadratic for $0 < u < 1$; namely, $\Lambda_{T(t+u)} = \sum_{i=0}^{2} a_i^{(t)} u^i$. The scaled variance is

$$\begin{aligned} \frac{C_T^{(y)}(t,t)}{\sigma_y^2 2H(2H-1)} &= T^{2H} \int_0^1 e^{-\rho w} w^{2H-2} \\ &\quad \left(\int_0^{1-w} \sum_{i=0}^{2} a_i^{(t)} v^i dv \right) dw. \end{aligned} \tag{13}$$

The parenthesized integral is cubic in w. Denote the coefficients by $b_i^{(t)}$:

$$\sum_{i=0}^{3} b_i^{(t)} w^i = \sum_{i=0}^{2} \left. \frac{a_i^{(t)} v^{i+1}}{i+1} \right|_{v=0}^{1-w}. \tag{14}$$

Substituting (14) into the scaled variance (13) gives

$$\frac{C_T^{(y)}(t,t)}{\sigma_y^2 2H(2H-1)} = T^{2H} \sum_{i=0}^{3} \frac{b_i^{(t)}}{\rho^{2H-1+i}} \Gamma_{2H-1+i}(0,\rho)$$

where $\Gamma_a(y,z)$ denotes the generalized incomplete gamma function

$$\Gamma_a(y,z) = \int_y^z t^{a-1} e^{-t} dt$$

for which $\int_y^z t^{a-1} e^{-bt} dt = b^{-a} \Gamma_a(by, bz)$. The final expression for $C_T^{(y)}(t,t)$, the variance of y_{Tt}, is thus

$$\sigma_y^2 2H(2H-1) \mu^{2H} \sum_{i=0}^{3} \frac{b_i^{(t)}}{\rho^{i-1}} \Gamma_{2H-1+i}(0,\rho). \tag{15}$$

The autocovariance of y_{Tt} is derived similarly. For $s \le t+1$

$$\begin{aligned} &\frac{2C_T^{(y)}(s,t)}{\sigma_y^2 2H(2H-1)T^{2H}} \\ &= \sum_{i=0}^{3} b_i^{s,t} \int_{t-s-1}^{t-s} e^{-\rho w} w^{2H-2+i} dw \\ &\quad + \sum_{i=0}^{3} c_i^{s,t} \int_{t-s}^{t-s+1} e^{-\rho w} w^{2H-2+i} dw \end{aligned} \tag{16}$$

where the parenthesized integrals have been expressed as cubic functions in w with coefficients $b_i^{s,t}$ and $c_i^{s,t}$; namely,

$$\sum_{i=0}^{3} b_i^{s,t} w^i = \sum_{i=0}^{2} \left. \frac{a_i^{(s)} u^{i+1}}{i+1} \right|_{u=t-s-w}^{1} \tag{17}$$

$$\sum_{i=0}^{3} c_i^{s,t} w^i = \sum_{i=0}^{2} \left. \frac{a_i^{(s)} u^{i+1}}{i+1} \right|_{u=0}^{1+t-s-w} . \tag{18}$$

Substituting (17) and (18) into (16) gives the final expression for the autocovariance of y_{Tt} (for $d \equiv t - s \geq 1$) as

$$\begin{aligned} C_T^{(y)}(s,t) &= \frac{\sigma_y^2 2H(2H-1)\mu^{2H}}{2} \\ &\quad \sum_{i=0}^{3} \left[\frac{b_i^{s,t}}{\rho^{i-1}} \Gamma_{2H-1+i}(\rho(d-1), \rho d) \right. \\ &\quad \left. + \frac{c_i^{s,t}}{\rho^{i-1}} \Gamma_{2H-1+i}(\rho d, \rho(d+1)) \right] \end{aligned} \tag{19}$$

REFERENCES

[1] Beran, J., Sherman, R., Taqqu, M. S., and Willinger, W., "Long range dependence in variable bit rate traffic," *IEEE Trans. Commun.*, vol. 43, pp. 1566-1579, 1995.

[2] Crovella, M. E. and Bestavros, A., "Self-similarity in World Wide Web traffic: evidence and possible causes," *IEEE/ACM Transactions on Networking*, vol. 5, pp. 835-846, 1997.

[3] Duffield, N., Massey, W. and Whitt, W., "A Nonstationary Offered-Load Model for Packet Networks," *Telecommunication Systems*, pp. 271–296, 2001.

[4] Efron, B. and Tibshirani, R. J., *An Introduction to the Bootstrap*. Chapman & Hall, New York, 1993.

[5] Eick, S. G., Massey, W. A. and Whitt, W., "The Physics of the $M_t/G/\infty/$ Queue," *Operations Research*, pp. 731–741, 1993.

[6] Feldmann, A., Gilbert, A. C. and Willinger W., "Data Networks as Cascades: Investigating the Multifractal Nature of Internet WAN Traffic," in: *Proc. ACM SIGCOMM* pp. 42–55, 1998.

[7] Feldmann, A., "Characteristics of TCP connection arrivals," in: *Self-similar Network Traffic and Performance Evaluation*, eds. K. Park and W. Willinger, John Wiley and Sons, pp. 367–399, 2000.

[8] Lehmann, E. L., *Testing Statistical Hypotheses*, John Wiley & Sons, New York, 1986.

[9] Leland, W. E., Taqqu, M. S., Willinger, W. and Wilson, D. V, "On the self-similar nature of Ethernet traffic," *IEEE/ACM Transactions on Networking*, vol. 2, pp. 1-15, 1994.

[10] Lucas, M. T. and Wrege, D. E. and Dempsey, Bert J. and Weaver, A. C., "Statistical Characterization of Wide-Area IP Traffic," *Proceedings of Sixth International Conference on Computer Communications and Networks (IC3N'97)*, pp. 442–447, 1997.

[11] Mandelbrot, B. B. and van Ness, J. W., "Fractional Brownian motions, fractional noises and applications," *SIAM Review*, vol. 10, pp. 422-37, 1968.

[12] Paxson, V. and Floyd, S., "Wide-Area Traffic: The Failure of Poisson Modeling," *IEEE/ACM Transactions on Networking*, 3(3), pp. 226–244, 1995.

[13] Thompson, K. and Miller, G. and Wilder, R., "Wide-Area Internet Traffic Patterns and Characteristics," *IEEE/ACM Transactions on Networking*, pp. 10–23, November 1997.

[14] Willinger, W., Taqqu, M. S., Sherman, R., and Wilson, D. V., "Self-similarity through high-variability: statistical analysis of Ethernet LAN traffic at source level,"" *IEEE/ACM Transactions on Networking*, vol. 5, pp. 71-86, 1997.

[15] Xiao, X. and Hanna, A. and Bailey, B. and Ni, L., "Traffic Engineering with MPLS in the Internet," *IEEE Network*, vol. 5, pp. 28–33, 2000.