A Fast Convolution Approximation Scheme for Estimating End-to-End Delay Performance *

Seung Y. Nam and Dan K. Sung

A simple convolution approximation scheme is proposed to reduce the number of operations and the required amount of data. This approximation method can be used to estimate the end-to-end cell delay variation (CDV) using local information about cell delays in Asynchronous Transfer Mode (ATM) networks.

Introduction: Direct convolution calculation of two N-point sequences requires a number of arithmetic operations of the order of N^2. For a large convolution, the corresponding processing load becomes rapidly excessive and, thus, considerable effort has been devoted to devising faster computational methods. Conventional approaches for speeding up convolution calculations are based on FFT, which requires a number of operations of the order of $N \log_2 N$, where N is the power of two [1].

This Letter introduces a simple convolution approximation method. The number of operations can be reduced at the expense of approximation errors. Another advantage of this approximation method is that it is possible to obtain an approximation result with a small amount of required information. This feature can be useful in estimating end-to-end CDV in ATM networks by successive convolutions of the delay distribution of each node [2]. Due to signaling constraints [3][4] the proposed approximation mechanism is preferable to conventional fast convolution methods.

*This work was supported in part by the Basic Research Program of the Korea Science and Engineering Foundation

Approximation Method: The proposed convolution approximation method uses the following property:

Property 1. Let X_1 and X_2 be discrete random variables that take integer values and satisfy the following condition for an integer k:

$$Pr(X_1 \geq k) = \sum_{i=k}^{\infty} Pr(X_1 = i) \leq \sum_{i=k}^{\infty} Pr(X_2 = i) = Pr(X_2 \geq k) \tag{1}$$

Letting $Z_1 = X_1 + Y$, $Z_2 = X_2 + Y$ for a nonnegative discrete random variable Y independent of both X_1 and X_2, the following relation holds:

$$Pr(Z_1 \geq k') \leq Pr(Z_2 \geq k'), \qquad \forall k'$$

For a random variable X_1 with a probability mass function (PMF) of

$$P_{X_1}(i) = Pr(X_1 = i) = \begin{cases} a_i, & \text{if } i \geq 0, \\ 0, & \text{otherwise,} \end{cases}$$

if we introduce X_2 with the following PMF using data compression factor Δ, which is a positive integer:

$$P_{X_2}(i) = Pr(X_2 = i) = \begin{cases} \sum_{h=i-\Delta+1}^{i} a_h, & \text{if } i = j\Delta - 1,\ j = 1, 2, \cdots, \\ 0, & \text{otherwise,} \end{cases} \tag{2}$$

then the following relation holds for a nonnegative integer k:

$$Pr(X_1 \geq k) = \sum_{i=k}^{\infty} a_i \leq \sum_{i=\lfloor k/\Delta \rfloor \Delta}^{\infty} a_i = Pr(X_2 \geq k) \tag{3}$$

X_1 and X_2 satisfy the condition of the Property 1. Thus, convolution of the PMFs of X_2 and a discrete random variable Y_1 yields the upper bound of the complementary cumulative distribution function (CDF) that can be obtained from convolution of the PMFs of X_1 and Y_1. Since the random variable X_2 retains compressed information about X_1, it is possible to reduce the amount of transmitted data by use of X_2 instead of X_1.

Information about random variable Y_1 can also be compressed into Y_2 by the same mechanism, as follows:

$$P_{Y_2}(i) = \begin{cases} \sum_{h=i-\Delta'+1}^{i} P_{Y_1}(h), & \text{if } i = j\Delta' - 1,\ j = 1, 2, \cdots, \\ 0, & \text{otherwise}, \end{cases} \tag{4}$$

where Δ' is a positive integer.

The convolution of PMFs of X_2 and Y_2 can be calculated as follows when Δ is equal to Δ':

$$P_{X_2} * P_{Y_2}(k) = \begin{cases} \sum_{i=0}^{k} P_{X_2}(i) P_{Y_2}(k-i), & \text{if } k = n\Delta - 2,\ n = 2, 3, 4, \cdots, \\ 0, & \text{otherwise}, \end{cases} \tag{5}$$

$$\begin{aligned} P_{X_2} * P_{Y_2}(n\Delta - 2) &= \sum_{i=0}^{n\Delta-2} P_{X_2}(i) P_{Y_2}(n\Delta - 2 - i) \\ &= \sum_{i=1}^{n-1} P_{X_2}(i\Delta - 1) P_{Y_2}((n-i)\Delta - 1) = \sum_{j=0}^{n-2} P_{X_2^*}(j) P_{Y_2^*}(n-2-j) \\ &= P_{X_2^*} * P_{Y_2^*}(n-2), \end{aligned} \tag{6}$$

where $P_{X_2^*}(i) = P_{X_2}((i+1)\Delta - 1)$ and $P_{Y_2^*}(i) = P_{Y_2}((i+1)\Delta - 1)$ for $i = 0, 1, 2, \cdots$.

This result indicates that convolution of the PMFs of X_2 and Y_2 can be obtained by convolution of $P_{X_2^*}(i)$ and $P_{Y_2^*}(i)$, followed by rescaling. Let the maximum values of X_1 and Y_1 be D_X and D_Y, respectively. The number of probability values of $P_{X_2^*}(i)$ is approximately $1/\Delta$ times less than for $P_{X_1}(i)$. Therefore, convolution of the PMFs of X_2 and Y_2 instead of for X_1 and Y_1 can reduce the number of multiplications from $(D_X + 1) \times (D_Y + 1)$ to $(\lfloor D_X/\Delta \rfloor + 1) \times (\lfloor D_Y/\Delta \rfloor + 1)$.

When Δ is different from Δ', Δ^* denotes the greatest common divisor (GCD) of Δ and Δ'. Then, the number of multiplications can be reduced from $(D_X + 1) \times (D_Y + 1)$ to $(\lfloor D_X/\Delta^* \rfloor + 1) \times (\lfloor D_Y/\Delta^* \rfloor + 1)$ by the proposed mechanism. Increasing Δ^* can reduce the number of operations at any level at the expense of approximation errors.

Now we consider the upper bound of approximation errors. Let X_1 and Y be nonnegative integer random variables. X_2 is obtained by Eqn. (2) from X_1. When we use X_2 instead of X_1 in the convolution with Y, the approximation error is expressed as

$$\frac{|Pr(X_2+Y\geq z)-Pr(X_1+Y\geq z)|}{Pr(X_1+Y\geq z)}\leq\frac{\sum_{i=0}^{z}\Big\{\sum_{j=\max(0,z-i-\Delta+1)}^{z-i-1}Pr(X_1=j)\Big\}Pr(Y=i)}{\sum_{i=0}^{z}Pr(X_1\geq z-i)Pr(Y=i)+\sum_{i=z+1}^{\infty}Pr(Y=i)} \tag{7}$$

As an example, let X_1 and Y be two waiting times in two consecutive M/M/1 systems. The waiting time distribution in the M/M/1 system is an exponential distribution [5]. Since a discrete version of an exponential distribution is a geometric distribution, the two geometric random variables X_1 and Y are introduced with parameters p and q, respectively.

$$Pr(X_1=i)=(1-p)^i p,\quad Pr(Y=i)=(1-q)^i q,\qquad i=0,1,2,\cdots. \tag{8}$$

In this case the upper bound of the convolution approximation error can be obtained by Eqns. (7) and (8), as follows:

$$\begin{aligned}&\frac{|Pr(X_2+Y\geq z)-Pr(X_1+Y\geq z)|}{Pr(X_1+Y\geq z)}\\ &\qquad\leq\begin{cases}\Big(\dfrac{1}{1-p}\Big)^{\Delta-1}\dfrac{p((1-q)/(1-p))^{z-\Delta+2}-q}{p((1-q)/(1-p))^{z+1}-q}-1, & \text{if } p\neq q,\\ \Big(\dfrac{1}{1-p}\Big)^{\Delta-1}\Big\{1-\dfrac{p(\Delta-1)}{1+pz}\Big\}-1, & \text{if } p=q.\end{cases}\end{aligned} \tag{9}$$

Regardless of the equality of p and q, the approximation error has the following looser upper bound.

$$\frac{|Pr(X_2+Y\geq z)-Pr(X_1+Y\geq z)|}{Pr(X_1+Y\geq z)}\leq\Big(\frac{1}{1-\min(p,q)}\Big)^{\Delta-1}-1. \tag{10}$$

This result indicates that the relative error is upper-bounded. The absolute error bound decreases if $Pr(X_1+Y\geq z)$ decreases as z increases. This error bound increases as Δ increases. Thus, there is a trade-off that the approximation error may increase while the required amount of data decreases as Δ increases.

Application and Simulation Result: The proposed fast convolution approximation method is applied to a serially connected three buffer model, as shown in Fig. 1. X_i $(i = 1, 2, 3)$ is the cell delay experienced in the i-th node and X_e is the end-to-end cell delay. X_e is equal to $\sum_{i=1}^{3} X_i$. Consequently, the distribution of X_e can be obtained by successive convolutions of the distributions of X_1, X_2, and X_3 if delays experienced in consecutive nodes are uncorrelated, or nearly so.

Constant bit rate (CBR) traffic and variable bit rate (VBR) traffic for foreground and background traffic are considered. VBR traffic is modeled as a two-state Markov process that consists of an active state and a silent state. Fig. 2 compares complementary CDFs obtained from three different methods. The first method is to obtain the delay distribution by measuring the delay of each cell through time stamps. The second method is to take successive convolutions of the delay distributions of the three nodes. The last method is to use the fast convolution approximation mechanism. The data compression factors of the three nodes are all Δ's. The amount of data required for fast convolution approximation is $1/\Delta$ times less than for conventional convolution mechanisms. The fast convolution approximation method yields a close estimate of the delay distribution when Δ is small, as shown in Fig. 2. Hence, the fast convolution approximation mechanism makes it possible to estimate the delay distribution with a reduced amount of data.

Conclusions: A simple convolution approximation method is proposed. Unlike conventional fast convolution mechanisms this method can reduce the number of operations at any low level at the expense of approximation errors. The proposed method closely estimates convolution results with a small amount of required data. This is a useful feature in calculating the end-to-end α-quantile CTD from the local delay information of each node in ATM networks.

References

1 Nussbaumer, H.J.: 'Fast Fourier Transform and Convolution Algorithms' (Springer-Verlag, 1982), 2nd edn.

2 Vleeschauwer, D.: 'Experimental Verification of the Statistical Independence of Cell Delays Introduced in Consecutive Switches,' B-ISDN Teletraffic Modeling Symposium, Feb. 1995, pp. 105-116

3 ATM Forum: 'User Network Interface Version 4.0', 1996.

4 ATM Forum: 'B-ISDN Inter-Carrier Interface (BICI) Specification Version 1.1', Sept. 1996.

5 Wolff, R.W.: 'Stochastic Modeling and The Theory of Queues' (Prentice-Hall, 1989)

Authors' affiliations:

Seung Y. Nam and Dan K. Sung (Department of Electrical Engineering and Computer Science, Korea Advanced Institute of Science and Technology, 373-1 Kusong-dong, Yusong-gu, Taejon 305-701, Korea, E-mail: dksung@ee.kaist.ac.kr)

Figure captions:

Fig. 1. A serially connected three buffer model.

Fig. 2. Comparison of the time stamp, convolution, and fast convolution approximation methods.

time stamp
convolution
fast conv. approx. ($\Delta = 2$)
fast conv. approx. ($\Delta = 4$)
fast conv. approx. ($\Delta = 8$)
fast conv. approx. ($\Delta = 16$)

Figure 1

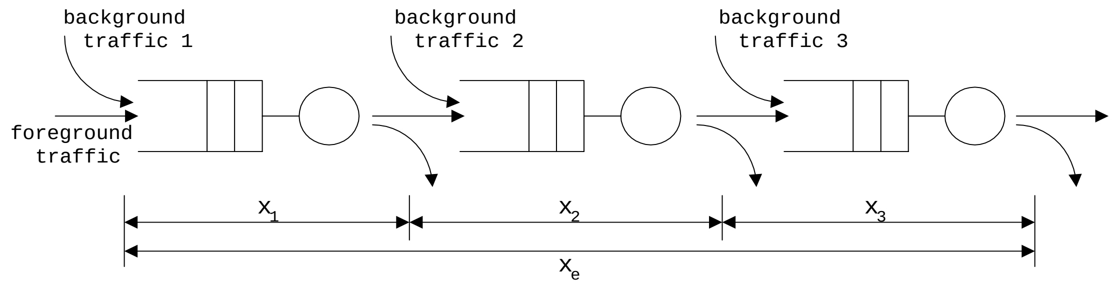

Figure 2