Spectral Analysis of Queue Response to Correlated Traffic

Zhiang Deng

Supervisors: Carl McCrosky, Carey Williamson

Abstract

Queueing theory is often used to model and analyze the traffic in computer networks. Most queueing model studies are based on the independent arrivals assumption to simplify the analysis. Recent studies have shown that many network traffic streams are highly correlated. The assumption of independence can lead to very poor network traffic modeling when the traffic is correlated. When the correlation is introduced into traditional queueing models, the analytical tractability is lost. The objective of this project is to study the queueing responses (i.e., the mean queue length) to correlated input process using spectral analysis, and to use this method to analyze video traffic. To overcome the difficulty of the queueing analysis, a complex correlated input function is decomposed into a series of simple element functions. Therefore a complex process can be approximated by constructing a superposition of a sequence of simple periodic chains. This spectral analysis approach provides a new viewpoint for network modeling and analysis.

1 Introduction

Queueing theory has been widely used in traffic analysis for network and telecommunication systems [1]. Traditional queueing theory offers concise closed form expressions for calculating the queue performance statistics (e.g., mean queue length, average waiting time, etc.) as a function of the input load to the queueing system. In the past few years, there have been many studies on high speed network traffic modeling, control, and analysis. Most of the studies are based on several traditional queueing model assumptions, such as independent processes to simplify the queueing analysis. For example, for a Poisson process, the arrivals are assumed to be independent. The performance of single queues with independent inter-arrival intervals and service demands is well understood, and often analytically tractable. Little is known when autocorrelation is introduced into inter-arrival times or service demands: analytical tractability is lost. Even the simple case of the M/M/1 queue with autocorrelation is not well understood. Such correlations do, in fact, abound in real-life systems. The self-similarity of network traffic is a typical example [5, 9]. Studies have shown that traffic with self-similarity property is highly correlated. Independence assumptions can lead to very poor network traffic modeling. If a process is independent, then it must be uncorrelated [13].

This project studies the queue response (i.e., the mean queue length) to correlated input processes in computer networks using a method called spectral analysis. An N-state MMPP (Markov Modulated Poisson Process) is used as a statistical model. MMPP is commonly used in modeling multimedia traffic in high speed networks [1, 5, 12]. The definition of MMPP is discussed in the next section. Section 2 also gives the relation between the input correlation function and the transition matrix of MMPP, the power spectrum represented by the input rate function, and the concept of periodic chains. Section 3 analyzes the queue response to periodic chains. Video traffic over ATM network is examined in Section 4. Section 5 concludes the paper.

2 Approach and Methodology

Many modeling and analysis studies of high speed network traffic are based on several traditional queueing model assumptions, such as independent arrivals, to simplify the queueing analysis. The simplest model in queueing theory is the M/M/1 queue, where the arrival process is Poisson and the service time is exponentially distributed.

From actual network traffic measurements, however, many studies have shown that network traffic is highly correlated [6, 9]. For example, video traffic has been found to be correlated at substantially different time intervals, such as between lines, frames, and scenes [6, 8, 11]. Its correlation function may appear to be periodic, time varying, and much dependent on scenes in a particular video segment. Another example is the self-similar (or fractal) phenomenon in Ethernet traffic [9]. The bursty pattern exhibits self-similarity across a wide range of time scales. The independent arrivals assumption is no longer valid for these cases. However, little is known when autocorrelation is introduced into queueing processes; this remains as a key unsolved issue in queueing theory.

To overcome the difficulty of queueing analysis, a Markov Chain is used in this paper as a statistical model to approximate the correlation structure of the input process. The input Markov Chain provides a means to analyze the queue response to the input correlation functions. A typical multimedia traffic stream is modeled by a Markov Modulated Poisson Process (MMPP). The MMPP is a doubly stochastic Poisson process with arrival rates depending on the phases (states) that constitute the continuous time Markov Chain.

The MMPP is constructed in two levels:

- **Macro-level:** At this level, an N-state Markov Chain is used. Each state has its own input rate. The N input rate can be correlated.
- **Micro-level:** Each phase of the Markov Chain is associated with a constant input rate. The local dynamics of individual packet arrivals at micro level - i.e., each phase of the Markov Chain - is characterized by a Poisson process.

From the theory of Fourier transforms, the correlation function (in this project, correlation always means autocorrelation) in the time domain is the equivalent of the power spectrum in the frequency domain. Hence, like the method used to study general-order autoregressive processes in signal processing, the general N-state Markov Chain process can be used to construct the input power spectrum in rational function form.

2.1 Input Correlation Functions

In this section, the relation between the input correlation function and the transition matrix is discussed. The input correlation function can be expressed by the spectral factor, which is a function of input rates and transition matrix elements.

For an N-state MMPP, there is a constant input rate γ_i for each state i. The input rate vector is:

$$\vec{\gamma} = [\gamma_0, \gamma_1, \cdots, \gamma_{N-1}] \tag{1}$$

Since the micro-level process is Poisson, the service rate μ is a constant. In the following discussion, each element of the input rate is divided by the service rate, i.e., it equals to the utilization factor ρ:

$$\gamma_k = \rho_k = \frac{input\ rate}{service\ rate} \tag{2}$$

and

$$0 \leq \gamma_k \leq 1$$

for $k = 0, 1, \cdots, N-1$.

The input rate process is then described by $\gamma(n) = \gamma_i$ for the Markov Chain in state i at the nth time unit. The input rate correlation function is defined by [12]:

$$R(n) \equiv E(\gamma(m)\gamma(m+n)) = \sum_i \sum_j \gamma_i \gamma_j p_{i,j}^{(n)} \pi_i \tag{3}$$

where the element of the transition matrix at step n is:

$$p_{i,j}^{(n)} = Prob(\gamma(m+n) = \gamma_j | \gamma(m) = \gamma_i) \tag{4}$$

and

$$\pi_i = Prob(\gamma(n) = \gamma_i)$$

The input correlation function can be rewritten as:

$$R(n) = \sum_k \psi_k \lambda_k^{|n|} \tag{5}$$

with the spectral factor being defined as:

$$\psi_k = \sum_i \sum_j \gamma_i \gamma_j g_{li} h_{lj} \pi_i$$

where g_{li} and h_{lj} are the components of vectors $\vec{g}$ and $\vec{h}$ respectively. Assuming that the state transition matrix is diagonalizable, by spectral decomposition, the nth step transition matrix is given by:

$$P^{(n)} = \sum_j \lambda_j^n \vec{g}_j \cdot \vec{h}_j$$

where $\vec{g}_j$ and $\vec{h}_j$ are the right column eigenvector and left row eigenvector for the eigenvalue λ_j:

$$\vec{g}_j = [g_{j0}, g_{j1}, \cdots, g_{j(N-1)}]^T$$

$$\vec{h}_j = [h_{j0}, h_{j1}, \cdots, h_{j(N-1)}]$$

with $[\cdots]^T$ representing the transposed matrix of $[\cdots]$.

Each complex eigenvalue can be expressed as:

$$\lambda_k = |\lambda_k| e^{i\omega_k t} \tag{6}$$

and the spectral factor is:

$$\psi_k = |\psi_k| e^{i\theta_k t} \tag{7}$$

where t is a time unit.

Since $R(n)$ is real, the input correlation function of the input rate process is [12]:

$$R(n) = \sum_{k=0}^{N-1} |\psi_k||\lambda_k|^{|n|} cos(|n|\omega_k t + \theta_k t) \tag{8}$$

2.2 The Power Spectrum

The power spectral function of the input rate correlation function can be derived through Fourier transformation. For $R(n)$, the power spectral function is [12, 13]:

$$P(\omega) = \sum_{n=-\infty}^{+\infty} R(n)e^{-in\omega t} = \sum_{k=0}^{N-1} \frac{\psi_k(1-\lambda_k^2)}{1-2\lambda_k cos(\omega t)+\lambda_k^2} \tag{9}$$

Note that the process discussed here has a correlated input rates. Since $R(n)$ is discrete in the time domain with unit t, $P(\omega)$ must be periodic.

Similarly, for the continuous spectral domain, we can derive the power spectrum [7, 11]:

$$P(\omega) = 2\pi\overline{\gamma}^2\delta(\omega) + \overline{\gamma} + \overline{\gamma}^2 \sum_{k=0}^{N-1} \frac{2\lambda_k C_k^2}{\lambda_k^2+\omega^2} \tag{10}$$

with:

$$C_k^2 = \frac{1}{\overline{\gamma}^2} \sum_i \sum_j \pi_i \gamma_i \gamma_j g_{ki} h_{kj} \tag{11}$$

where $\delta(\omega)$ is the Dirac δ-function, which is defined as:

$$\delta(x) = \begin{cases} 1 & (x=0) \\ 0 & (otherwise) \end{cases}$$

The first term in Equation 10 is the *dc* term, i.e., the constant input rate term. Since the average arrival rate in the queueing system must be positive, the *dc* term should always exist in the input power spectrum. The second term is the “white noise”, representing the micro-level Poisson arrivals in each input state. The last term is the average power spectrum contributed by all λ_k.

Fluid flow approximation [2, 4, 5] is another widely used method of modeling network traffic and queueing analysis. Using the spectral analysis method one can easily show that the fluid flow approximation removes the micro-level effect (or “white noise”) from the input power spectrum.

The superposition of independent MMPPs is still an MMPP. For input traffic consisting of M independent MMPPs, the mathematical expression is [7]:

$$\begin{aligned} Q &= Q_1 \oplus Q_2 \oplus \cdots \oplus Q_M \\ \vec{\gamma} &= \vec{\gamma}_1 \oplus \vec{\gamma}_2 \oplus \cdots \oplus \vec{\gamma}_M \end{aligned} \tag{12}$$

where the i-th MMPP is defined by transition rate matrix Q_i and input rate $\vec{\gamma}_i$, and $\oplus$ is the Kronecker sum operator. The Kronecker sum operator is defined for two matrices (or vectors) A and B, as:

$$A \oplus B = [a_{ij} + b_{kl}]$$

If the state space of the i-th MMPP is N_i, the overall input state space is $N = N_1 \times N_2 \times \cdots \times N_M$. For traffic modeling in high speed networks, M usually represents the number of sources multiplexed on each link. The superposition leads to [7, 13]:

$$\begin{aligned} P(\omega) &= P_1(\omega) + P_2(\omega) + \cdots + P_M(\omega) \\ f(x) &= f_1(x) \star f_2(x) \star \cdots \star f_M(x) \end{aligned} \tag{13}$$

where the power spectrum $P_i(\omega)$ and the input rate function $(f_i(x)$ characterize the i-th MMPP and $\star$ is the convolution operator. The convolution of two (or more) functions is defined by:

$$f(x) = f_1 \star f_2 = \int_0^x f_1(x-t)df_2(t)$$

for continuous distributions, and

$$f(x = k) = f_1 \star f_2 = \sum_{i=0}^{k} f_1(x_1 = i) f_2(x_2 = k - i)$$

for discrete distributions.

2.3 The Input Periodic Chains

In the time domain, spectral decomposition is equivalent to replacing the original Markov Chain by a periodic chain, or a sequence of sinusoidal functions by using Fourier expansion. The advantage of this approach is that construction of a periodic chain from a given discrete input spectral function is quite simple. Hence, the periodic chain can be used as a handy method of modeling the input correlation functions. The terminology used to define the periodic chain is analogous to those basic periodic functions used in linear systems in signal processing, such as dc, sinusoidal, impulse (represented by δ-function), triangle pulse, rectangular pulse, and their superpositions. The queueing response to basic input correlation functions will be examined later.

The transition matrix P of an N-state periodic chain is defined by [12]:

$$p_{i,j} = \begin{cases} 1 & (for\ 0 \leq i \leq N-1,\ j = mod(i+1, N)) \\ 0 & (elsewhere) \end{cases}$$

Denote the input power spectral vector by:

$$\vec{\psi} = [\psi_0, \psi_1, \cdots, \psi_{N-1}]$$

which characterize the corresponding discrete input power spectrum. The spectral vector can be expressed in general format [12]:

$$\vec{\psi} = \frac{1}{N^2} |DFT(\vec{\gamma})|^2 \tag{14}$$

i.e., the power spectral vector is derived from the discrete Fourier transform of the input rate vector.

The correlation function of the periodic chain be expressed by the discrete Fourier expansion form:

$$R(n) = \sum_{k=0}^{N-1} \psi_k cos(n\omega_k t)$$

Therefore, the corresponding power spectrum is expressed by a sequence of discrete impulses [12, 13]:

$$P(\omega) = \frac{\pi}{t} \sum_{k=0}^{N-1} \psi_k (\delta(\omega - \omega_k) + \delta(\omega + \omega_k)) \tag{15}$$

The N discrete frequencies in $P(w)$ are evenly divided in the fundamental frequency band $(-\pi/t, \pi/t)$. The fundamental frequency is defined by

$$\omega_0 = \frac{2\pi}{tN}$$

The power at each discrete frequency is assigned by the power spectral vector $\vec{\psi}$. The average input rate is $\overline{\gamma} = \frac{1}{N} \sum_{k=0}^{N-1} \gamma_k$.

Once the power spectrum is transferred from the continuous frequency domain to the discrete frequency domain, the input Markov Chain is replaced by a periodic chain.

Note that the main interest here is to design a Markov Chain, or a superposition of Markov Chains, which match the input correlation function. The Markov Chain is just a means to analyze the queue response to input correlation functions.

Queueing performance is mainly dominated by input with high positive correlations [12]. From the spectral analysis we know that the higher the positive correlation in the time domain, the more low frequency powers in the frequency domain [13].

3 Queue Response to Periodic Chain Input

For a high speed network, the input rate process can be modelled by an N-state periodic chain. At the n-th state of the macro-level, the micro-level packet arrivals are Poisson with input rate γ_n.

Using the generating function method in queueing theory [1, 12, 10, 15], one can compute the mean queue length of the periodic chain [12]:

$$L = \frac{\overline{\gamma}^2}{2(1-\overline{\gamma})} + \frac{1}{1-\overline{\gamma}} \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} c_k \gamma(j,k) \tag{16}$$

where

$$\gamma(j,k) = \sum_{l=0}^{j} (\gamma_l - \overline{\gamma}) - \sum_{l=0}^{k} (\gamma_l - \overline{\gamma}) \tag{17}$$

and c_k is the element of the boundary vector $\vec{c}$.

In order to compute the queue response to a given input rate function numerically, the key issue is to determine the boundary vector. The boundary equation is [12, 15]:

$$\vec{c} \cdot \vec{g_k}(z_k) = 0 \tag{18}$$

i.e.

$$A\vec{c} = 0 \tag{19}$$

where A is an $N \times N$ matrix

$$A = [A_{kj}]$$

and

$$A_{kj} = Re(e^{\frac{i2\pi kj}{N}} e^{(1-z_k)\sum_{m=o}^{j}(\gamma_m - \overline{\gamma})}) \tag{20}$$

If the z_k's are known, given the input rate function, then the boundary equation is a group of linear equations and can be solved numerically. That is, one can get all c_k's.

To calculate the z_k, let:

$$z_k = u_k + iv_k \tag{21}$$

one can get

$$v_k = e^{[v_k ctg(\frac{2\pi k}{N} + v_k\overline{\gamma}) - 1]\overline{\gamma}} sin(\frac{2\pi k}{N} + v_k) \tag{22}$$

which can be solved numerically using the iteration method, and

$$u_k = v_k ctg(\frac{2\pi k}{N} + v_k\overline{\gamma}) \tag{23}$$

for $k = 0, 1, \cdots, N-1$.

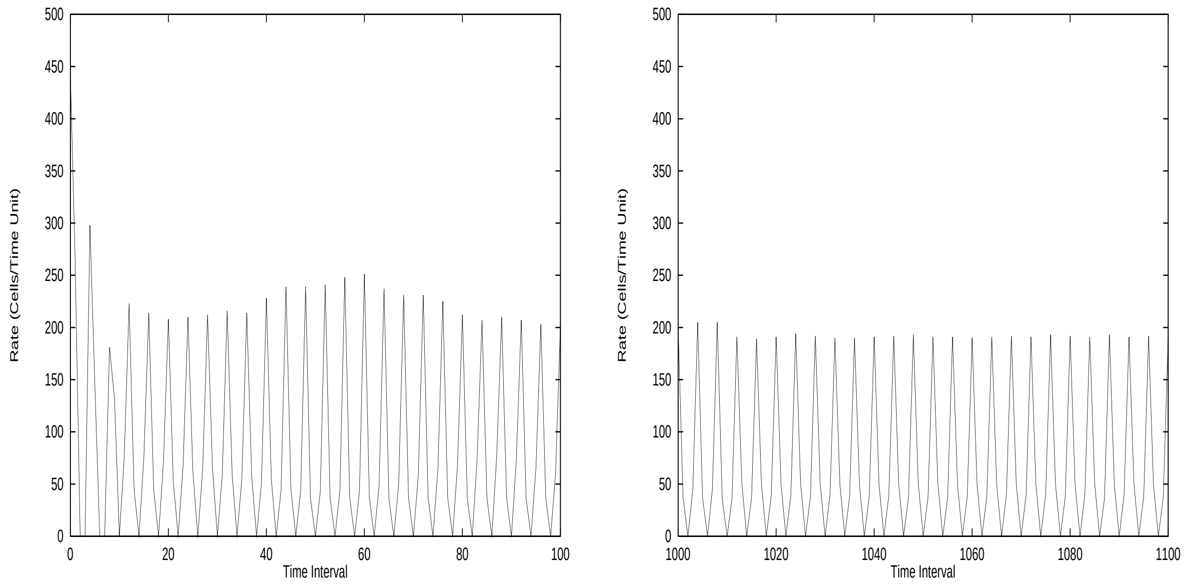

Figure 1: Cell input rate of the ATM video trace (Time unit = 0.01 second).

4 ATM VBR Video Traffic: An Example

Now let's look at an example of queue response to ATM video traffic. Suppose there is a FIFO buffer. The input is the video stream. The relation between the mean queue length and the input rate can be found.

The video trace used here was collected by the System Research Group at the University of Cambridge. The data and detail information can be found at *ftp.cl.cam.ac.uk* in the directory *fairdata/avatrace*. The data contains a 999,958 cell trace of JPEG compressed video. The duration is 155.45 seconds. Using the traffic analysis method in [8, 9, 16], one can compute the input cell rate function (Figure 1) and the cell rate correlation function (Figure 2).

Hurst parameter H $(0.5 \leq H < 1)$ is used to measure the degree of self-similarity of network traffic [9]. For independent traffic, $H = 0.5$. H is the slope of the least-squares fit of pox plot of R/S data on log-log axes [9]. For this video trace, the Hurst parameter is 0.83, which is larger than 0.5. Therefore this video traffic trace is self-similar. According to the property of self-similar traffic, it is a highly correlated process, as indicated by the cell rate correlation function (Figure 2).

Figure 1 shows two parts of the input cell rate. The input rate shows a sinusoidal pattern. Figure 2 (a) is the correlation of the input rate, which also shows a sinusoidal pattern. Figure 2 (b) is the correlation of a sinusoidal function $1 + sin(x)$.

A sinusoidal input rate process is defined as:

$$\gamma_k = \overline{\gamma}[1 + cos(\omega_i kT)]$$

for $k = 0, 1, \cdots, N-1$. The oscillation frequency ω_i is given by $i\omega_0$ for $i = 0, 1, \cdots, N-1$. All components of the input rate vector are positive.

The correlation function of the sinusoidal input is also sinusoidal. It can be expressed as [12]:

$$R(n) = \overline{\gamma}^2[1 + \frac{1}{2}cos(\omega_i nT)]$$

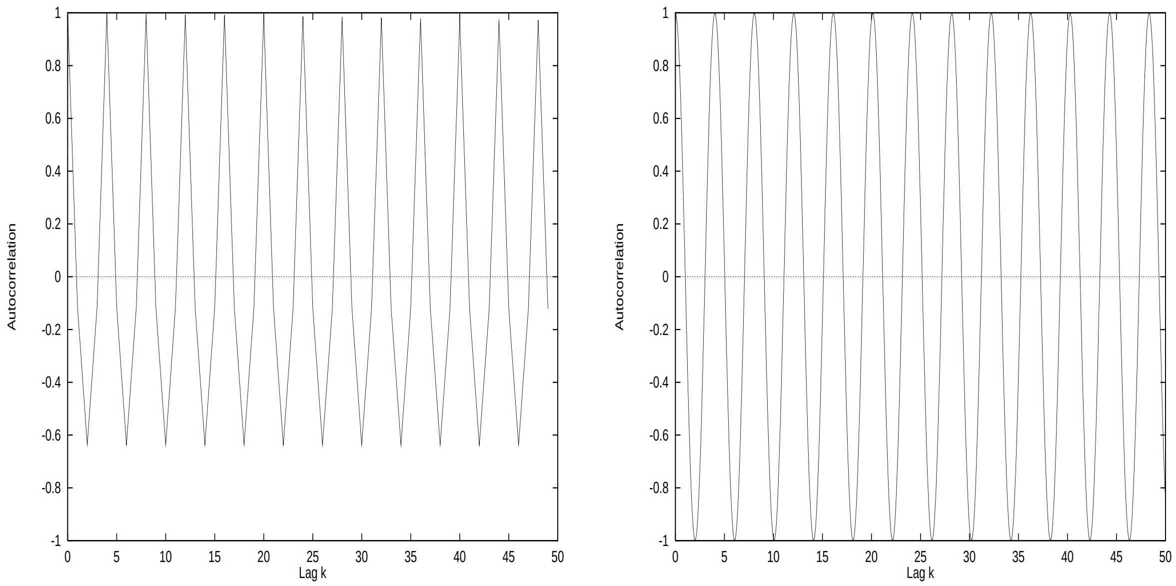

Figure 2: The autocorrelation of the input rate of the ATM video trace and function 1 + sin(kx).

and the power spectrum in the frequency domain is:

$$P(\omega) = \frac{2\pi\overline{\gamma}^2}{T}[\delta(\omega) + \frac{1}{4}\delta(\omega - \omega_i) + \frac{1}{2}\delta(\omega + \omega_i)]$$

If the correlation function has a sinusoidal form, it indicates that the input rate function is also sinusoidal. Both functions have the same sinusoidal frequency. For the ATM video trace, the input rate can be approximated by a sinusoidal periodic chain.

Further calculation [12] shows that the mean queue length is proportional to the inverse of sinusoidal frequency for the low frequency band. Also, the higher the average input rate $\overline{\gamma}$, the larger the mean queue length.

In Figure 3, mean queue lengths are given for two processes, one is MMPP with correlated input rate, another is an M/M/1 process which assumes the arrivals are independent. The results are significantly different. In general, if correlation is introduced, then the mean queue length is much longer. The result of mean queue lengths is consistent with the simulation results in [3].

Table 1: Mean queue length L (in cells) for MMPP and M/M/1 processes.

Utilization Factor	0.7	0.8	0.9	0.95
L (MMPP)	12	22	33	44
L (M/M/1)	2.33	4	9	19

For more complex input rate functions, such as in [8], the input rate can be expressed by the superposition of two or more sinusoidal or triangular functions.

Many studies have shown that high speed network traffic and queueing performance are mainly dominated by input streams with high positive correlations [6, 9]. The spectral analysis shows that the higher the positive correlation in the time domain, the more the input power is in low-frequency band [12, 11, 13]. Therefore, the input power in lower-frequency band dominates the queue performance.

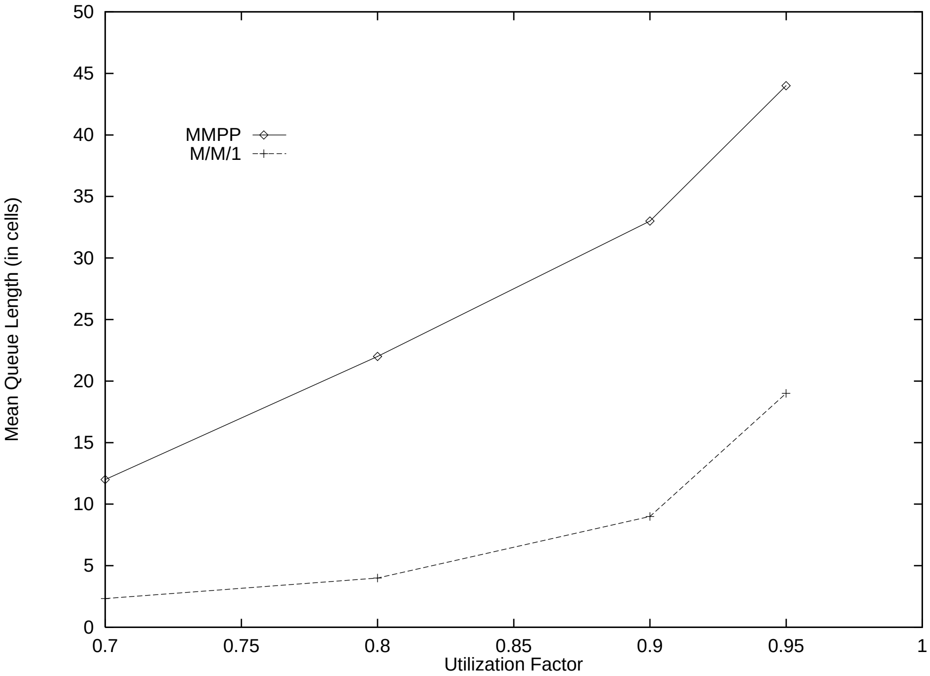

Figure 3: Mean queue lengths of the video trace for two processes.

This result guides us to focus on the low-frequency section in frequency domain of traffic modeling and analysis. The above comparison indicates that the queue length of the correlated input process tends to be longer than the independent process. For ATM networks with finite buffers, the simplified independent models such as Poisson may underestimate the queue length and the cell loss ratio.

5 Conclusion

This project studies the relation between the input rate function and the power spectrum. A complex input spectrum can be decomposed to simple elements. These elements in the frequency domain correspond to periodic chains in the time domain. Therefore, through the study of queue response to the simple periodic chains, and use of the spectral decomposition method, one can get the queue response to simple and complex correlated input. An ATM video trace is studied and the result suggests that the mean queue length of the correlated input of the self-similar traffic tends to be longer than the independent input process. It indicates that the self-similar traffic (which is highly correlated) has potentially serious implications on ATM networks. Understanding the queueing response to the correlated input will provide us a great deal of knowledge to develop advanced network control, measurement, and resources management. This method can be used to study many issues in high speed networks, such as admission control, routing, bandwidth allocation, network-wide traffic analysis [12, 14], and modeling the self-similar signals in broadband communication networks.

References

[1] Akimaru, H., and Kawashima, K. *Teletraffic: Theory and Applications.* Springer-Verlag, 1993.

[2] Anick, D., Mitra, D., and Sondhi, M. M. Stochastic theory of a data-handling system with multiple sources. *The Bell System Technical Journal 61*, 8 (Oct. 1982), 1871–1894.

[3] Chen, Y., Deng, Z., and Williamson, C. A model for self-similar ethernet traffic: design, implementation, and performance implications. To appear, Mar. 1995.

[4] Dravida, S., and Sasksena, V. R. Analysis and engineering of a voice/data packet multiplexer. *IEEE Transactions on Communications 41*, 11 (Nov. 1993), 1656–1667.

[5] Frost, V. S., and Melamed, B. Traffic modeling for telecommunications networks. *IEEE Communications Magazine 32*, 3 (Mar. 1994), 70–81.

[6] Garret, M., and Willinger, W. Analysis, modeling and generation of self-similar vbr video traffic. In *Proceedings of 1994 ACM SIGCOMM Conference on Communication Architectures and Protocols* (Aug. 1994), pp. 269–280.

[7] Hwang, C. L., and Li, S. Q. On input state space reduction and buffer non-effective region. In *Proc. IEEE INFOCOM '94* (Jun. 1994), vol. 3, pp. 1018–1028.

[8] Knightly, E. W., and Zhang, H. Traffic characterization and switch utilization using a deterministic bounding interval dependent traffic model. Tech. Rep. TR-94-047, ICSI and EECS Department, U. C. Berkeley, 1994.

[9] Leland, W., Taqqu, M., and Wilson, D. On self-similar nature of ethernet traffic (extended version). *IEEE/ACM Transactions on Networking 2*, 1 (Feb. 1994), 1–15.

[10] Li, S. Q. Generating function approach for discrete queueing analysis with decomposable arrival and service markov chains. In *Proc. IEEE INFOCOM '92* (May 1992), vol. 3, pp. 2168–2177.

[11] Li, S. Q., and Hwang, C. L. Queue response to input correlation functions: continuous spectral analysis. *IEEE/ACM Transactions on Networking 1*, 6 (Dec. 1993), 678–692.

[12] Li, S. Q., and Hwang, C. L. Queue response to input correlation functions: discrete spectral analysis. *IEEE/ACM Transactions on Networking 1*, 5 (Oct. 1993), 522–533.

[13] Priestley, M. B. *Spectral Analysis and Time Series*, vol. 1. Academic Press, 1981.

[14] Sheng, H. D., and Li, S. Q. Spectral analysis of packet loss rate at statistic multiplexer for multimedia services. In *Proc. IEEE Globecom '92* (1992), pp. 1524–1532.

[15] Sohraby, K., and Zhang, J. Spectral decomposition approach for transient analysis of multi-server discrete-time queues. *Performance Evaluation 21* (Nov. 1994), 131–150.

[16] Williamson, C., Arlitt, M., Chen, Y., Choy, H., Deng, Z., and Gurski, R. Model validation and test suite for atm-tn traffic models. *TeleSim Project* (Nov. 1994). Draft Version 2.0.