A Simple Formula for End-to-End Jitter Estimation in Packet-Switching Networks

Olivier Brun, Charles Bockstal, Jean-Marie Garcia

LAAS-CNRS

Toulouse, France

{brun,jmg,cbocksta}@laas.fr

Abstract

In this paper, an analytic approximation is derived for the end-to-end delay-jitter incurred by a periodic traffic with constant packet size. It is assumed that the periodic traffic is multiplexed with a background packet stream under the FCFS service discipline in each queue along the path to its destination. The processes governing the packet arrivals and the packet sizes of the background traffics are assumed to be general renewal processes. A very simple analytical approximation is derived and its accuracy is assessed by means of event-driven simulations.

1 Introduction

The Internet is expected to become a global communication infrastructure supporting real-time multimedia applications with stringent Quality-of-Service (QoS) requirements, and in particular real-time audio and video applications. Many real-time speech or video applications are generating Constant Bit Rate (CBR) traffic streams. An important performance measure for such traffics is the jitter which can be defined as the distortion on the periodic nature of the packet stream between the source and the destination. This distortion of the original pattern is due to the series of multiplexing operations performed by network routers.

In the last decade, many studies have been devoted to the estimation of jitter in packet-switched networks. Most of these works have been done in the context of ATM networks and therefore assume discrete time models. Jitter has been analyzed by means of event-driven simulations [3, 2]. Bounds have also been derived, either in a deterministic setting [4, 5, 6, 7] or in a stochastic setting [8, 9], but they are considered too loose for practical purposes. Other previous works are based on queueing theory [10, 11, 12, 13, 14, 15, 16, 17, 18, 20, 21, 22]. Exact and approximate numerical methods have been developed for the single node case. In [15], the node-by-node approximation of jitter is used for the numerical evaluation of the end-to-end jitter. The main drawback of these numerical models is that computations are typically very expensive. To the extend of our knowledge, only [19] proposes an analytic formula of end-to-end jitter based on heavy and light traffic approximations. The range of validity of this formula is however limited.

In this paper, a simple analytic approximation is derived for the end-to-end delay-jitter incurred by a periodic traffic having a constant packet size. We assume that the periodic traffic is multiplexed with a background traffic under the FCFS service discipline. The only assumption regarding the background traffic is that the processes governing packet arrivals and packet sizes are renewal processes. The accuracy of the approximation is assessed by means of numerous event-driven simulations.

The paper is organized as follows. Our assumptions and notations are introduced in the following section. Section 3 introduces the exponential approximation for tail probabilities of the steady-state waiting time that will be used throughout the paper. Section 4 is devoted to the single-node case and section 5 extends the approximation of the single-node jitter to the multiple-node case. Finaly, some conclusions are drawn in section 6.

2 Problem statement

We consider a tagged traffic passing through n tandem queues. This tagged traffic is assumed to be initialy periodic, i.e. we assume a constant time $T = 1/\lambda_T$ between packet arrivals at node 1. All the packets of the tagged traffic have the same size: D data units. The service rate of queue $k = 1 \dots n$ is μ_k data units per second.

In each queue $k = 1, \dots, n$, the tagged traffic is multiplexed under the FCFS discipline with a background traffic. Packet arrivals of the background traffic as well as packet sizes are assumed to be governed by general re-

newal processes, with respective probability density functions (pdf) $a_k^{\text{bg}}(t)$ and $b_k^{\text{bg}}(x)$. Let λ_k^{bg} denotes the packet rate of the background traffic at queue k and let x_k^{bg} denotes its average packet size (in data units). The global utilization factor of queue k is therefore $\rho_k = \rho_k^{\text{bg}} + (\lambda_T D)/\mu_k$ where $\rho_k^{\text{bg}} = \lambda_k^{\text{bg}} x_k^{\text{bg}}/\mu_k$ is the offered load of the background traffic. In the following, $\rho_k < 1$ is assumed for $k = 1, \ldots, n$.

Let us consider two consecutive packets of the tagged traffic, C_0 and C_1. For packet C_j, $j = 0, 1$ and node $k = 1, \ldots, n$, let $\tau_j^{\text{in}}(k)$ and $\tau_j^{\text{out}}(k)$ be the arrival and departure times of C_j at node k, let $W_j(k)$ be the waiting time of C_j at node k, and finally let $\Delta_k = W_1(k) - W_0(k)$ be the variation of the inter-packet delay at node k.

The end-to-end jitter is defined as,

$$J_{[1\ldots n]}(T) = E\left[\,|(\tau_1^{\text{out}}(n) - \tau_0^{\text{out}}(n)) \,-\, T\,|\,\right]$$

Under our assumptions, it is easy to show that,

$$
\begin{aligned}
J_{[1\ldots n]}(T) &= E\left[\,|\sum_{k=1}^{n}(W_1(k) - W_0(k))\,|\,\right] \\
&= E\left[\,|\sum_{k=1}^{n}\Delta_k\,|\,\right]
\end{aligned}
\tag{1}
$$

The end-to-end jitter is therefore given by the expected absolute value of the sum of inter-packet delay variations introduced by each node along the path between the source and the destination.

3 Exponential approximation for tail probabilities of the waiting time

It is well-known that tail probabilities of the steady-state waiting time W in GI/GI/1 queues often have approximately an exponential form, i.e.

$$P[W > x] = \alpha\, e^{-\eta x} \tag{2}$$

for suitably large x, where the decay rate η and the constant α are fixed positive real numbers. Such approximations are largely available in the queueing litterature and are known to often have a remarkable quality even for small values of x. In the sequel, we will use the approximation proposed by Abate et al. [1]. Due to space limitations, we omit the details of this approximation.

To assess the quality of this approximation, let us consider the simple case of a M/D3/1 queue, i.e. $a(t) = \lambda\, e^{-\lambda t}$ and $b(x) = p_1\, u_0(x - T_1) + p_2\, u_0(x - T_2) + p_3\, u_0(x - T_3)$, where u_0 denotes the unit impulse. The following values are assumed: $T_1 = 1$, $p_1 = 0.6$, $T_2 = 5$, $p_2 = 0.3$, $T_3 = 10$

and $p_3 = 0.1$. For these values, the first three moments of the service time distribution are 3.1, 18.1 and 138.1, and the SCV is 0.88345. Figure 1 show the quality of the approximation for $\rho = 0.8$.
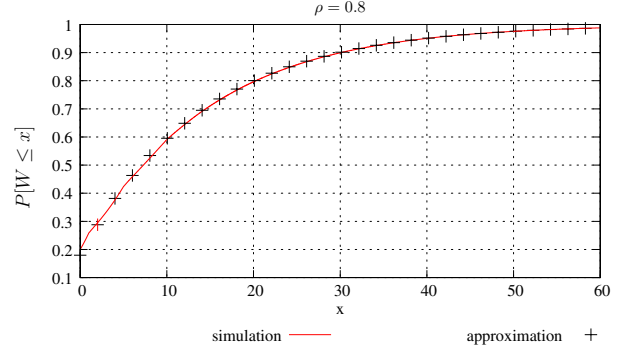

Figure 1. Exponential approximation of $P(W \leq x)$ for $\rho = 0.8$.

4 Single node case

In this section, we consider a single queue and, to simplify notations, we drop the index k of the queue. If the queue is not heavily loaded, the packet C_1 enters the queue when C_0 has already departed, i.e. $W_0 \leq T'$ where $T' = T - D/\mu$. In this case, W_1 and W_0 can be considered as two i.i.d random variables whose distribution is the distribution of the steady-state waiting time of the queue. According to eq. 2, the pdf of the waiting time is,

$$f_W(x) = (1 - \alpha)\, u_0(x) \,+\, \alpha\,\eta\, e^{-\eta x} \tag{3}$$

where η and α depend on the first three moments of the arrival and service time processes of the background traffic. Using this approximation, it can be shown that the pdf of Δ can be approximated as follows,

$$f_\Delta(x) \approx h(T') \begin{cases} 0 & , x \leq -T' \\ \frac{\phi\eta}{2}\, e^{\eta x} - \frac{\alpha^2\eta}{2}\, e^{-\eta(2T'+x)} & , -T' \leq x < 0 \\ 1 - \phi & , x = 0 \\ \left[\frac{\phi\eta}{2} - \frac{\alpha^2\eta}{2}\, e^{-2\eta T'}\right] e^{-\eta x} & , x > 0 \end{cases} \tag{4}$$

where $\phi = 2\,\alpha\,(1 - \alpha/2)$ and $h(T') = 1/[1 - \alpha\, e^{-\eta T'}]$. By definition, the jitter of the tagged packet stream is,

$$
\begin{aligned}
J &= E\left[\,|\,W_1 - W_0\,|\,\right] \\
&= -\int_{-T'}^{0} x\, f_\Delta(x)\, dx \,+\, \int_{0}^{\infty} x\, f_\Delta(x)\, dx
\end{aligned}
$$

After some algebra, it yields,

$$J = \frac{\alpha}{\eta} \, \frac{(2-\alpha) \, - \, e^{-\eta T'} \left[\eta T' + 1 - \alpha + \alpha \, e^{-\eta T'} \right]}{1 - \alpha \, e^{-\eta T'}} \quad (5)$$

where $T' = T - D/\mu$.

4.1 Results

4.1.1 Example 1

Let us consider again the M/D3/1 queue described in section 3. Assume a periodic packet stream, with constant interarrival time T and constant packet size $D = 1$ data unit, is multiplexed with the packets of the Poisson background traffic. In the figures below, we present comparisons of the jitter given by formula 5 and the empirical jitter given by event-driven simulations. We assume that $\mu = 1$ data unit/s and that the utilization rate of the queue is kept constant by increasing $\lambda_T = 1/T$ and decreasing accordingly the intensity $\lambda^{bg} = \rho - \lambda_T$ of the background traffic.

Figures 2-3 show that the jitter approximation is very accurate for light or moderate utilization rates. The approximation is rather crude when the queue is saturated, as illustrated by the case $\rho = 0.8$.
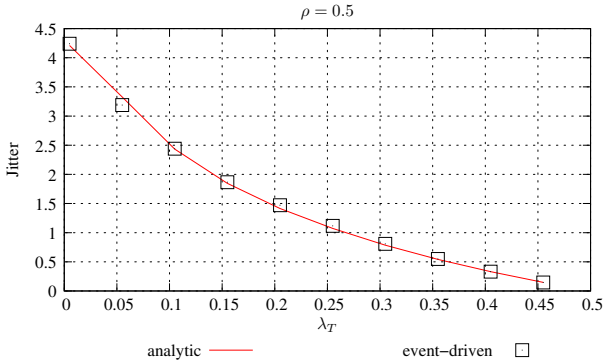

Figure 3. Single M+D/D3+D/1 queue: $\rho = 0.8$.
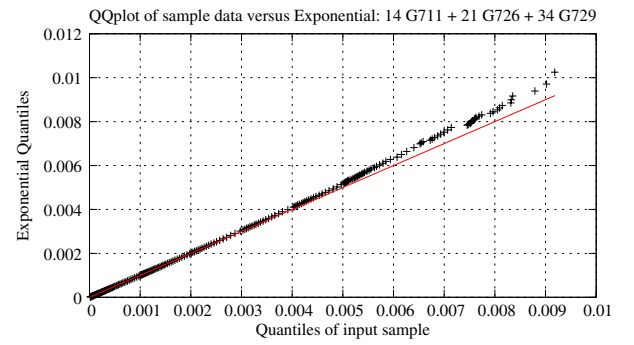
Figure 4. QQ plot for a mixture of 14 G711, 21 G726 and 34 G729.

Figure 2. Single M+D/D3+D/1 queue: $\rho = 0.5$.

4.1.2 Example 2

We assume that the background traffic is the aggregation of several digitized voice traffics, generated using two different standard speech coders : G.726 and G.729. The standard practice is to model voice packet streams as ON-OFF processes [23], with parameters depending on the codec.

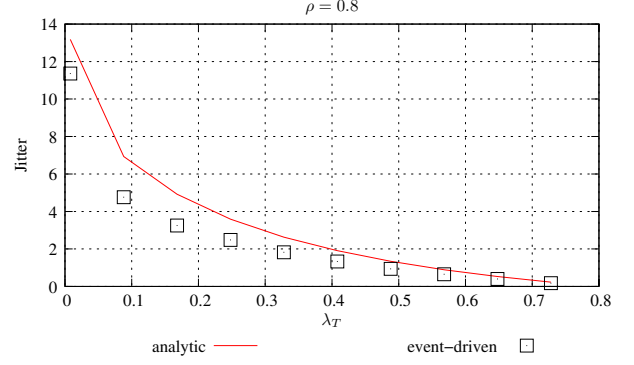In the following, we consider two different mixtures of voice traffics. The first mixture assumes 10% of G.726 voice flows and 90% of G.729 voice flows. The second mixture assumes 50% of each type.

We assume that the background traffic can be modeled as a Poisson packet stream. This assumption is a rather crude approximation when the number of voice flows is low, but it becomes more and more accurate as the number of voice flows is increased. This is illustrated by the QQ plot in figure 4, where the sample data are the interarrivals of a superposition of voice traffics.

The packet size distribution is a two-point distribution. The parameters of this distribution are given by the packet size of each codec and by the probability for an incoming packet to be of type G.726 or G.729.

The packets of this aggregation of voice traffics are multiplexed with a single G.711 voice traffic on a link. The capacity of the link is $\mu = 1920$ kbps.

For several packet-rates of the background traffic, the jitter of the single G.711 voice traffic has been computed using approximation 5 (J_{ana}), and compared to the empirical jitter (J_{eve}) obtained by event-driven simulations. Results are synthetized in figure 5. For the first (resp. second) mixture,

the relative error is below 1.7% (resp. 2.8%) if the utilization rate is less than 54%. It is 6.5% (resp. 13.5%) if the utilization rate is 73.6%.
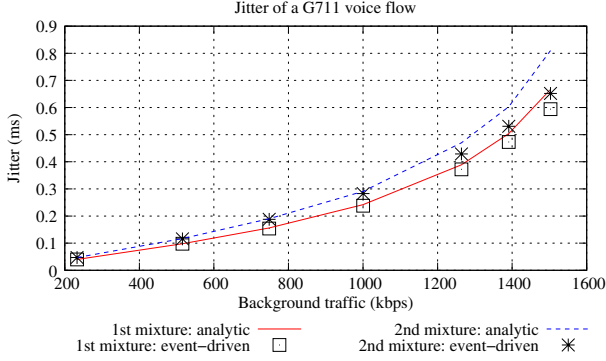

Figure 5. Jitter of a G.711 voice flow multiplexed with a mixture of voice traffics .

4.2 Approximation of the probability density function

Recall that $T' = T - D/\mu$. In the sequel, we make the assumption that the terms weighted by $e^{-\eta T'}$ can be neglected. This assumption will hold if the offered load of the tagged traffic is low with respect to the service rate and if the queue is not saturated. If it holds, eq. 4 becomes,

$$f_\Delta(x) \approx \frac{\phi\eta}{2} e^{-\eta|x|} \left[1 - u_0(x) \right] + (1-\phi)u_0(x) \, , \, x > -T'$$
$$(6)$$

and $f_\Delta(x) = 0$ if $x \le -T'$. This approximation will be used in the next section for the analysis of the multiple node case.

5 Multiple node case

In this section, we consider the multiple-node case. Knowing the arrival distribution and the packet size distribution of the background traffic for each node $k = 1, \ldots, n$, the parameters α_k, $\phi_k = 2\alpha_k(1-\alpha_k/2)$ and η_k can be computed for each node.

Let T_k, $k = 1, \ldots, n$ be the time between arrivals of C_0 and C_1 at node k. We obviously have,

$$T_1 = T \quad \text{and} \quad T_k = T + \sum_{j=1}^{k-1} \Delta_j \quad k = 2 \ldots n$$

Let us denote by $f_{\Delta_k|T}(x)$ the pdf of Δ_k conditioned on $T_k = T$ for $k = 1, \ldots, n$. The pdf $f_{\Delta_k|T}(x)$ can be approximated by eq. 6 with $T'_k = T - D/\mu_k$.

In the sequel, we assume $\eta_i \ne \eta_j$, $i, j = 1, \ldots, n$. The case $\eta_i = \eta_j$ can be handled by continuity. Let us define, for $j = 1, \ldots, n$, the coefficients K_j^n by:

$$K_j^n = \prod_{i=1, i \ne j}^{n} \left(1 - \phi_i \frac{\eta_j^2}{\eta_j^2 - \eta_i^2} \right)$$

To compute the end-to-end jitter, we will derive an analytical approximation for the pdf $f_{\Delta_1 + \ldots + \Delta_n}(x)$ of the sum $\Delta_1 + \ldots + \Delta_n$. The key observation is that the following recursion holds,

$$f_{\Delta_1 + \ldots + \Delta_{n+1}}(x) = \int_{-\infty}^{+\infty} f_{\Delta_1 + \ldots + \Delta_n}(y) \times$$
$$f_{\Delta_{n+1}|T + \Delta_1 + \ldots + \Delta_n}(x - y|T + y) \, dy$$

Using this recusion and approximation 6, the following proposition can be proved.

Proposition 1 *The pdf $f_{\Delta_1 + \ldots + \Delta_n}(x)$ is given by,*

$$f_{\Delta_1 + \ldots + \Delta_n}(x) \approx \begin{cases} 1 - \sum_{j=1}^{n} \phi_j K_j^n & , \, x = 0 \\ \sum_{j=1}^{n} K_j^n f_{\Delta_j | T}(x) & , \, x \ne 0 \end{cases}$$

Proposition 1 allows to derive a very simple expression of the end-to-end jitter, as stated below.

Proposition 2 *The jitter introduced by nodes $1, \ldots, n$ can be approximated by,*

$$J_{[1 \ldots n]}(T) = E\left[\, | \sum_{k=1}^{n} \Delta_k \, | \, \right] \approx \sum_{j=1}^{n} K_j^n J_j(T) \qquad (7)$$

where $J_j(T)$ is the jitter that would be introduced by node j if the inter-arrival time at that node was T,

$$J_j(T) = \frac{\phi_j}{\eta_j} \left[1 - \frac{1}{2}(1 + \eta_j T'_j) e^{-\eta_j T'_j} \right]$$

with $T'_j = T - D/\mu_j$.

5.1 Results

5.2 Example 1

Let us consider $n = 5$ nodes in tandem and a G.711 voice flow sent from node 1 towards node 5. The first and last links have the same bandwidth, $\mu_1 = \mu_4 = 1920$ kbps.

The bandwidth of the two other links is $\mu_2 = \mu_3$=7680 kbps.

We assume that the background traffic on the first and last link is the aggregation of 10% of G.726 voice flows and 90% of G.729 voice flows. The traffic on the two other links is the aggregation of 50% of each type. Moreover, since links 2 and 3 have a much higher bandwidth, it is assumed that the rate of their background traffic is approximately 3 times the rate of the background traffic on links 1 and 4. As in section 4.1.2, the analytic model approximates the aggregation of voice flows by a Poisson arrival process and a two-point service time distribution.

Figure 6 plots both the jitter given by the analytic model and the empirical jitter given by event-driven simulations according to the rate of the background traffic on the first (and last) link. It can be seen that the analytic approximation is fairly accurate. The error is below 1% when the utilization rate of the first queue is less than 46.2%. It is only 2.08% if this utilization rate is 53.4%, and 3.46% if the utilization rate is 59.9%.
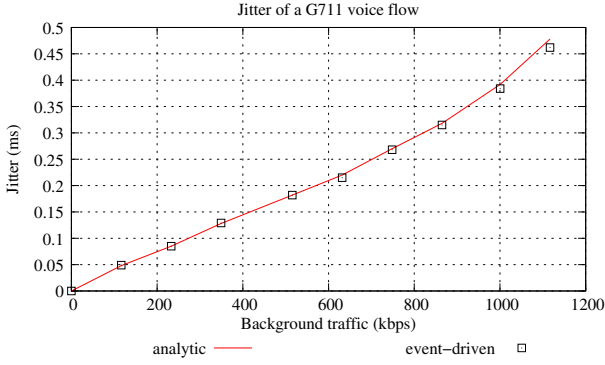

Figure 6. Jitter approximation for a G.711 voice flow passing through 4 queues.

5.3 Example 2

Let us consider $n = 5$ queues in tandem. All queues have a unit service rate, i.e. $\mu_k = 1$ data unit/s, $k = 1 \ldots n$. We assume a two-phase hyperexponential arrival distribution H2 with balanced means. As for the SCV, we use c_a^2=2.0. The packet sizes of the background traffic have a Gamma distribution with mean 1 and shape parameter 0.5. The first three moments of the packet size distribution are 1.0, 3.0 and 15.0, and its SCV is c_s^2=2.0.

The constant packet-size D of the tagged traffic is 1 data unit. All queues have the same utilization rate $\rho = 0.4$. Figure 7 compares the jitter obtained with formula 7 and the jitter obtained by event-driven simulations. The approximation is satisfactory.
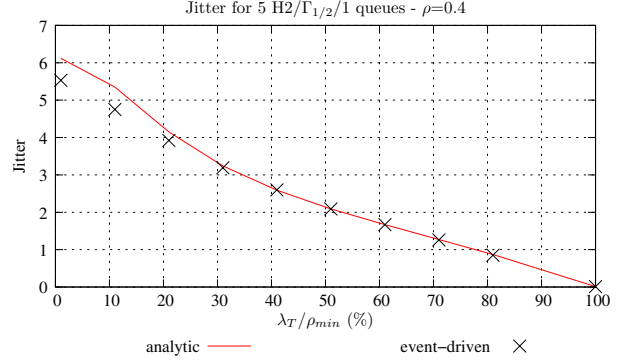

Figure 7. Jitter approximation for 5 $H_2/\Gamma_{1/2}/1$ queues in tandem.

It can be argued that the $\Gamma_{1/2}$ packet size density is reasonably close to an exponential packet size density. To make a more challenging comparison, we change the packet size distribution to a two-point distribution D2 and keep the H2 arrival process unchanged. The SCV of the packet size is still 2.0, but the first three moments are now 1.0, 3.0 and 26.6. Figure 8 shows that the approximation is still satisfactory.
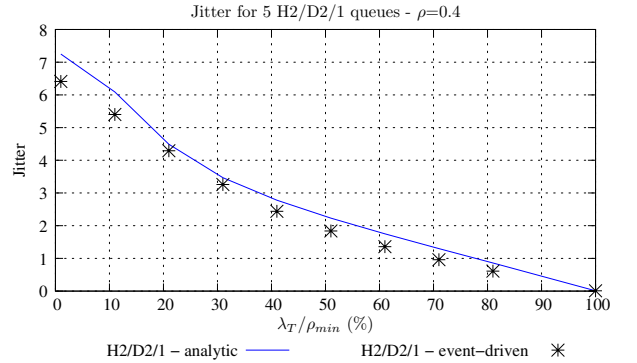

Figure 8. Jitter approximation for 5 H2/D2/1 queues in tandem.

6 Conclusion

Jitter is a crucial performance indicator for real-time audio and video traffics in the Internet. We have derived an analytic approximation for the evaluation of end-to-end jitter incurred by a periodic packet stream with a constant packet size. Comparisons with event-driven simulations have shown that the proposed approximation is fairly accurate. With respect to the numerical methods proposed by prior works, our approximation has the advantage of being

very simple and thus very well suited to the simulation and optimization of large networks. Our approximation is also fairly general since the only assumption regarding the background traffic is that arrivals and service-times are governed by renewal processes.

The main limit of our approximation lies in the independance assumption regarding the waiting times of two consecutive packets. Our results show that this assumption holds in light to moderate traffic conditions, i.e. for utilization rates lower than 70% or 80%. Our future research will focus on extending this approximation to heavy traffic conditions.

References

[1] Abate J., Choudhury G.L., Whitt W. *Exponential approximations for tail probabilities in queues, I: waiting times* . Operations Research, vol 43, No 5, 1995.

[2] Alshaer, H., Horlait, E. *Expedited Forwarding End to End Delay Jitter in the Differentiated Services Networks*, Proceedings of HSNMC 2004, 2004.

[3] Boyer, P. E., Guillemin, F. M., Servel, M. J., Coudreuse, J.-P. *Spacing Cells Protects and Enhances Utilization of ATM Network Links* , IEEE Network, vol. 6, 1992.

[4] Cruz, R., *A calculus for network delay, Part I: network elements in isolation*, IEEE Trans. on Inf. Theory, vol. 37, 1991.

[5] Cruz, R., *A calculus for network delay, Part II: network analysis*, IEEE Trans. on Inf. Theory, vol. 37, 1991.

[6] Parekh, A., Gallager, R., *A generalized processor sharing approach to flow control in integrated services networks: the single node case*, IEEE/ACM Trans. on Networking, vol. 1, 1993.

[7] Parekh, A., Gallager, R., *A generalized processor sharing approach to flow control in integrated services networks: the multiple node case*, IEEE/ACM Trans. on Networking, vol. 1, 1993.

[8] Kurose, J., *On computing per-session performance bounds in high-speed multi-hop computer networks*, Proceedings ACM SIGMETRICS'92, 1992.

[9] Yaron, O., Sidi, M., *Calculating performance bounds in communication networks*, Proceedings IEEE INFOCOM'93, San Francisco, 1993.

[10] Cidon, I., Khamisy, A., Sidi, M., *Dispersed messages in discrete-time queues: delay, jitter and threshold crossing*, IEEE INFOCOM, Toronto, Canada, 1994.

[11] Ferrari, D., *Delay jitter control scheme for packet-switching internetworks*, Computer Communications, 15(6):367-373, 1992.

[12] Guillemin, F., Boyer, P., Dupuis, A, Romoeuf, L., *Peak rate enforcement in ATM networks*, IEEE INFOCOM, Florence, Italy, 1992.

[13] Lam, S., Xie, G., *Burst scheduling: architecture and algorithm for switching packet video*, IEEE INFOCOM, Boston, Massachussets, 1995.

[14] Landry, R., Stavrakakis, I., *Traffic shaping of a tagged stream in an ATM network: approximate end-to-end analysis*. IEEE INFOCOM, Boston, Massachussets, 1995.

[15] Landry, R., Stavrakakis, I., *Study of delay jitter with and without peak rate enforcement*. IEEE/ACM Transactions on Networking, vol. 5:4, 1997.

[16] Li, S. Q., Fulton, C., *Delay jitter first-order statistics of general traffic on high-speed networks*. IEEE ICC, Dallas, Texas, 1996.

[17] Matragi, W., Bisdikian, C., Sohraby, K., *Jitter calculus in ATM networks: single node case*, Proceedings IEEE Infocom, Toronto, Canada, 1994.

[18] Matragi, W., Sohraby, K., Bisdikian, C., *A Framework for Jitter Analysis in Cell Based Multiplexors*, Performance Evaluation, vol. 22, 1996.

[19] Matragi, W., Sohraby, K., Bisdikian, C., *Jitter calculus in ATM networks: multiple node case*. IEEE/ACM Transactions on Networking, vol. 5, 1997.

[20] Privalov, A., Sohraby, K., *Per-stream jitter analysis in CBR ATM multiplexors*, IEEE/ACM Transactions on Networking, vol. 6:2, 1998.

[21] Roberts, J. W., *Jitter due to an ATM multiplex - application to peak rate policing*, Technical Report COST 224-35, RACE Cost Project, 1989.

[22] Roberts, J. W., Guillemin, F., *Jitter in ATM networks and its impact on peak rate enforcement*, Performance Evaluation, 16:35-48, 1992.

[23] Sriram, K., Whitt, W., *Characterizing Superposition Arrival Processes in Packet Multiplexers for Voice and Data*, IEEE Journal of Selected Areas in Communications, pp. 833-846. 16:35-48, Sept. 1986.