Some Models Parameters Calculation for Simulation of Network Traffic Marginal Distribution and Self-similarity

Alexander Yu. Privalov
Samara State Aerospace University
Moskovskoe shausse, 34
Samara, 443086, Russia
+7(846) 335-1920, +7(846) 267-4331
privalov@smr.ru

Alexander V. Blagov
Samara State Aerospace University
Moskovskoe shausse, 34
Samara, 443086, Russia
+7(846) 335-1837, +7(906) 340-9091
alex_ssauprof@mail.ru

ABSTRACT
In this paper, we present a simple numerical method of parameters calculation for well known traffic models M/G/∞ Input and On-Off Sources, which allows to simulate both self-similar properties and marginal distribution of real network traffic. Some modifications of these models are considered. These modifications allow to simulate network traffic with several important statistical characterizes, such as marginal distribution, autocorrelation function and Hurst parameter like in real network traffic.

Categories and Subject Descriptors
I.6.5. [**Computing Methodologies**]: Simulation and Modeling – *Model Development*

General Terms
Algorithms, Experimentation

Keywords
Self-similar Traffic, Traffic Simulation, M/G/∞ Input, On-Off Sources, Long Range Dependence, Heavy-tailed Distribution

1. INTRODUCTION
We present simple algorithms of parameters calculation for two structural models, which allows to simulate both self-similar properties and marginal distribution of real network traffic. By structural models we assume models which is not only mathematical constructs to yield necessary characteristics, but construction which reflects nature of network traffic. Parameters of such models usually have clear practical sense, suitable for purposes of simulation, prediction, performance evaluation and so on.

Modern network traffic has two important features: first is that the traffic is result of simultaneous information exchange between large number of sources and destinations, and second is that this information exchange can have very different parameters for different pairs source-destination.

We consider two of well known modern traffic models, which directly reflects these features: M/G/∞ Input and On-Off Sources. In both of them network traffic is the sum of information flows from many sources, and total traffic characteristics depend upon of features of individual sources parameters. For example, such interesting feature of modern network traffic as self-similarity is the consequence of a fact, that duration of information exchange between one source and its destination can be considered as random value with heavy tailed distribution.

Goal of this report is to present simple numerical algorithms for calculation of individual source characteristics, which provides given marginal distribution and self-similarity of total traffic. Examples of marginal distribution and Hurst parameter of real traffic we get from some traces of Internet Traffic Archive [1].

Self-similarity of modern network traffic attracts a lot of researches attention during last two decades (see, for example, comprehensive review and collection of important results [2]). There are many successful and well known approaches to self-similar traffic modeling, like multi-scale models and fractal noise models (see, for example, resent works [3-4]). But among them M/G/∞ Input and On-Off Sources models remain important and attractive for practice.

One of the first articles about models considered in this report is [5]. It is impossible to mention all the important works in this field, so just for examples of article about self-similar traffic modeling by M/G/∞ Input and ON-OFF sources models let us mention a few: [6-10].

2. EXPERIMENTAL MATERIAL PREPARATION
For our experiments we took four well known traces: BC-pAug89.TL, BC-pOct89.TL, LBL-PKT-5.TCP, dec-pkt-2.tcp. First two of them are Ethernet traces from Bell Laboratories [11], others – TCP traces from Lawrence Berkeley Laboratory and Digital Equipment Corporation [12].

All these traces are a sets of records about information packets transfers through some point of appropriate networks. We consider discrete time models, therefore we need to transform these traces into realizations of discrete time random processes. We take time slot duration of 10 ms, and a value of the process at some discrete time moment is a sum of sizes of all packets arrived during

Proceedings 23rd European Conference on Modelling and Simulation ©ECMS Javier Otamendi, Andrzej Bargiela, José Luis Montes, Luis Miguel Doncel Pedrera (Editors)
ISBN: 978-0-9553018-8-9 / ISBN: 978-0-9553018-9-6 (CD)

appropriate time slot. For practical purposes and for simplification of calculations we also scale these processes by division its values on some constant and round up the results. In practice of modern networks we don't need a one byte accuracy, so for this work our scale coefficient was 64. So, random processes under consideration are discrete time integer valued processes.

Marginal distributions of these processes are estimated by histogram method, and Hurst parameters are estimated by aggregate variance method. So, below we will assume, that Hurst parameter H and marginal distribution $\Pr\{Y_t=k\}$ for k=0,1,… of traffic process Y_t are given.

3. M/G/∞ INPUT MODEL

In this model, a random value of new sources arrives into system each time slot. Numbers of new sources in all time slots are i.i.d. random values with generic random value ξ, which has Poisson distribution with parameter λ.

In opposite to the simplest case, when all sources have the same information rate, we consider the case, when the rate of different sources can be different. It gives opportunity to get necessary marginal distribution of total traffic. So, rates of each new source at the moment of its arrival are realizations of i.i.d. random values with generic integer random value ζ, with some distribution, which we are going to find below. Each source keeps this information rate during all its lifetime in the system. After end of its lifetime the source leaves the system forever.

Lifetimes of all sources are realization of i.i.d. random values with generic integer random value τ, which has heavy tailed distribution

$$\Pr\{\tau = k\} = \frac{A}{k^\alpha}, \quad 2 < \alpha < 3, \quad k = 1,2,\ldots \qquad (1)$$

As shown by many authors (see, for example, [2] and [8]), such lifetime distribution provides self-similarity (second order asymptotical self-similarity) of total traffic (which is the sum of the flows from all active sources) with Hurst parameter H=2-α/2. So from given H we get α as

$$\alpha = 4 - 2H . \qquad (2)$$

3.1 Some previous analytical results

As shown in [13], it is easy to get equation for probability generation functions (p.g.f.) of traffic distribution $\Phi_Y(z)$ and source rate distribution $\Phi_\zeta(z)$. If

$$\Phi_Y(z) = \sum_{k=0}^{\infty} \Pr\{Y_t = k\} z^k \,, \Phi_\zeta(z) = \sum_{k=0}^{\infty} \Pr\{\zeta = k\} z^k$$

then taking under consideration that for stationary traffic process in each time slot the number of active sources N_t is Poisson random value with parameter $\Lambda=\lambda^*E\tau$, it is easy to get

$$\Phi_Y(z) = \sum_{k=0}^{\infty} \Pr\{N_t = k\} (\Phi_\zeta(z))^k = e^{\Lambda(\Phi_\zeta(z)-1)} ,$$

and

$$\Phi_\zeta(z) = 1 + \frac{1}{\Lambda} \ln \Phi_Y(z)$$

If *no fictive sources* condition is assumed, which means that

$$\Pr\{\zeta = 0\} = 0 \qquad (3)$$

and leads to $\Phi_\zeta(z) = 0$, then in some particular cases this equation can be solved analytically and has unique solution . For example, if Y_t is assumed to have Pascal distribution with parameters m and θ, then

$$\Phi_\zeta(z) = 1 + \frac{1}{\Lambda} \ln \frac{(1-\theta)^m}{(1-\theta z)^m} = 1 + \frac{m}{\Lambda} \ln \frac{1-\theta}{1-\theta z} ,$$

and *no fictive sources* condition gives Λ= -m ln(1-θ) and

$$\Pr\{\zeta = k\} = -\frac{1}{\ln(1-\theta)} \frac{\theta^k}{k}, \quad k = 1,2,\ldots .$$

Unfortunately, marginal distribution of real network traffic is not even close to Pascal or any other analytically defined distribution, so below algorithm from [13] for numerical solution is presented.

3.2 Numerical calculation of source rate distribution

Let us denote

$$p_k = e^{-\Lambda} \frac{\Lambda^k}{k!} . \qquad (4)$$

Then *no fictive sources* condition gives

$$\Pr\{Y_t = 0\} = \Pr\{N_t = 0\} = p_0 \qquad (5)$$

from which

$$\Lambda = -\ln(\Pr\{Y_t = 0\}) \qquad (6)$$

By definition and according to *no fictive sources* condition, random value ζ can take only integer values starting from 1. Taking this fact into consideration, we can write down that

$$\Pr\{Y_t = 1\} = p_1 \Pr\{\zeta = 1\}$$

and

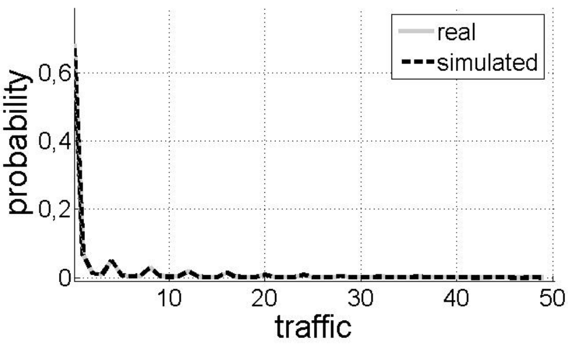

Figure 1. Traffic Distribution for LBL-PKT-5.TCP trace

$$\Pr\{\zeta = 1\} = \frac{\Pr\{Y_t = 1\}}{p_1}.$$

On the same way we have

$$\Pr\{Y_t = 2\} = p_1 \Pr\{\zeta = 2\} + p_2 \Pr\{\zeta_1 + \zeta_2 = 2\}$$

where ζ_1 and ζ_2 are i.i.d. random values with the same distribution as ζ. It means that minimal value both of ζ_1 and ζ_2 is 1, so

$$\Pr\{\zeta_1 + \zeta_2 = 2\} = \Pr\{\zeta_1 = 1, \zeta_2 = 1\} = (\Pr\{\zeta = 1\})^2$$

and

$$\Pr\{\zeta = 2\} = \frac{\Pr\{Y_t = 2\} - p_2(\Pr\{\zeta = 1\})^2}{p_1}.$$

Similar considerations give expression for general case:

$$\Pr\{\zeta = k\} = \frac{\Pr\{Y_t = k\} - \sum_{m=2}^{k} p_m \Pr\left\{\sum_{n=1}^{m} \zeta_n = k\right\}}{p_1}, \quad (7)$$

where $\Pr\{\sum_{n=1}^{m} \zeta_n = k\}$ can be found as appropriate term in m-times discrete convolution of sequence $\Pr\{\zeta=1\}$, $\Pr\{\zeta=2\}$, …, $\Pr\{\zeta=k-m\}$, which is initial part of source rate distribution found on previous iterations (7).

Equations (5)-(7) are iterative procedure to find source rate distribution numerically. Last parameter, which is necessary to find for the model is

$$\lambda = \frac{\Lambda}{E\tau}, \quad (8)$$

where $E\tau$ is calculated also numerically from (1) and (2).

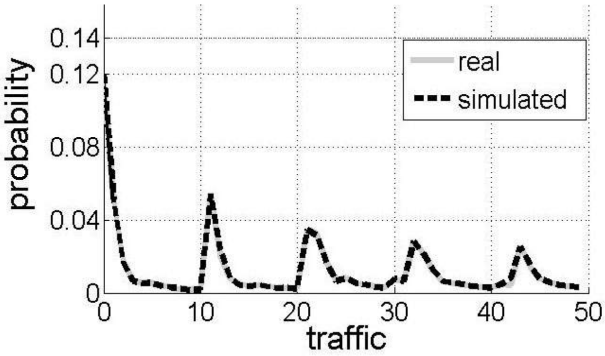

Figure 2. Traffic Distribution for DEC-PKT-2.TCP trace

There are two problems in practical calculation according (7): possible negative results of some iterations and when to stop these iterations. As solution for first problem, we used obvious approach: if some iteration according (7) gives negative value, appropriate probability (which is used for next iterations) is made equal to 0. Iterations are stopped, when sum of probabilities becomes close to 1 less then on 10^{-6}, and last probability is made equal to difference between 1 and sum of previous probabilities.

3.3 Numerical results for M/G/∞ Input model

Using H and $\Pr\{Y_t=k\}$ from real traffic traces in equations (1)-(8) we found α, λ and sequence of source rate probabilities $\Pr\{\zeta=k\}$. All these parameters were substituted into program of traffic simulation by M/G/∞ Input model. For simulated traces of length 360000 time slots, Hurst parameters and marginal distribution were estimated by the same methods as for real ones. For TCP traces comparison of these results with real ones are shown on Figures 1 and 2, and in Tables 1 and 2. Source rates for both traces are shown on Figures 3 and 4. It looks like M/G/∞ Input model works well for these kind of traces.

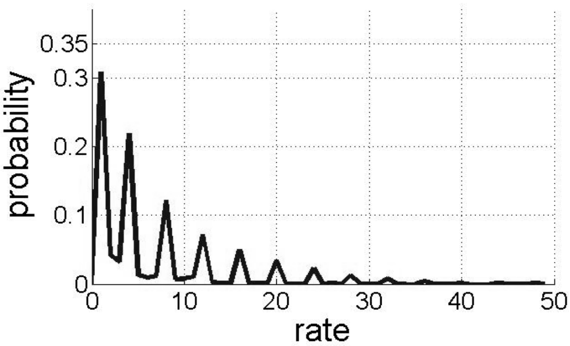

Figure 3. Source Rate Distribution for LBL-PKT-5.TCP trace

But for Ethernet traces the situation is much worse (see Figures 5 and 7 in next section). It could be several possible explanations of these facts. One of them is follow: modeling of new sources arrival as Poisson process is not suitable enough for local networks. In next section we consider another well known model, which is close to M/G/∞ Input model, but with different assumptions about number of sources in the system.

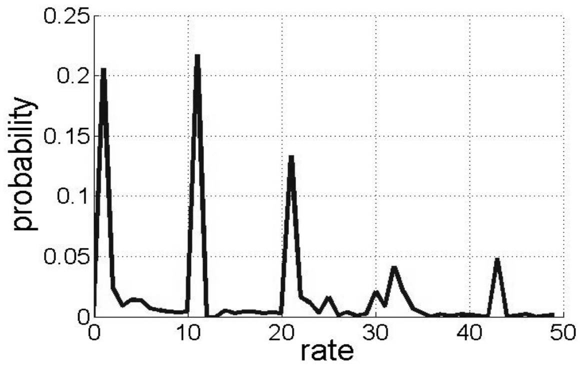

Figure 4. Source Rate Distribution for DEC-PKT-2.TCP trace

Table 1. LBL-PKT-5.TCP trace parameters comparison

Trace	H	Average	Variance
Real	0,666	2,017	30,545
Simulation	0,748	2,076	31,112

Table 2. DEC-PKT-2.TCP trace parameters comparison

Trace	H	Average	Variance
Real	0,755	49,528	2384,076
Simulation	0,771	46,592	2095,209

3.4 Approximation of normalized autocorrelation function

In addition to Hurst parameter and marginal distribution, M/G/∞ Input model can simultaneously approximate autocorrelation function of real traffic. To do this, we use source lifetime distribution in form

$$\Pr\{\tau = k\} = \begin{cases} y, & k = 1, \\ \dfrac{A}{(k+x)^{\alpha}}, & k = 2,3\ldots \end{cases}$$

instead of (1), x and y are fitting parameters.
Using

$$y + \sum_{k=2}^{\infty} \frac{A}{(k+x)^{\alpha}} = 1$$

we get

$$A = \frac{1-y}{\sum_{k=2}^{\infty} (k+x)^{-\alpha}}$$

For normalized autocorrelation function of M/G/∞ Input traffic (see for example [2])

$$r(k) = \frac{\sum_{i=k}^{\infty} \sum_{n=i+1}^{\infty} \Pr\{\tau = n\}}{y + \sum_{n=2}^{\infty} \Pr\{\tau = n\} + \sum_{i=1}^{\infty} \sum_{n=i+1}^{\infty} \Pr\{\tau = n\}}$$

After some transformations we get:

$$r(k) = \frac{(1-y)\sum_{i=k}^{\infty} \sum_{n=i+1}^{\infty} (n+x)^{-\alpha}}{\sum_{n=2}^{\infty} (n+x)^{-\alpha} + (1-y)\sum_{i=1}^{\infty} \sum_{n=i+1}^{\infty} (n+x)^{-\alpha}}$$

Using denotation $\hat{r}_k$ for autocorrelation function of the real traffic trace, and performing minimization of

$$S(x,y) = \sum_{k=1}^{N} \left(r(k) - \hat{r}_k \right)^2$$

by x and y, we get this results.
In our experiments we made minimization by using MatLab.
Let us note, that after modification of lifetime distribution, it is necessary to reculculate λ.
Finally, autocorrelation function of simulated traffic is much close to real then before. (see on Figures 5).

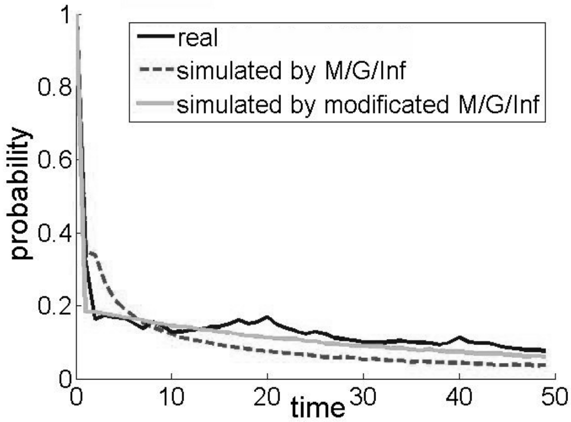

Figure 5. Normalized autocorrelation functions for DEC-PKT-2.TCP trace rate

4. ON-OFF SOURCES MODEL

In opposite to previous model, here we assume, that there are M sources of traffic in the system and each of them has two possible states: ON and OFF. In OFF state the source keeps silence, and in ON state it generates information flow. Each source switches between these states from time to time.

We consider the case, when each source during different ON periods can have different information rates. As before, rate for each new ON period are chosen at the beginning of the period and equal to realization of one of i.i.d. random values with generic integer random value ζ.

Durations of all ON periods of all sources are realizations of i.i.d. random values with generic integer random value τ, which has heavy tailed distribution (1). Again, α is calculated according to (2). As before, it provides second order asymptotical self-similarity of total traffic with given Hurst parameter. Durations of OFF periods of all sources are i.i.d with generic integer random value υ, which has geometric type distribution

$$\Pr\{\upsilon = k\} = p(1-p)^{k-1}, \quad k = 1,2,\ldots \tag{9}$$

where p is necessary to find.

4.1 ON-OFF version of source rate distribution calculation

From theory of alternating renewal processes stationary probabilities of each state for single source are given by

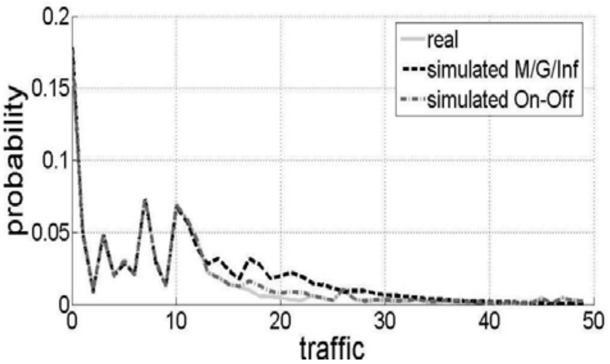

Figure 6. Traffic Distribution for BC-pAug89.TL trace

$$P_{ON} = \frac{E\tau}{E\tau + E\upsilon} = \frac{E\tau}{E\tau + 1/p} = \frac{p\,E\tau}{p\,E\tau + 1} \tag{10}$$

$$P_{OFF} = \frac{E\upsilon}{E\tau + E\upsilon} = \frac{1}{p\,E\tau + 1} \tag{11}$$

$E\tau$ can be calculated numerically as before. Let us denote here p_i as probability to have i sources in ON state in some time slot. Independence of all sources gives, obviously, that

$$p_i = \binom{M}{i} P_{ON}^{i}\ P_{OFF}^{M-i}, \quad i = 0,1,\ldots M \tag{12}$$

No fictive sources condition gives, again, expression (5), and from (12) here we have

$$P_{OFF} = \sqrt[M]{\Pr\{Y_t = 0\}} \tag{13}$$

Using (11) and (13), we get value of p for distribution (9):

$$p = \frac{1}{E\tau}\left(\frac{1}{\sqrt[M]{\Pr\{Y_t = 0\}}} - 1\right) \tag{14}$$

To get procedure for source rate distribution, we can follow exactly the same way as for M/G/∞ Input model, with only one difference: number of sources now can not be more, then M. So, it is easy to see, that now expression (7) is valid only for $k \le M$ (where p_i are calculated by (12)). For $k > M$ it must be

$$\Pr\{\zeta = k\} = \frac{\Pr\{Y_t = k\} - \sum_{m=2}^{M} p_m \Pr\left\{\sum_{n=1}^{m} \zeta_n = k\right\}}{p_1} \tag{15}$$

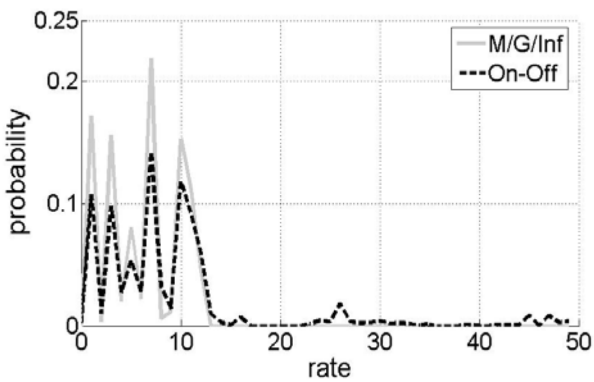

Figure 7. Source Rate Distribution for BC-pAug89.TL trace

4.2 Numerical results for ON-OFF sources model

Using H and $\Pr\{Y_t=k\}$ from Ethernet traffic traces in equations (1), (2), (5), (9), (14), (15) for different values of M=2,…10 we found α, p and sequence of source rate probabilities $\Pr\{\zeta=k\}$, and made the same numerical experiments for each M used, as for M/G/∞ Input model. Then we choose the best M in terms of closest of real and simulated traffic distributions.

For both Ethernet traces the best M is 2. (It is necessary to note, that the best M is 1, obviously, but this case is trivial. This case does not give any inside in network behavior, it just repeats traffic distribution as unique source distribution, therefore, we don't consider it.)

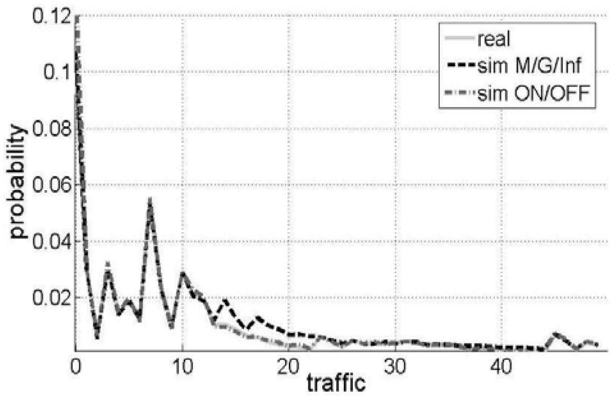

Figure 8. Traffic Distribution for BC-pOct89.TL trace

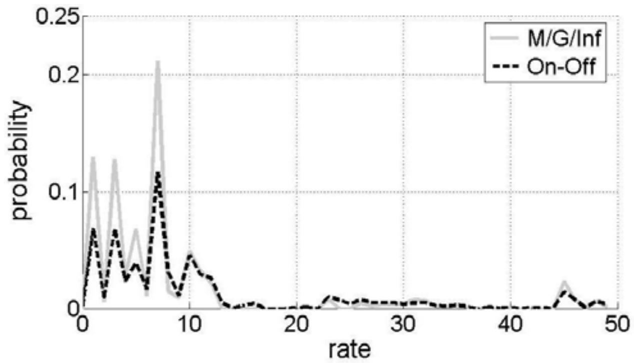

Figure 9. Source Rate Distribution for BC-pOct89.TL trace

The results for both models are shown on Figures 8, 9 and in Tables 3, 4. Source rates for both traces are shown on Figures 8 and 9.

Table 3. BC-pAug89.TL trace parameters comparison

Trace	H	Average	Variance
Real	0,786	28,786	2158,722
Sim. M/G/∞	0,792	11,069	96,212
Sim ON-OFF	0,781	23,055	1315,212

Table 4. BC-pOct89.TL trace parameters comparison

Trace	H	Average	Variance
Real	0,802	75,576	5948,424
Sim. M/G/∞	0,807	62,603	5286,218
Sim ON-OFF	0,789	71,755	6520,264

5. REMARKS AND CONCLUSIONS

Presented above results show that M/G/∞ Input model works better for TCP traces then for Ethernet traces. The explanations of this fact can be different. One of possible explanation is follow: local network of large company usually consists of a few big segments connected to central cable by bridges. From central cable "point of view" these bridges are looked like main sources of traffic. Therefore On-Off sources model with small number of big sources is more adequate, then assumption about large number of small sources. But actual reason of this situation is still open problem.

Another problem is not very good accuracy in simulation of average and variance of real traffic. It is because algorithms presented here "pay most attention" to initial part of marginal distribution, but not to tail. But tail of distribution can affect much on first moments, especially on variance. So, some possible direction of future work is algorithms modification for more accurate simulation of marginal distribution tail. Another way to improve accuracy is more accurate estimation of $E\tau$. One possible such improvement was presented above. More accurate estimation of $E\tau$ of real traffic, for example, by estimation of initial part of autocorrelation function (see [2]), can improve accuracy of simulation marginal distribution and autocorrelation function simultaneously.

In conclusion we would like to note that models considered here is very useful for practice. For example, prediction by simulation of some growing organization network behavior can be made easily, if we take current organization traffic, find such parameters as λ (for M/G/∞ Input model) or p (for On-Off sources model), and increase it. In case of more complex analysis, it can be possible to identify the nature of individual source rate distribution in terms of mostly used network applications and make some prediction of this evolution.

6. REFERENCES

[1] Internet Traffic Archive. DOI= http://ita.ee.lbl.gov

[2] Self-similar Network Traffic and Performance Evaluation. 2000. K. Park and W. Willinger Ed. Wiley. New-York.

[3] Liu N.X. and Baras J.S. 2003. Statistical modeling and performance analysis of multi-scale traffic. In Proceedings of IEEE INFOCOM'03. (March 2003) 1837-1847.

[4] Zeng Y. and Chen T.M. 2004. Measurement-based Real Time Traffic Model Classification. In Proceedings of IEEE ICC. (June 2004), 1857-1861.

[5] Cox D. R. 1984. Long Range Dependence: a review. In Statistics: an Appraisal. H. A. David and H. T.

David Ed. The Iowa State University Press. Ames. 55-74.

[6] Jelenkovic P.R. and Lazar A.A. 1997. Multiplexing On-Off Sources with Subexponential On periods. Part II. In Proceedings of the 15th International Teletraffic Congress (Washington, DC, June 1997), 965-974.

[7] Liu Z and all. 1998. Asymptotic Behavior of a multiplexer Fed by a Long Range Dependent Process. Applied Probability, 36(1), 105-118.

[8] Tsybakov B. and Georganas N.D. 1998. Self-similar processes in communications network. IEEE Trans. On Information Theory. 44(5) (September 1998), 1713-1725.

[9] Krunz M.M. and Makowski A.A. 1998. Modeling Video Traffic Using M/G/∞ Input Process: A Compromise Between Markovian and LRD Models. IEEE Journal of Selected Areas in Communications. 16 (June 1998), 733-748.

[10] Adas A. 1997. Traffic Models in Broadband Networks. IEEE Communications Magazine. 35 (July 1997), 82-89.

[11] Fowler H.J. and Leland W.E. 1991. Local Area Network Traffic Characteristics with Implications for Broadband Network Congestion Management. IEEE Journal of Selected Areas in Communications. 9(7) (September 1991), 1139-1149.

[12] Paxson V. and Floyd S. 1995. Wide-Area Traffic: The Failure of Poisson Modeling. IEEE/ACM Transactions on Networking. 3(3) (June 1995), 226-244.

[13] Privalov A.Yu. and Baeva M.V. 2007. (in Russian). Some Approaches to Self-similar Traffic Modeling. In “Information technologies and system mathematical modeling 2006-2007”. Radiotechnika. Moscow. 81-84.

AUTHOR BIOGRAPHIES

Alexander Yu. Privalov received the M.S. degree from Moscow Institute of Physics and Technology, Moscow, Russia, in 1990, and the Ph.D. degree from the Institute of Information Transmission Problems, Moscow, Russia, in 1993.

He was with the Department of Technical Cybernetics Samara State Aerospace University (SSAU) as a Associate Professor in 1993 – 1996.

He was a Visiting Research scholar in Computer Science and Telecommunication with University of Missouri-Kansas City in 1996 – 1998.

Since 1998 he has been with SSAU (since 2002 as an Full Professor).

His research interests include modeling and performance evaluation of communication networks.

His e-mail address is: privalov@smr.ru

Alexander V. Blagov received the M.S. degree from Samara State Aerospace University (SSAU), Samara, Russia, in 2008

Since 2008 he is the Post-graduate (student) of the Department of Technical Cybernetics SSAU.

His research interests include modeling and telecommunications.

His e-mail address is: alex_ssauprof@mail.ru