Link Capacity Allocation by Input Power Spectrum

Chia-Lin Hwang and San-qi Li

Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, Texas 78712, USA

Abstract

Consider a multimedia input traffic stream, described by a general stationary random process. Our previous study indicates that second-order input statistics, which are described by power spectrum in the frequency domain, play a dominant role in queueing performance as compared to higher-order input statistics. In this short paper we show the significant effect of input power spectrum on link capacity allocation in a finite buffer system, to achieve a desired performance criterion. The study also explores a fundamental limit of buffer sizing to the link capacity improvement when more input power is in the low frequency band.

1. INTRODUCTION

Link capacity allocation in high speed multimedia service networks represents one of the most challenging tasks in advanced telecommunication research. The central objective of link capacity allocation is two-fold: to take advantage of statistical multiplexing for transmission efficiency, and to avoid nodal congestion caused by the arrival of unpredicted bursty traffic. The existing studies [1, 2, 3, 4] used approximation analysis, asymptotic analysis and computer simulations for the link capacity allocation, mainly based on two-state Markov chain input models.

Traffic modeling consists of two basic components: traffic measurement and queueing analysis. In the network performance evaluation field, traffic measurement has long been neglected. Consider a multimedia input stream, generally described by a stationary random process. It is difficult to obtain an exact description of random traffic; only its steady state, second and higher order statistics are measurable. Steady state statistics are defined by the distribution function. Second and higher order statistics are described

---

1The research reported here was supported by NSF under grants NCR-9015757 and by Texas Advanced Research Program under grant TARP-129.
by the autocorrelation functions in the time domain, or by equivalent spectral functions in the frequency domain such as power spectrum, bispectrum and trispectrum. There are well known signal processing techniques available to measure traffic statistics. Second-order statistics, i.e., the power spectrum, can be measured by many sophisticated software packages or in hardware chips.

The concept of spectral representation of multimedia traffic in queueing analysis was first introduced in [5, 6, 7]. In [7] we used a special class of Markov chain, circulants, to construct input processes. The construction allows us to individually tune the power spectrum, bispectrum, trispectrum and distribution of the input traffic. In [7] we examined the effect of each individual input statistical function on characteristics of queue and loss rate. The study indicates that the queueing performance is much more sensitive to the second order input statistics than that to the higher order ones. Especially the low frequency input power has a dominant impact on queueing performance.

In this short paper we will assess the minimum link capacity requirement by a given input power spectrum to achieve a desired performance criterion. Such an issue is directly related to the design of call admission and resource management in high speed networks for support of multimedia services. Consider a single-server finite-buffer system, as shown in Figure 1. Denote the link capacity by $\mu$ and buffer size by $K$ measured in packets. For simplicity, consider a single bell power spectrum, defined by the central frequency $\omega_1$, bell bandwidth $B_1$ and normalized average power $C_\gamma$. The effect of $\omega_1$, $B_1$ and $C_\gamma$ on the minimum link capacity will then be examined, subject to a given mean queue length or average loss rate in a finite buffer system. The study reveals the significance of input power spectrum in link capacity allocation to provide guaranteed quality services.

The paper is organized as follows. Section 2 shows the construction of a single-bell input power spectrum by an MMPP. Link capacity allocation by the input power spectrum will then be examined in Section 3. Section 4 is the conclusion.

2. INPUT POWER SPECTRUM

Let input traffic be modeled by a Markov modulated Poisson Process (MMPP). The underlying $N$-state homogeneous Markov chain is described by a transition rate matrix $Q$. 

Figure 1: A single bell input power spectrum
The Poisson input rate at each state is defined by an input rate vector $\bar{\gamma} = [\gamma_0, \gamma_1, \ldots, \gamma_{N-1}]$. Assume that $Q$ is diagonalizable. The autocorrelation function of MMPP is then expressed by

$$R(\tau) = \bar{\gamma} \delta(\tau) + \bar{\gamma}^2 + \sum_{l=1}^{N-1} \psi_l e^{\lambda_l \tau}$$

where $\lambda_l$ is the $l$-th eigenvalue of $Q$ which can be complex and $\delta(\tau)$ is an impulse function. Taking the Fourier transform, we have the power spectrum of MMPP equal to

$$P(\omega) = \bar{\gamma} + 2\pi \bar{\gamma}^2 \delta(\omega) + \sum_{l=1}^{N-1} \psi_l b_l(\omega)$$

with

$$b_l(\omega) = \frac{-2\lambda_l}{\lambda_l^2 + \omega^2} \quad \text{and} \quad \frac{1}{2\pi} \int_{-\infty}^{+\infty} b_l(\omega) d\omega = 1.$$  

Refer to [7] for the details. The first component in (1), $\bar{\gamma}$, is described as the background white noise, which is attributed to the local dynamics of Poisson arrivals in each input state. The second component, $2\pi \bar{\gamma}^2 \delta(\omega)$, represents the DC term which is attributed to the average input rate. If the arrival process is purely Poisson, the power spectrum will only consist of white noise and DC. In general, each eigenvalue $\lambda_l$ for $l > 0$ will contribute a new component $\psi_l b_l(\omega)$ to $P(\omega)$. From (2) one can interpret $\psi_l$ as the average power contributed by $\lambda_l$. Each component represents a bell-shaped curve located at the central frequency $\omega_l = \text{Im}\{\lambda_l\}$ and weighted by the average power $\psi_l$. The shape of each bell, before being weighted by $\psi_l$, is measured by its half power bandwidth $B_l = \frac{\text{Re}\{\lambda_l\}}{\omega_l}$. Both $\omega_l$ and $B_l$ are defined by the radian frequency. In principle, we can construct a desired input power spectrum in rational function form with a sufficient number of eigenvalues. It is the eigenstructure of $Q$ that captures the input spectral properties. In general, finding the spectral functions from a given $(\bar{\gamma}, Q)$ is not difficult. The difficult part is its inverse: constructing $(\bar{\gamma}, Q)$ from given spectral functions. The inverse eigenvalue problem is much more difficult than the problem of finding eigenvalues in $Q$ matrix.

Recently we have introduced a new technique to construct a special Markov chain, called circulant, from a given pair of complex eigenvalues [7]. In a circulant matrix, each row circulates one element to the right to form the next row. Denote the first row in $Q$ by vector $\vec{a} = [a_0, a_1, \ldots]$, where $a_k$ is the transition rate from state $i$ to state $(i + k) \mod N$, $\forall i$. As a circulant, $Q$ will be replaced by $\vec{a}$. Let us denote the single pair of complex eigenvalues by $\lambda_i$ and its conjugate $\lambda_{N-i}$, where $i$ and $N - i$ are the index of the effective eigenvalue in $\vec{\lambda}$. The rest of the eigenvalues in $\vec{\lambda}$ can be made ineffective in the power spectrum simply by assigning the corresponding $\psi_l$'s equal to zero in (1). The corresponding bell component for $\omega \geq 0$ will then be described by its central frequency $\omega_i = \text{Im}\{\lambda_i\}$ and half-power bandwidth $B_i = 2|\text{Re}\{\lambda_i\}|$, as described in Figure 1. Taking
\[ a_k = \begin{cases} 
-\alpha_1 - \alpha_{N-1} & \text{at } k = 0 \\
\frac{1}{2} \left[ B_1/2 \cos(2\pi k/N) + \frac{\omega_1}{\sin(2\pi k/N)} \right] & \text{at } k = 1 \\
\frac{1}{2} \left[ \frac{B_1/2}{\sin(2\pi k/N)} - \frac{\omega_1}{\sin(2\pi k/N)} \right] & \text{at } k = N - 1 \\
0 & \text{else} 
\end{cases} \]  
\tag{3}

To ensure that both \( a_1 \) and \( a_{N-1} \) are non-negative, the size \( N \) must be large enough to satisfy

\[ \frac{B_1}{\omega_1} \geq 2|\tan(\pi/N)|. \]

The corresponding input rate vector \( \tilde{\gamma} \) is given by

\[ \gamma_k = \overline{\gamma}[1 + \sqrt{2}C_\gamma \cos(2\pi k/N - \theta)], \quad \text{for } 0 \leq k < N \]  
\tag{4}

where \( \theta \) is a phase angle, \( \overline{\gamma} \) is the average input rate and \( C_\gamma \) is the input rate variation coefficient. We also have \( \psi_1 = \psi_{N-1} = \frac{1}{2} \overline{\gamma} C_\gamma^2 \) for the average power of each bell component. Subject to \( \gamma_k \geq 0 \), we must have \( C_\gamma \leq \frac{1}{\overline{\gamma}} \). The corresponding input rate distribution in steady state is expressed by a sinusoidal plus the DC term.

For a real eigenvalue, one can use a number of i.i.d. 2-state Markov chains to construct a bell component centered at zero frequency (i.e., \( \omega_1 = 0 \)). Each 2-state Markov chain is defined by

\[ Q = \begin{bmatrix} -\beta & \beta \\
\alpha & -\alpha \end{bmatrix}, \quad \tilde{\gamma} = [0, \gamma_{on}], \quad \tilde{\pi} = [1 - \epsilon, \epsilon] \]  
\tag{5}

where \( \gamma_{on} \) is the Poisson input rate while in ON-state and \( \epsilon = \frac{\beta}{\beta + \alpha} \) is the probability in ON-period. The power spectrum superimposed by \( M \) i.i.d. 2-state Markov chains is then expressed by

\[ P(\Omega) = 2\pi \overline{\gamma}^2 \delta(\Omega) + \overline{\gamma} + \frac{B_1 C_\gamma^2 \overline{\gamma}^2}{(B_1/2)^2 + \omega^2} \]  
\tag{6}

with respect to

\[ \overline{\gamma} = M \epsilon \gamma_{on}, \quad B_1 = 2(\alpha + \beta), \quad C_\gamma = \sqrt{\frac{1 - \epsilon}{M \epsilon}} \]  
\tag{7}

The average power of the bell is measured by \( \overline{\gamma}^2 C_\gamma^2 \); the shape of the bell is described by the bandwidth \( B_1 \). \( B_1 \) is also directly related to the single non-zero eigenvalue of each 2-state MC: \( \lambda = -(\alpha + \beta) = -B_1/2 \). The distribution function of the aggregate input rate is binomial, as expressed by

\[ Pr(\tilde{\gamma} = k \gamma_{on}) = \binom{M}{k} \epsilon^k (1 - \epsilon)^{M-k} \quad \text{for } k = 0, 1, \ldots, M \]  
\tag{8}
Figure 2: Normalized link capacity for a single bell input power spectrum at $\bar{q} = 30$

Note that for each given $M$, a 2-state MC will be uniquely determined by $P(\omega)$:

$$
\beta = \frac{B_1}{2(1 + MC_\gamma^2)}, \quad \alpha = \frac{B_1MC_\gamma^2}{2(1 + MC_\gamma^2)}, \quad \gamma_{on} = \left(\frac{1}{M} + C_\gamma^2\right)\bar{\gamma}
$$

(9)

Hence, the freedom in selection of $M$, after matching with the power spectrum, is used to tune the input rate resolution of the binomial distribution function. Note that both the mean and variation coefficient of the distribution function are already fixed by the power spectrum. As $M \to \infty$, the binomial function will approach to a normal distribution. The study in [8] indicates that choosing $M = 30$ is sufficient for queueing analysis. In other words, the queueing solutions will hardly be affected by further increasing $M$ once $P(\omega)$ is fixed. Note that one can also use a circulant to construct a central-bell power spectrum.

3. LINK CAPACITY ALLOCATION

Let us model the queueing system by an MMPP/M/1/K queue which has the structure of finite quasi-birth-death (QBD) process. One can then use the QBD-Folding algorithm [9, 10] to evaluate both queueing delay and loss rate performance. The buffer size is assumed to be $K = 255$. Denote the normalized link capacity by $\rho^{-1} = \mu/\bar{\gamma}$.

In the numerical study of link capacity allocation we first consider a single bell input power spectrum as constructed by the circulant in Section 2, described by the central frequency $\omega_1$ and bell bandwidth $B_1$. The input variation coefficient is fixed at $C_\gamma = 0.7$. Assume the performance criterion is set by mean queue length $\bar{q} = 30$. Figure 2 shows the normalized minimum link capacity required by a single bell input power spectrum as a function of $\omega_1/2\pi\mu$ and $B_1/2\pi\mu$. Note that the queue is dependent on $\omega_1/2\pi\mu$ and $B_1/2\pi\mu$ instead of $\omega_1$ and $B_1$, due to the necessary frequency normalization for queueing analysis [7]. The contour in Figure 2 explore the remarkable impact of the input power spectrum on link capacity allocation. For each given $B_1$, the normalized link capacity $1/\rho$ increases by the reduction of $\omega_1$, which essentially shifts the bell in $P(\omega)$ to the low frequency band. As a result, more input power is moved from the high-frequency band to the low frequency band, which inherently causes the queueing performance to
deteriorate [5]. The impact of the low central frequency $\omega_1$ is especially strong when the bandwidth $B_1$ is also small. This is because the smaller the $B_1$, the more the input power is concentrated in the neighborhood of the lower $\omega_1$. For each given $\omega_1$, we also observe that the link capacity reaches its maximum around $B_1 = 2\omega_1$. This is because the input power at zero frequency, $P(0)$, reaches its maximum when $B_1 = 2\omega_1$ for each given $\omega_1$. In other words, as $B_1$ is close to $2\omega_1$, more input power will be located around the zero frequency, which causes the link capacity to increase. Also displayed in Figure 2 is the loss rate performance subject to $\bar{q} = 30$. As one can see, for a given average queue length, the loss rate performance will still be strongly dependent on the input power spectrum.

Some results are not available in Figure 2 at the corner where $B_1$ is small and $\omega_1$ is large, due to the limit of the circulant size. Similar results are obtained in Figure 3 when the performance criterion $\bar{q} = 30$ is replaced by the average loss rate $L = 10^{-6}$. The greater the input power in the low frequency band, the less the average queue should be selected for a given loss rate.

The study reveals the significant difference in the selection of delay and loss rate as a performance criterion. It should be pointed out that virtually no buffering is needed in Figures 2 and 3 as the link capacity increases up to $1/\rho = 2$. From (4), having $1/\rho = 2$ at $C_\gamma = 0.7$ is equivalent to assigning the link capacity by the maximum input rate $2\bar{q}$. Both results in Figures 2 and 3 indicate that the statistic gain through buffering diminishes rapidly as the input power increases in the low frequency band.

Next, consider a central-bell power spectrum as constructed by the superposition of 30 i.i.d. 2-state MCs in Section 2. Similarly, the performance criterion is set by $L = 10^{-6}$ at $K = 255$. Figure 4 plots are the results of the maximum throughput $\rho_{\text{max}}$ as a function of the normalized bandwidth $B_{\frac{1}{2\pi \mu}}$ and input rate variation coefficient $C_\gamma$. Notice that the maximum throughput is equivalent to the minimum link capacity. Examine the effect of $B_{\frac{1}{2\pi \mu}}$ on $\rho_{\text{max}}$ at each given $C_\gamma$. One can divide the value of $B_{\frac{1}{2\pi \mu}}$ into three regions: $(0, 1e^{-4}], (1e^{-4}, 1), [1, \infty]$, which are referred to as low-, mid- and high-frequency regions, respectively. When $B_{\frac{1}{2\pi \mu}} \in [1, \infty]$, most input power is located in the high frequency band. As a result, $\rho_{\text{max}}$ is close to 1 and much less dependent on $C_\gamma$. Especially as $B_{\frac{1}{2\pi \mu}} \to \infty$, $P(\omega)$ becomes a white noise. Correspondingly, the Poisson input rate vector $\bar{\gamma}$ will be governed by a renewal process, instead of MC. In this high-frequency region, a queuing system designed at $K = 255$ for $L \leq 1e^{-6}$ can always operate in the nearly full load
range. In contrast, when $\frac{B_1}{2\pi\mu} \in (1e^{-4}, 1)$, $\rho_{\max}$ becomes strongly affected by $\frac{B_1}{2\pi\mu}$. In this mid-frequency region, the system may operate in a much low throughput range, especially when more input power is moved in the low frequency band as $\frac{B_1}{2\pi\mu}$ reduces. Yet, once $\frac{B_1}{2\pi\mu} \in (0, 1e^{-4})$ in the low frequency region, $\rho_{\max}$ converges rapidly at each given $C_\gamma$.

In practice, one can expect that the condition $\frac{B_1}{2\pi\mu} < 1e^{-5}$ to a large extent can be held for the transmission of multimedia traffic on a high-speed link, especially as the link capacity ever increases and the multimedia traffic tends to be more correlated. For example, from the existing voice/video statistics one can always show $B_1 < 10$ radians based on the 2-state Markov chain modeling. Further, a 100Mbps ATM link is equivalent to $\frac{1}{2\pi\mu} = 6.7e^{-7}$, where $\mu$ is measured by number of cells per second.

Let us explain why $\rho_{\max}$ converges quickly once $\frac{B_1}{2\pi\mu} \leq 1e^{-4}$, and where it converges. Consider an extreme case, called a “zero-buffer” system, where the average loss rate $L$ is determined by the input rate density distribution function $f(x)$. In reality, the actual buffer size of this “zero-buffer” system should never be zero. We always need a small amount of buffer capacity to smooth out the local dynamics of Poisson arrivals and exponential service time, in order for $L$ to be determined by $f(x)$. For simplicity we refer to such a system as $K \approx 0$. This is analogous to the fluid-flow approximation in queueing analysis [11]. As described in Figure 5a, the loss occurs whenever the input rate $x$ exceeds the service rate $\mu$. $L$ will equal the portion of the tail distribution for $x > \mu$. Finding $\rho_{\max}$ for such a “zero-buffer” system is equivalent to finding the minimum service rate $\mu_{\min}$ at a fixed $L$, i.e., $\rho_{\max} = \frac{L}{\mu_{\min}}$. In our case, $f(x)$ is binomial and uniquely determined by $P(\omega)$ and $M$. It is thus straightforward to find $\rho_{\max}$ at $K \approx 0$. The two solution curves at $K \approx 0$ and $K = 255$ are plotted in Figure 5b for comparison. The one at $K = 255$ is directly copied from Figure 4 in the low frequency region at $\frac{B_1}{2\pi\mu} = 1e^{-7}$; the one at $K \approx 0$ is found from the corresponding $f(x)$ at $M = 30$. The small margin between the two curves shows that the throughput is marginally improved by increasing the buffer size from $K \approx 0$ to $K = 255$, under the same condition $L = 1e^{-6}$. It explores a fundamental limit of buffer sizing to the maximum throughput improvement, especially when more input power is in the low frequency band.

Similar results can be obtained for a multi-bell input power spectrum. One way to
Figure 5: Link capacity allocation at $L = 1e^{-6}$ (a) input rate density function, (b) comparison of $\rho_{\text{max}}$ at $K \approx 0$ and $K = 255$

construct a multi-bell input power spectrum is by the superposition of circulants. For instance, describe the $l$-th bell component by $(\psi_l, \lambda_l)$. The circulant for the $l$-th bell component is denoted by $(Q_l, \tilde{y}_l)$ with $Q_l = \text{circ}(\tilde{a}_l)$, which can be constructed as in the single-bell case. For an $M$-bell power spectrum, the superimposed circulant will then be given by

$$Q = \bigoplus_{l=1}^{M} Q_l, \quad \tilde{y} = \bigoplus_{l=1}^{M} \tilde{y}_l$$

where the notation “$\bigoplus$” is a Kronecker sum operator of vectors. If the size of $Q_l$ is $N_l \times N_l$, the size of $Q$ will be $N \times N$ with $N = \prod_l N_l$, which can be too large to numerically evaluate the queueing solutions. In practice we have used the QBD-Folding algorithm to find queueing solutions for $N < 1200$ on a SUN workstation. Refer to [7] for the details.

4. CONCLUSION

Traffic modeling consists of two basic components: traffic measurement and queueing analysis. In the network performance evaluation field, traffic measurement has long been neglected. In [7] we introduced a new technique to model input traffic statistics by a special class of Markov chains, and examined the individual effect of input statistics on characteristics of queue and loss rate in a finite buffer system. The input power spectrum, i.e., second order statistics, was identified to have a dominant impact on queueing system performance. In this paper we have further examined the significant effect of input power spectrum on link capacity allocation design subject to a certain delay/loss criterion. The study has explored a fundamental limit of buffer sizing to the throughput improvement when more input power is in the low frequency band. Recently the concept of spectral representation was used to characterize the transient loss behavior in system congestion period [12], and also to capture the performance constraints of input rate control in high speed networks [13].
References


