14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP

ESTIMATION OF PARAMETERS OF INPUT TRAFFIC STREAMS FROM STATISTICALLY MULTIPLEXED OUTPUT [1]

Rajesh Narasimha, Raghuveer M Rao#, Sohail Dianat#*

*Georgia Institute of Technology, Atlanta GA 30332, #Rochester Institute of Technology, Rochester, NY 14623

rajesh@ece.gatech.edu, {mrreee, sadeee@rit.edu}

ABSTRACT

Given the statistically multiplexed stream observations of two independent and different types of traffic streams, this paper examines the problem of determining the degree of mixing. In data networks a common example of such a pair of different stream would be one conforming to the traditional Poisson model with an exponential inter-arrival distribution and the other obeying long-range dependent traffic characterized by a heavy-tailed distribution. The paper provides an expression for the probability density function of the inter-arrival time of the mixed stream in terms of those of the input streams for the general case. As an example, we consider a mixed output traffic stream for the specific case of multiplexed Poisson and heavy-tailed processes and an approach is provided to estimate input parameters from the first and second order statistics. For arrival rate estimation of the input streams, we propose a look–up table approach based on nearest neighbor search. The simulation results demonstrate that the estimated arrival rates and the moments are indeed close to their respective true values.

Keywords: Statistical Multiplexing, parameter estimation, heavy-tailed.

1. INTRODUCTION

In data networks, statistical multiplexers have extensively been a part of packet switches and routers. In the statistical multiplexing scheme, various traffic streams that arrive on links of fixed or varying bandwidth are served on a first-come first-serve (FCFS) basis, where a finite buffer space is allocated for queuing purposes. Under heavy traffic volume, packet loss/delay occurs at the outgoing link and hence statistical multiplexers are intended to decrease packet loss/delay using better queuing strategies thereby increasing link utilization. Estimation of the arrival rates of the input traffic streams from the knowledge of the multiplexed stream can facilitate in improving the link utilization. Studies in traffic engineering have demonstrated extensively over the last decade that various instances of packet traffic such as Ethernet LAN traffic [1], aggregate packet streams [2] and Variable Bit-Rate (VBR) video [3, 4] exhibit statistical self-similarity and long-range dependence, typically attributed to the aggregation of heavy-tailed ON-OFF processes. Recently, it has also been shown that superposition of many voice streams possess burstiness which leads to packet delays under heavy loads [5]. On the other hand, there are instances of traffic that fit conventional Poisson queuing models. Packets of different traffic streams are often mixed onto a single outgoing link through statistical multiplexing or superposition [6]. Thus, it is conceivable that there are instances of mixing of various types of traffic streams through this process.

This paper examines the problem of determining the degree of mixing of two independent streams of traffic from observations of their statistically multiplexed stream. The independent streams are assumed to arise from different statistical models. Initially, the paper provides a general expression for the PDF of the inter-arrival time mixed stream in terms of the PDF of the input inter-arrival times. An example of a pair of different stream types would be one conforming to the conventional Poisson model and the other obeying long-range dependence (LRD). The latter type of traffic is typically characterized by heavy tailed distributions such as a Pareto distribution. Both types of traffic are known to occur typically in data networks. The Poisson model is prevalent in telephone-type traffic while LRD has been found to be widespread in various instances of data traffic such as Ethernets, variable bit rate video and other types of compressed traffic streams.

The situation considered is depicted in Fig. 1. If l is the rate of packet arrival for stream 1and m is that for stream 2, then the fractional contribution of stream 1 to the mixed stream is $\frac{l}{l+m}$, while that of the second stream is $\frac{m}{l+m}$. The problem addressed amounts to determining these fractions from the statistics of the output stream. It is assumed that the form of the PDFs of the two traffic streams are know but their parameters are not. Thus the problem can alternatively be seen as estimating these parameters from the mixed stream.

Various possibilities arise as to how one might estimate the statistics of the mixed stream. The most obvious route is to estimate the PDF of the mixed traffic and then fit the PDF model to the estimate. This clearly means estimating l and m according to some criterion such as least squares, where l and m are the parameters of interest of the input streams respectively. Another possibility is to estimate traffic statistics and determine the values of l and m that best generate these statistics according to the model. Since two parameters are involved, we need at least two statistics, say the mean and the variance to obtain the estimates. We take the latter course in this work.

[1] This work was supported in part by the IISc-DRDO program on Mathematical Engineering. A short version of the paper was presented at CISS 2006.

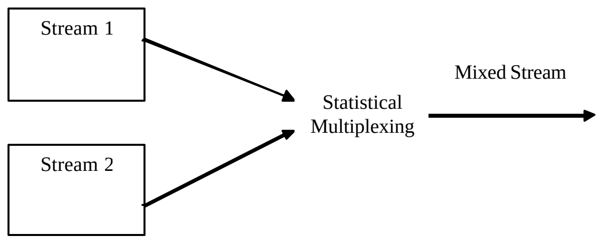

Fig. 1. Statistical Multiplexing of independent streams.

2. DISTRIBUTION OF MULTIPLEXED STREAM

In this section we derive the expression for the probability density function (PDF) of the time interval between any random instant of non-arrival to the next arrival of a traffic stream. We then provide an expression for the PDF of the inter-arrival time of the mixed stream in terms of the PDFs of the input the inter-arrival times.

Suppose X denotes the inter-arrival time for a random arrival process *A*. Let $f_X(x)$ be the PDF of *X*. Let $f_{X'}(x)$ be the PDF of the random variable X' denoting the time interval between an arbitrary instant of no arrival of the process A and its next arrival as shown in Fig. 2 .

Lemma 1: With $F_X(x)$ denoting the cumulative distribution function (CDF) of *X* and $E\{X\}$ its expected value, the PDF of the random variable X' is given by

$$f_{X'}(x) = \frac{(1-F_X(x))}{E\{X\}} \; . \tag{1}$$

Proof: Let l be the average arrival rate for the process *A*. Then

$$P(\text{arrivalinelementdxatx}) = l\,dx \tag{2}$$

Therefore

$$P\left(x \le X' \le x+dx\right) = f_{X'}(x)dx \tag{3}$$

~~For any arrival process~~, one should note that the time reversal does not change the distribution of the inter-arrival time. In other words, if *B* is an arrival process such that for every time instant *t* for which there is an arrival for process *A*, there is an arrival at –*t* for *B* and vice versa, then the inter-arrival time for B also has PDF $f_X(x)$. Thus

$$\begin{aligned} P\left(x \le X' \le x+dx\right) &= P\begin{pmatrix}\text{onearrivalofprocess } A \text{ininterval} \\ \text{of } dx \text{secandnoarrivalinnext } x \text{sec}\end{pmatrix} \\ &= \left(1-F_X(x)\right) l\,dx \end{aligned} \tag{4}$$

Also, the arrival rate for stream 1, l is related to the expected value of inter-arrival time as

$$l = \frac{1}{E\{X\}} \tag{5}$$

Hence from (3) and (4) we have

$$f_{X'}(x)dx = \left(1-F_X(x)\right) l\,dx \tag{6}$$

Equation (1) is obtained by substituting (5) in (6).

That the right-hand-side of (1) integrates to a value of one is assured by the fact that for a random variable taking only positive values, the condition $\lim_{x\to\infty} x^2 f_X(x) \to 0$ holds, and thus $E\{X\} = \int_0^\infty (1-F_X(x))dx$ [7]. It can also be verified from (1) that for Poisson arrival with exponential inter-arrival distribution $f_X(x) = l e^{-lx} u(x)$, where $u(x)$ is the unit step function, $f_{X'}(x) = f_X(x)$ as is to be expected for a memoryless process.

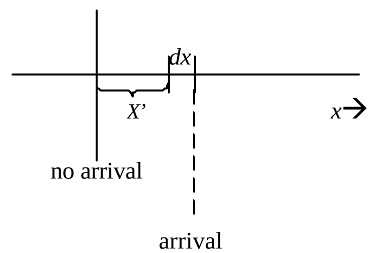

Fig. 2. Time interval between arbitrary instant and next arrival

Corollary to Lemma 1:

$$f_X(x) = -E\{X\} f'_{X'}(x) \tag{7}$$

where $f'_{X'}(x)$ is the derivative of $f_{X'}(x)$.

Proof: Differentiating both sides of Equation (1) we have the result.

Next we derive the PDF of the inter-arrival of the multiplexed stream. Suppose X is the inter-arrival time for stream 1, Y is the inter-arrival time for stream 2 and Z is the inter-arrival time for the multiplexed stream. Also, let X', Y' and Z' be the respective times for these processes between an arbitrary instant of no arrival and the next arrival. Then,

Theorem 1: *The probability density function of the inter-arrival time of the multiplexed stream Z is given by*

$$f_Z(x) = E\{Z\}\left\{\begin{aligned} &\frac{2}{E\{X\}E\{Y\}}(1-F_X(x))(1-F_Y(x)) + \\ &\frac{f_X(x)}{E\{X\}}A(x) + \frac{f_Y(x)}{E\{Y\}}B(x) \end{aligned}\right\} \tag{8}$$

where $A(x) = 1 - \int_0^x \frac{(1-F_Y(u))du}{E\{Y\}}$, $B(x) = 1 - \int_0^x \frac{(1-F_X(u))du}{E\{X\}}$,

$E\{X\} = \frac{1}{l}$, $E\{Y\} = \frac{1}{m}$ *and* $E\{Z\} = \frac{1}{l+m}$ *.Here* $F_X(x)$ *and* $F_Y(x)$ *are the cumulative distribution functions of X and Y respectively.*

Proof:

$$\begin{aligned} P\left(x \le Z' \le x+dx\right) &= P\left(x \le X' \le x+dx\right)P\left(Y' > x\right) \\ &+ P\left(x \le Y \le x+dx\right)P\left(X' > x\right) \end{aligned} \tag{9}$$

$$\Rightarrow f_{Z'}(x)dx = f_{X'}(x)dx\int_x^{\infty} f_{Y'}(u)du + f_{Y'}(x)dx\int_x^{\infty} f_{X'}(u)du \qquad (10)$$

$$\Rightarrow f_{Z'}(x) = f_{X'}(x)\int_x^{\infty} f_{Y'}(u)du + f_{Y'}(x)\int_x^{\infty} f_{X'}(u)du \qquad (11)$$

Differentiating (11) and substituting it in equation (7) we have,

$$f_Z(x) = -E\{Z\} f_{Z'}^{'}(x)$$

$$= E\{Z\}\left[2 f_{X'}(x) f_{Y'}(x) - f_{X'}^{'}(x)\int_x^{\infty} f_{Y'}(u)du - f_{Y'}^{'}(x)\int_x^{\infty} f_{X'}(u)du\right] \qquad (12)$$

Using (1), we have

$$f_{X'}(x) = \frac{(1-F_X(x))}{E\{X\}},\ f_{Y'}(x) = \frac{(1-F_Y(x))}{E\{Y\}},\ f_{Z'}(x) = \frac{(1-F_Z(x))}{E\{Z\}} \qquad (13)$$

Equation (8) follows by substituting (13) in (12) where $A(x) = 1 - \int_0^x \frac{(1-F_Y(u))du}{E\{Y\}}$, $B(x) = 1 - \int_0^x \frac{(1-F_X(u))du}{E\{X\}}$. In chapter 7 of [8] the author evaluates the PDF of the inter-arrival time of N pooled streams. For two memoryless arrivals with PDF $f_X(x) = l e^{-lx} u(x)$ and $f_Y(x) = m e^{-mx} u(x)$ the output PDF is given as

$$f_Z(x) = E(Z)\left[(l+m)^2 e^{-(l+m)x}\right]. \qquad (14)$$

Integrating both sides of (14) we have $E(Z) = 1/(l+m)$ which confirms the form of (12). For e.g. if all of arrivals have the same exponential inter-arrival distribution $f_X(x) = l e^{-lx} u(x)$, then from (8) we have $f_Z(x) = N l e^{-Nlx} u(x)$ and hence is a special case of (12).

3. ESTIMATION OF THE ARRIVAL RATES

Knowing the form of the PDFs of the input streams, determines the form of the PDF of the multiplexed stream as shown in (12). In principle, estimating the statistics of Z based on its observations can therefore be used to determine the parameters of its PDF, namely, l and m. It is difficult to derive expressions for these parameters for the general case. We therefore develop expressions for one case for illustration, namely, that of statistical multiplexing of Poisson and heavy-tailed traffic.

We will assume that the Poisson traffic to have an inter-arrival time X with an exponential PDF $f_X(x) = l e^{-lx}$, where the average packet arrival rate for this stream is l and $E\{X\} = \frac{1}{l}$. The CDF is given by $F_X(x) = \left(1 - e^{-lx}\right)u(x)$. Suppose the inter-arrival time distribution Y for the heavy-tailed traffic is Pareto-distributed as

$$f_Y(x) = \begin{cases} (k-1)x^{-k} & x \ge 1 \\ 0 & \text{otherwise} \end{cases} \qquad (15)$$

where $k > 3$. The average time interval between packets of the heavy-tailed traffic is then given by $E\{Y\} = \frac{k-1}{k-2}$. The average arrival rate (denoted by m) is given by $m = 1/E\{Y\}$. The CDF is given by $F_Y(x) = \left(1 - x^{1-k}\right)u(x-1)$. Thus the average packet rate of the mixed stream traffic (denoted by n) is given by $n = l + m$.

Proposition 1: *The PDF of the inter-arrival time of the mixed stream consisting of a Poisson traffic with inter-arrival time X that is exponentially distributed,* $f_X(x) = l e^{-lx}$ *and heavy-tailed traffic with the inter-arrival time Y that is Pareto-distributed as in (15), is given as*

$$f_Z(x) = \begin{cases} \dfrac{l e^{-lx}}{(l+m)}(2m + l - lmx) & 0 < x < 1 \\ \dfrac{m x^{-k} e^{-lx}}{(l+m)}\left(k - 1 + 2lx - \dfrac{l^2}{2-k}x^2\right) & x > 1 \end{cases} \qquad (16)$$

Proof: Substituting for *A(x)* and *B(x)* in (8), for $0 < x < 1$, we obtain $A(x) = 1 - mx$ and $B(x) = e^{-lx}$

Substituting these back into equation (8), we have the PDF of Z as

$$f_Z(x) = \left\{\frac{l e^{-lx}}{(l+m)}(2m + l - lmx) \quad 0 < x < 1\right. \qquad (17)$$

For $x > 1$, $A(x) = \frac{-m}{2-k}x^{2-k}$ and $B(x) = e^{-lx}$,

$$f_Z(x) = \left\{\frac{m x^{-k} e^{-lx}}{(l+m)}\left(k - 1 + 2lx - \frac{l^2}{2-k}x^2\right) \quad x > 1.\right. \qquad (18)$$

Hence the result.

The valid range of $0.5 < m < 1$ and $l > 0$. The CDF of Z is given by integrating (15) and (16) over the valid range as follows,

$$F_Z(x) = \begin{cases} \dfrac{lmx e^{-lx}}{(l+m)} + (1 - e^{-lx}) & 0 < x < 1 \\ \dfrac{m}{(l+m)}\left(e^{-l} - x^{1-k}e^{-lx}\right) + \dfrac{l(1-m)}{(l+m)} & \\ \left(e^{-l} - x^{2-k}e^{-lx}\right) & x > 1 \end{cases}. \qquad (19)$$

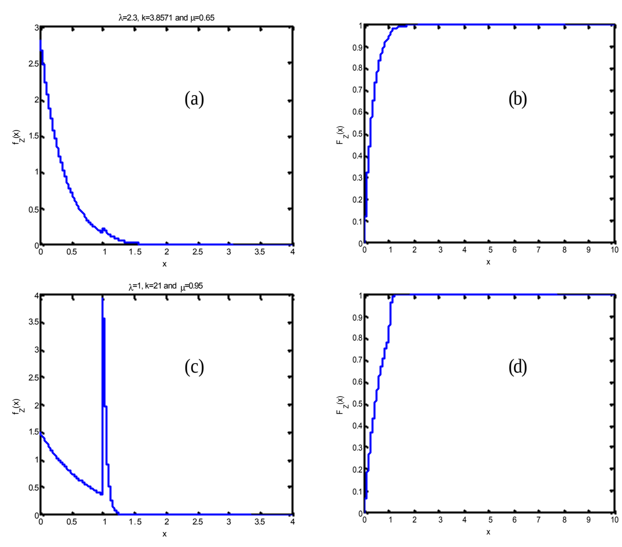

Fig. 3. PDFs of the mixed stream (a), (c) and their respective CDFs (b), (d).

Examples of the PDFs are shown in Fig. 3(a) & (c) for the pairs $l = 2.3, m = 0.65$ and $l = 1, m = 0.95$ respectively. The function $f_Z(x)$ has a discontinuity at *x=1* as seen from Fig. 3(a) & (c). This discontinuity is not obvious when $l \gg m$ as shown in Fig. 3(a). The CDFs of the density functions are depicted in Fig. 3(b) & (d) respectively. Fig. 4(b) & (d) shows the histogram of the mixed stream obtained by statistical multiplexing along with the true PDF for $l = 2.3, m = 0.65$ and $l = 1, m = 0.85$ respectively. The closeness of the histogram to the true PDF provides a strong verification of the accuracy of the derivation leading to the results in (8) and (16). The corresponding multiplexed stream is depicted in Fig. 4(a) & (c)

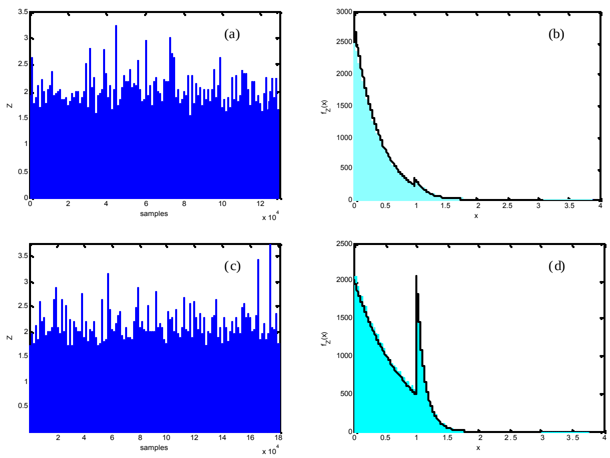

Fig. 4. (a)&(c) Multiplexed stream, (b)&(d) Histogram of multiplexed stream Z superimposed on the output PDF generated using equation (16) for various arrival rates.

Proposition 2: *The n^{th} moment for the mixed stream Z knowing the form of the PDF of the input streams is given as*

$$\begin{aligned} E\{Z^n\} &= a(n,l)\left\{\frac{l}{(l+m)}(l+m-mn)\right\} + \\ &\left(\frac{m}{(1-m)} + 2m(n+1-b) + (1-m)(n+2-b)(n+1-b)\right) \\ &\left\{\frac{\Gamma(n-b+1)}{l^{n-b+1}(l+m)} - \frac{a(n-b,l)}{(l+m)}\right\} + \frac{e^{-l}}{(l+m)} \\ &\{2m + ml + (1-m)(l+n+2-b)\} \end{aligned} \tag{20}$$

where $a(n,l) = \int_0^1 x^n e^{-lx} dx$ and $b = \dfrac{m-2}{m-1}$. The recursive equation for $a(n,l)$ is given as

$$a(n+1,l) = \frac{n+1}{l} a(n,l) - \frac{e^{-l}}{l} \tag{21}$$

with the initial condition

$$a(0,l) = \int_0^1 e^{-lx} dx = \frac{1-e^{-l}}{l}. \tag{22}$$

Proof: Follows directly upon evaluating the integral for $E\{Z^n\}$ using (16)

for $0 < x < 1$

$$\begin{aligned} E\{Z^n\} &= \frac{l(2m+l)}{(l+m)}\int_0^1 x^n e^{-lx} - \frac{ml^2}{(l+m)}\int_0^1 x^{n+1} e^{-lx} \\ &= \frac{l(2m+l)}{(l+m)} a(n,l) - \frac{ml^2}{(l+m)} a(n+1,l) \end{aligned} \tag{23}$$

for $x > 1$

$$\begin{aligned} E\{Z^n\} &= \frac{m}{(1-m)(l+m)}\left[\frac{\Gamma(n-b+1)}{l^{n-b+1}} - a(n-b,l)\right] + \\ &\frac{2ml}{(l+m)}\left[\frac{\Gamma(n-b+2)}{l^{n-b+2}} - a(n+1-b,l)\right] + \\ &\frac{l^2(1-m)}{(l+m)}\left[\frac{\Gamma(n-b+3)}{l^{n-b+3}} - a(n+2-b,l)\right] \end{aligned} \tag{24}$$

Using $\Gamma(n+1) = n\Gamma(n)$ and $a(n+1,l) = \dfrac{n+1}{l} a(n,l) - \dfrac{e^{-l}}{l}$

we have

$$\begin{aligned} E\{Z^n\} &= a(n,l)\left\{\frac{l}{(l+m)}(l+m-mn)\right\} + \\ &\left(\frac{m}{(1-m)} + 2m(n+1-b) + (1-m)(n+2-b)(n+1-b)\right) \\ &\left\{\frac{\Gamma(n-b+1)}{l^{n-b+1}(l+m)} - \frac{a(n-b,l)}{(l+m)}\right\} + \frac{e^{-l}}{(l+m)} \\ &\{2m + ml + (1-m)(l+n+2-b)\} \end{aligned} \tag{25}$$

To evaluate the correctness of proposition 2 we consider an example for *n=1*. For this example we have, $E\{Z\} = a(1,l)\left\{\dfrac{l^2}{(l+m)}\right\} + \dfrac{e^{-l}\{1+l\}}{(l+m)}$ which simplifies to $E\{Z\} = 1/(l+m)$ where $a(1,l) = 1/l^2(1-e^{-l}) - e^{-l}/l$.

3.1 Algorithm

In this section we provide an estimation procedure for estimating the arrival rates of the exponential and Pareto density functions, $[\hat{m}, \hat{l}]$ for measured statistics of the mixed stream. The measured statistics can be any moment of the mixed stream which is a function of $[m, l]$. For simplicity, we choose the first and the second moments (fmom, smom) in this paper. The following procedure explains the estimation algorithm. Initially we form a look-up table with coarse steps of $[m, l]$ and we compute the Euclidean distance to choose the minimum distance points. The surface of *E(X)* and *E(X^2)* for all pairs of $[m, l]$ is shown in Fig. 5. Then we form look-up table around the minimum distance points and the algorithm is iterated until the required accuracy d is reached. The algorithm is given as follows

$[\hat{m}, \hat{l}]$ = **procedure** FindParam(fmom, smom)

(a) **Compute** $E\{X\}_{ij}$ and $E\{X^2\}_{ij}$ using equation (7) $\forall$ (m_i, l_j) with a given *step size*, where $0.5 < m < 1$ and $l > 0$.

(b) **Calculate** Euclidian distance between (fmom, smom) and $\left(E\{X\}_{ij}, E\{X^2\}_{ij}\right)$ $\forall$ (m_i, l_j).

(c) **if** $\min_{i,j}(fmom - E\{X\}_{ij})^2 \le d$ and

$\min_{i,j}(smom - E\{X^2\}_{ij})^2 \le d$ **then** $[\hat{m}, \hat{l}] = (m_i, l_j)$

else decrease the *step size*

end

4. SIMULATION RESULTS

In the simulation setup, we estimate the parameters of the input streams from the moments of the multiplexed stream. Initially we generate synthetic traffic based on Poisson and heavy-tailed streams with the true parameters $[m, l]$ and statistically multiplex them to obtain the output stream *Z*. The input streams consisted of 100,000 samples. We perform three set of experiments where the true values of $[m, l]$ are [0.035, 0.5]. [0.95, 1] and [2, 0.75] respectively. In particular simulation was run for 200,000 time units (This can be expressed in terms milliseconds, microseconds or seconds based on whether $[m, l]$ is expressed on the basis of per milliseconds, microseconds or seconds). Once the samples of the output stream are generated we obtain the true moments that are depicted in columns 1 and 2 of Table 1 using the moment equations. Using the estimation algorithm we estimate the arrival rates of the input streams as shown in columns 3 and 4 of the table. The surface of *E(X)* vs. $E(X^2)$ generated for the valid ranges of $[m, l]$, $0.5 < m < 1$ and $l > 0$ is shown in Fig.5. Using these estimated values, $[\hat{m}, \hat{l}]$, synthetic multiplexed traffic are generated and the first and second moments are calculated and this is depicted as the estimated *E(X)* and $E(X^2)$ in the table. Results demonstrate that the estimated arrival rates are indeed close to the true values. The standard deviation of the estimates averaged over 100 runs was of the order 10^{-4} .

5. CONCLUSION AND FUTURE WORK

A problem of determining the degree of mixing of two independent streams of traffic from observations of their statistically multiplexed stream was presented. The paper provides a general expression for the PDF of the inter-arrival time mixed stream in terms of the PDF of the input inter-arrival times. The paper provides an expression to compute the moments of the output stream. An approach was provided to estimate input parameters from the first and second order statistics of the output traffic for the specific case of multiplexing Poisson and heavy-tailed processes. For arrival rate estimation of the input streams, we have proposed a look–up table approach based on nearest neighbor search. The results demonstrate that the estimated arrival rates are indeed close to the true values. Attempting solution to the problem through recursive means such as EM algorithm can be a motivating extension to compare against the current approach.

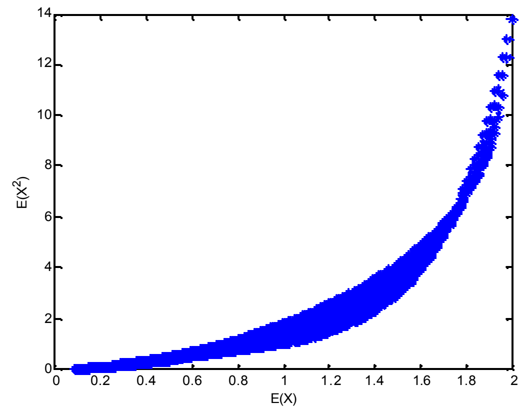

Fig. 5. Surface of E(X) vs. $E(X^2)$.

E(X)	$E(X^2)$	Estimated l	Estimated m	Estimated E(X)	Estimated $E(X^2)$
1.87	8	0.0352	0.5	1.8679	8.0181
0.5128	0.3931	0.9456	1.0050	0.5130	0.3930
0.3636	0.2389	2	0.75	0.3635	0.2391

Table 1. Estimated arrival rates from the moments

REFERENCES

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

[2] V. Paxson and S. Floyd, "Wide-area traffic: the failure of Poisson Modeling," presented at Proceedings of ACM SIGCOMM'94, 1994.

[3] J. Beran, R. Sherman, M. S. Taqqu, and W. Willinger, "Long-Range Dependence in Variable-Bit-Rate Video Traffic," *IEICE Transactions on Communications*, vol. 43, pp. 1566-1579, 1995.

[4] M. W. Garrett and W. Willinger, "Analysis, Modeling, and Generation of Self-Similar VBR Video Traffic," presented at Proceedings of SIGCOMM Conference, London, 1994.

[5] K. Sriram and Ward Whitt, "Characterizing superposition arrival processes in packet multiplexers for voice and data," *IEEE Journal on Selected Areas in Communications*, vol. SAC4, pp. 833--846, September 1986.

[6] J. Cao, W. S. Cleveland, D. Lin, and D. X. Sun, "The Effect of Statistical Multiplexing on the Long-Range Dependence of Internet Packet Traffic," *Bell Labs Technical Journal*, pp. 86, 2001.

[7] V. K. Rohatgi, *An Introduction to Probability Theory and Mathematical Statistics*, 1976.

[8] V. K. Murthy, *The General Point Process*: Addison-Wesley, 1974.