Analysis and Simulation of WAN Traffic by Self-Similar Traffic Model with OMNET

Alexander Yu. Privalov
Computer Science Department
Samara State Aerospace University
Samara, Russia
privalov1967@gmail.com

Alexander Tsarev
Computer Science Department
Samara State Aerospace University
Samara, Russia
al-xandr1@yandex.ru

***Abstract*—An OMNET package with an implementation of a useful modification of a well-known self-similar traffic model is presented. The package is designed for the analysis of real WAN traffic traces, for the identification of model parameters and for the generation of synthetic traffic, which is close to real one by a number of characteristics simultaneously. These characteristics include traffic value probability distribution, normalized autocorrelation function and Hurst parameter. The performance of the package was tested with traffic traces from CAIDA and Internet Traffic Archive collections. These experiments revealed that the probability distribution of the modern WAN traffic has multimodal "saw-like" form. The reason for that interesting phenomenon is yet to be discovered.**

Keywords—telecommunication traffic analysis;telecommunication traffic simulation;self-similarity models;OMNET++.

I. INTRODUCTION

During the last two decades, following the publication of famous papers like [1], analysis and modeling of self-similar telecommunication traffic became very popular topics in theoretical and applied researches. Representative collection of theoretical results and practical achievements in this field during the first decades of these efforts can be found e.g. in book [2]. Currently, the impact of traffic self-similarity on different aspects of telecommunications is still an important research topic (see e.g. [3-6]).

One of the important problems in this field is developing tractable traffic models and evaluating their ability to match closely the real traffic behavior. The first purpose of this paper is to present OMNET++ realization of our modification of the well-known self-similar traffic model "M/G/∞ Input". The modified model is able to capture several characteristics of real WAN traffic simultaneously, including the Hurst parameter, the probability distribution of traffic value and the autocorrelation function (ACF) of traffic value. Another purpose of this paper is to make the parameters of the modified "M/G/∞ Input" model more tractable and explaining of the traffic behavior.

We would like to explain the difference between our work and the well-known paper [7], which provided deep theoretical analysis and a collection of practical results related to the ability of the "M/G/∞ Input" model to capture self-similar behavior, autocorrelation function and probability distribution of real traffic. The main difference between our work and [7] is that in [7] the information rates of the individual sources were assumed the same for all individual sources and equal to 1. Therefore the probability distribution of the original traffic of the "M/G/∞ Input" model was Poisson. To transform the Poisson distribution to the required probability distribution, a special transformation function was used. In our opinion, the main disadvantage of this approach is that it faces difficulties in tractability of the transformation function. The introduction of such a function is a mathematical trick, which solves the problem of the traffic simulation, but does not explain the traffic behavior. In our work, we use another approach. We assume that each individual source can have its own information rate, and the superposition of such nonuniform sources creates an appropriate probability distribution of the total traffic.

The rest of the paper is structured as follows. In Section 2 we provide an overview of the basic "M/G/∞ Input" model and present the modification implemented in the package. In Section 3 we describe the package and present the results of the real traffic simulation. Section 4 concludes the paper.

II. OVERVIEW OF THE "M/G/∞ INPUT" MODEL

For simplicity, we consider a discrete time situation. Let the time be divided into equal time slots, and let the size of each slot be a time unit. Also let the amount of information be measured in cells, and let the size of the each cell (in bytes) be determined out of accuracy considerations (in modern network analysis, the accuracy up to one byte usually is not necessary). Therefore, the traffic and the source rate are measured in information cells per time slot.

In the "M/G/∞ Input" model, a random number of new individual sources arrives into the system in each time slot. The amounts of new sources in all time slots are i.i.d. (independent identically distributed) random values with Poisson distribution. The parameter of the Poisson distribution is denoted as λ.

It is well-known that the self-similarity of the model traffic is determined by the heavy-tailed distribution of the individual sources lifetime. If lifetimes are i.i.d. random values with a generic random value T and probability distribution of T obeys

$$\lim_{k\to\infty} \Pr\{T=k\}/k^{-(4-2H)} = const > 0\,, \qquad (1)$$

then the Hurst parameter of the "M/G/∞ Input" model traffic equals to H.

To fit also a traffic autocorrelation function, we will use the following lifetime probability distribution:

$$\Pr\{T=k\} = \begin{cases} y, & k=1 \\ C(k+x)^{-(4-2H)}, & k=2,3\ldots \end{cases}. \qquad (2)$$

where x and y are fitting parameters, C is a normalized constant. As shown in [8] and [9], in a case of such parameterization, the normalized ACF of the model traffic value is

$$r(k) = \frac{(1-y)\sum_{i=k}^{\infty}\sum_{n=i+1}^{\infty}(n+x)^{-(4-2H)}}{\sum_{n=2}^{\infty}(n+x)^{-(4-2H)} + (1-y)\sum_{i=1}^{\infty}\sum_{n=i+1}^{\infty}(n+x)^{-(4-2H)}}\,. \qquad (3)$$

Having several values r_k, $k=1,2,\ldots,K$ of the real traffic normalized ACF, we can fit (3) by minimizing over x and y a mean square deviation

$$D(x,y) = \sum_{k=1}^{K}(r(k)-r_k)^2\,. \qquad (4)$$

At every timeslot during its lifetime, each individual source produces traffic (some amount of cells). The sum of streams from all active sources is the resulting traffic of the model. In basic "M/G/∞ Input" model, it is assumed usually that all sources have the same rate of traffic generation (information rate). In stationary state of the system it makes the value of total traffic Y_t at some random moment t strictly proportional to the number of active sources N_t in the system at the same moment, which is Poisson random value with parameter

$$L=\lambda * \mathrm{E}(T) \qquad (5)$$

(E() is used for mean operation). So, let us denote

$$p_k = \Pr\{N_t = k\} = \exp(-L)L^k/k!, \quad k=0,1,\ldots\,. \qquad (6)$$

Our main modification of the basic model is an assumption that source rates are i.i.d. random values with a generic random value S. If at some moment t there are N_t active sources in the system, let us denote their rates as S_k, $k=1,2,\ldots N_t$

As shown in [8] and [9], knowing probability distribution $\{\Pr\{Y=k\}\}_{k=0,1,2,\ldots}$ of traffic value Y, we can find the probability distribution of S, which provides this distribution of Y. Let us describe this procedure with some explanations. First of all, let us assume, that

$$\Pr\{S=0\} = 0\,. \qquad (7)$$

This "no dummy sources condition" provides that Y_t=0 if and only if N_t=0, so

$$\Pr\{Y=0\} = p_0 = e^{-L} \qquad (8)$$

and

$$L = -\ln(\Pr\{Y=0\})\,. \qquad (9)$$

The individual source rate S (in cells per slot) is integer valued random value, therefore Y_t=1 if and only if N_t=1 and S_1=1, so

$$\Pr\{Y=1\} = p_1 \Pr\{S=1\}, \qquad (10)$$

and we get Pr{S=1}. Then Y_t=2 if and only if N_t=1 and S_1=2 or N_t=2 and S_1+S_2=2 (which means $S_1=S_2$=1), so

$$\Pr\{Y=2\} = p_1\Pr\{S=2\} + p_2\Pr\{S_1+S_2=2\}, \qquad (11)$$

and knowing Pr{S=1}, we get

$$\Pr\{S=2\} = \frac{\Pr\{Y=2\} - p_2(\Pr\{S=1\})^2}{p_1} \qquad (12)$$

In the general case we get

$$\Pr\{S=k\} = \frac{\Pr\{Y=k\} - \sum_{m=2}^{k} p_m \Pr\{\sum_{n=1}^{m} S_n = k\}}{p_1}\,, \qquad (13)$$

where $\Pr\{\sum_{n=1}^{m} S_n = k\}$ can be found from m-times discrete convolution of sequence $\{\Pr\{S=i\}\}_{i=1,2,\ldots k-1}$, already known from previous iterations. In practice, calculation of these convolutions is the most time consuming part of the iteration procedure (7) and has exponential complexity when k grows.

From (5) we can find

$$\lambda = L/\mathrm{E}(T). \qquad (14)$$

So, in our model we use very tractable parameters: the new sources arriving rate λ, the probability distribution of individual source lifetime and the probability distribution of individual source information rate. The appropriate choice of these parameters allows us to simulate traffic with given probability distribution, normalized ACF and Hurst parameter.

III. Simulation of WAN traffic.

The presented OMNET++ package is available at [12] as SelfSimMGI.

First purpose of the package is the traffic analysis. The package provides a way to collect statistical data and to compute statistical characteristics and Hurst parameter from these data.

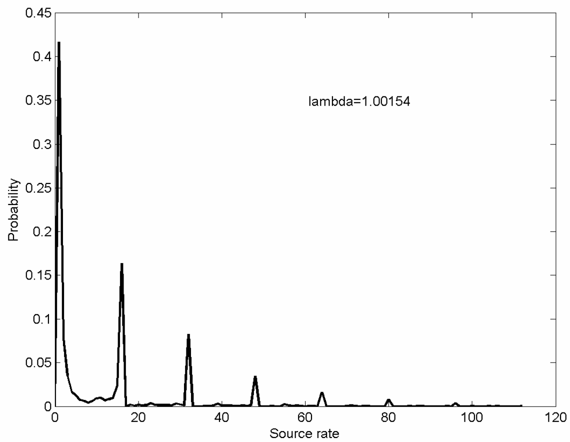

Figure 1. Source rate probability distribution for trace 1

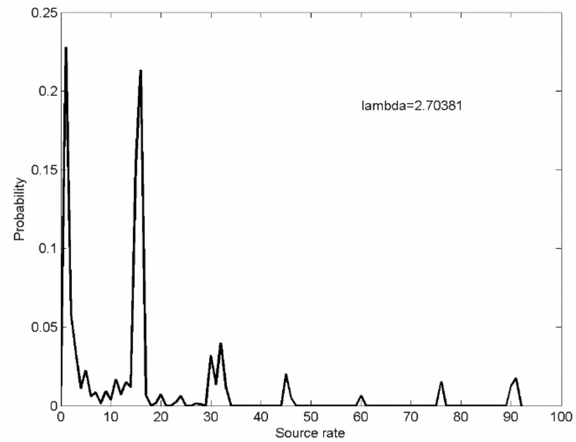

Figure 4. Source rate probability distribution for trace 2

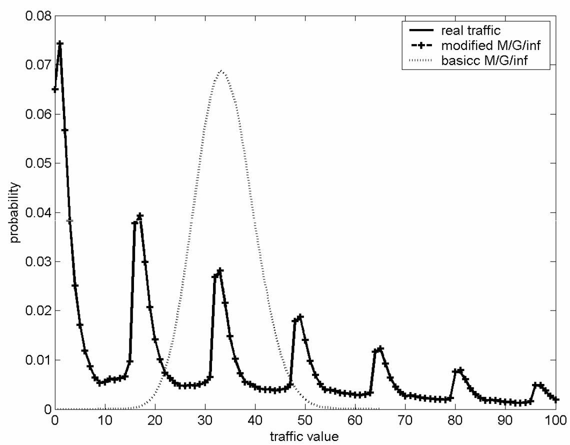

Figure 2. Traffic value probability distribution for trace 1

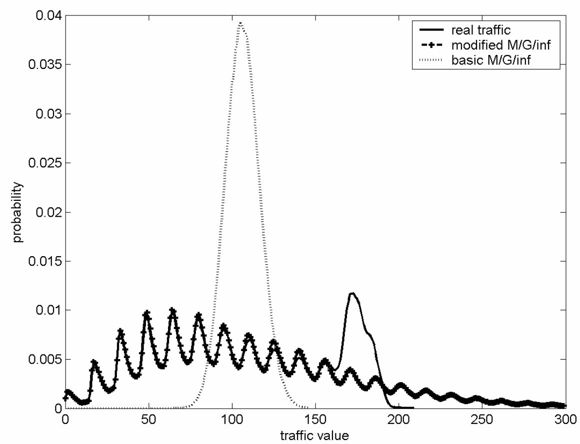

Figure 5. Traffic value probability distribution for trace 2

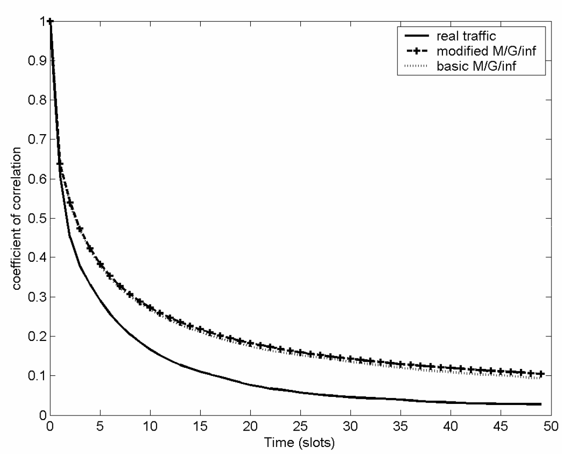

Figure 3. Normalized ACF for trace 1

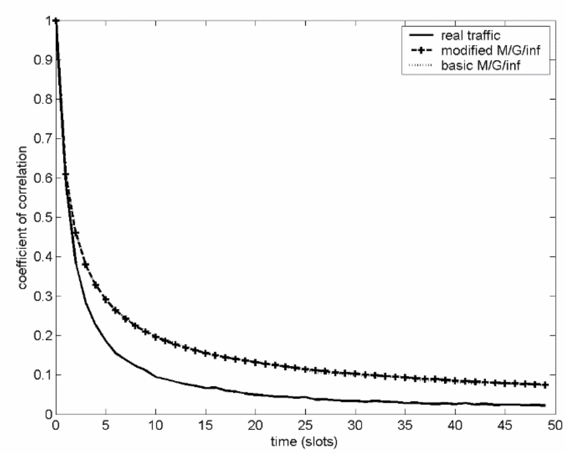

Figure 6. Normalized ACF for trace 2

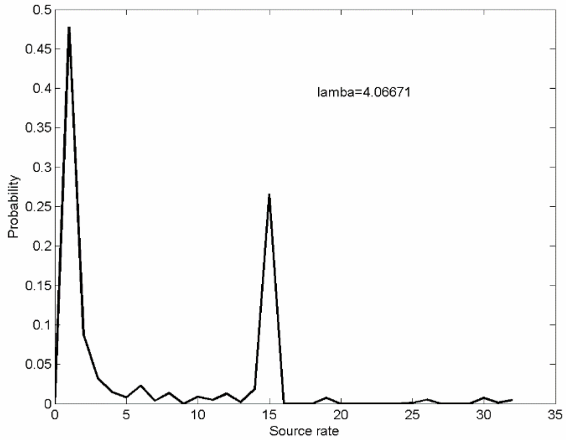

Figure 7. Source rate probability distribution for trace 3

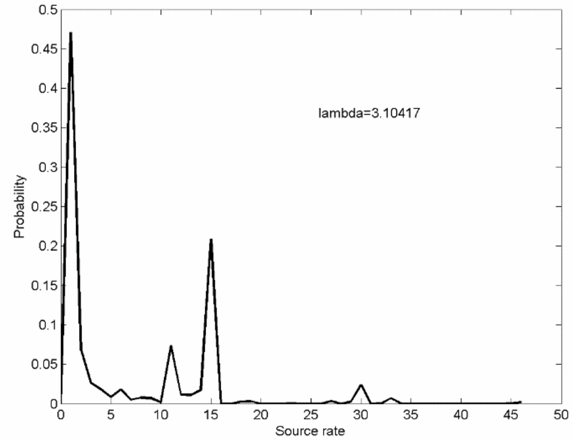

Figure 10. Source rate probability distribution for trace 4

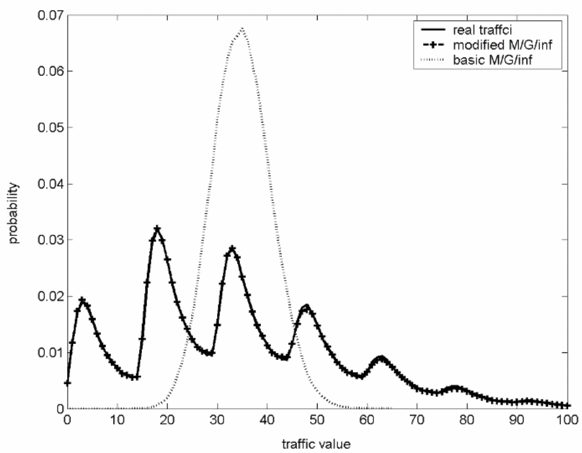

Figure 8. Traffic value probability distribution for trace 3

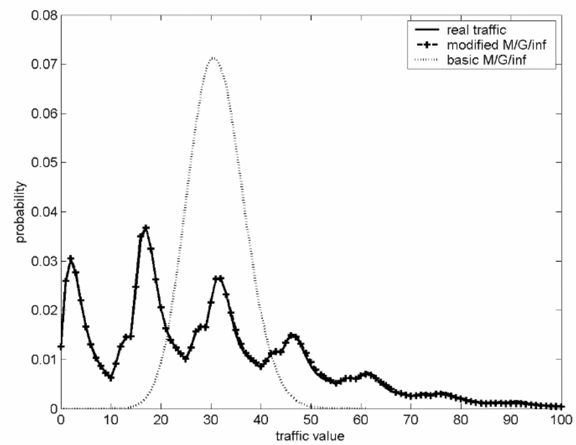

Figure 11. Traffic value probability distribution for trace 4

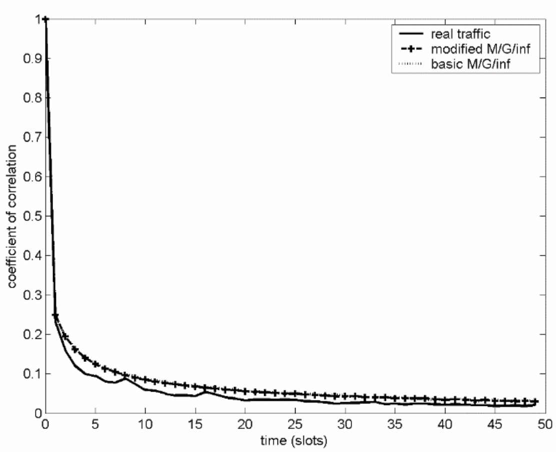

Figure 9. Normalized ACF for trace 3

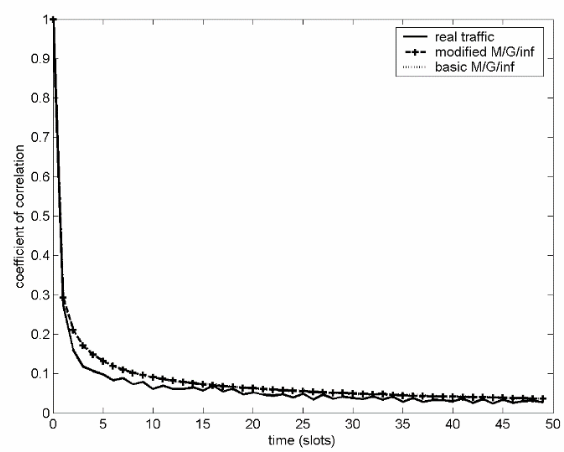

Figure 12. Normalized ACF for trace 4

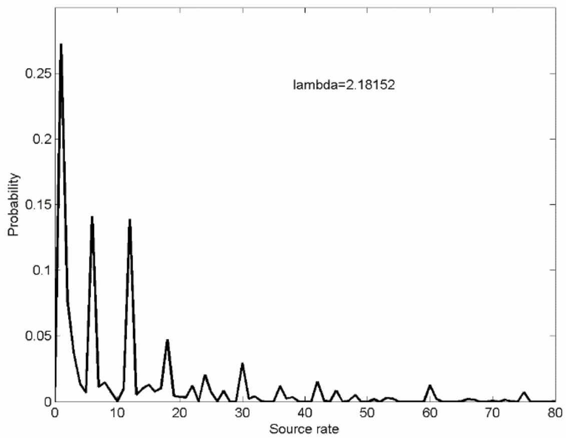

Figure 13. Source rate probability distribution for trace 5

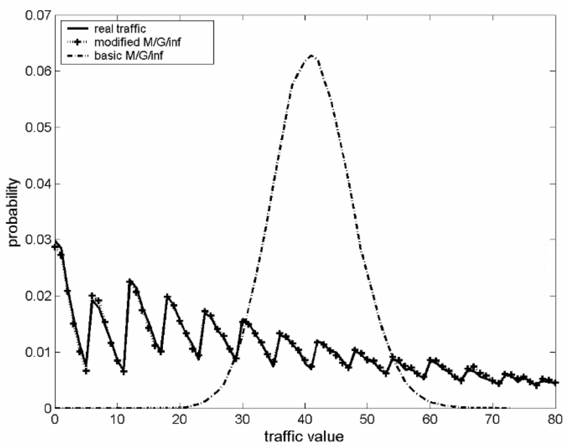

Figure 14. Traffic value probability distribution for trace 5

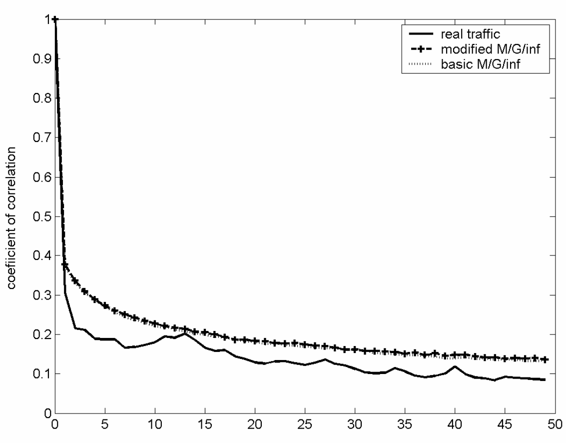

Figure 15. Normalized ACF for trace 5

There is a module in the package, which reads a traffic trace file, generates OMNET packets according to records in the file and sends packets to outgoing link (packet in OMNET is a special kind of message, which has a size attribute; by these messages information packets are modeled in OMNET).

There is an another module, statistics collector, which collects all necessarily statistical data from the stream of OMNET packets, computes probability distribution, normalized ACF and Hurst parameter of the traffic passed through, and writes results to an XML file.

A second purpose of the package is the traffic simulation. There is a module in the package, a parameter calculator, which reads the statistical information from the above mentioned XML file, computes parameters of "M/G/∞ Input" model and writes them into another XML file.

Another module, a traffic generator, reads these parameters from the file at the beginning of every simulation run and generates traffic as the stream of OMNET packets according to appropriate "M/G/∞ Input" model behavior. For testing the performance of the generator in our experiments, statistical data of the simulated traffic are collected by the above mentioned statistics collector.

Let us make some notes about computations inside the parameter calculator module. First of all, after substitution of the required H into (3), the mean square deviation (4) (K=4 gave the best results) is minimized over the variables x and y, making up the source lifetime probability distribution (2). Then L from (9) and λ from (14) are calculated. After that the iteration procedure is performed according to a following pseudo code:

Pr{S=0}**:=** 0; sum**:=** 0; k**:=** 0;
do {
    sum**:=** sum + Pr{S=k};
    k**:=**k+1; calculate Pr{S=k} according expression (13);
    if (Pr{S=k}<0) Pr{S=k}**:=**0;
} **while** (sum+Pr{S=k}<1 **and** k<maximal value of Y)**;**
Pr{S=k}**:=**1-sum; sum**:=**1; k is the maximal value of the source rate.

We present results of experiments for four traces from CAIDA collection [10] and one trace from Internet Traffic Archive [11]. Traces 1 and 2 are 1 minute traces from CAIDA equinix-chicago Internet data collection monitor in two opposite directions - direction A and direction B respectively. Both traces were taken on December 19, 2013, at 13:00. Traces 3 and 4 are 1 minute traces from CAIDA equinix-sanjose Internet data collection monitor in two opposite directions - direction A and direction B respectively. Trace 3 (direction A) was taken on November 20, 2008, at 13:00. Trace 4 (direction B) was taken on December 18, 2008, at 13:00. Processing of these traces was done using the timeslot equal to 15 microseconds and the cell size equal to 100 bytes.

Trace 5 is dec-pkt-2 trace from Internet Traffic Archive [11]. The trace was taken on March 9, 1995 at 02:00 and contains an hour's worth of all wide-area traffic between Digital Equipment Corporation and the rest of the world. Processing of this trace was done using the timeslot equal to 15 milliseconds and the cell size equal to 100 bytes.

Let us make some notes about practical choice of fundamental parameters of our model – the cell size and the time slot.

The cell size is determined out of required accuracy of traffic representation. In all our experiments presented here, the cell size was the same, and equal to 100 bytes. The time slot size is determined by computability of model parameters in a reasonable time. The most time consuming operation is multiple times convolution in (13). When maximal source rate becomes about thousand (cells per slot), then the calculation time becomes unacceptably long. So, the choice of an appropriate time slot depends on real traffic value and in practice we tried different values before a good one was found. Our "rule of thumb" is that after processing with appropriate time slot value, the maximal traffic value should be somewhere between several dozen and several hundred cells per slot. In this case procedure (13) works well and does not take too much time (several minutes on Pentium 2.8Gg for presented examples). So, for modern traces from CAIDA collection we took the time slot of 15 microseconds, and for the old trace from Internet traffic archive we used the time slot of 15 milliseconds.

Results about ability of the simulated traffic to approximate the real one are presented in a graphical form. Some quantitative data about approximation ability of the model can be found in [9], but those data are results of experiments with a previous implementation of the model, which was not OMNET package and traces for those experiments were not modern WAN traffic.

For each trace under consideration we present the source rate probability distribution and the source born rate λ, the traffic value probability distribution and the normalized ACF of real and simulated traffic. For comparison, the probability distribution and the normalized ACF of the traffic simulated by basic M/G/∞ Input model (where rates of all sources are equal to 1) are presented too.

It is interesting that all source probability distributions are multimodal, and it provides a very good approximation of the real traffic probability distribution, which is also multimodal. There is an interesting artifact on fig. 5 in the real trace probability distribution near values 150-200 (cells per slot). Taking under consideration that near the x-axis mark of 200 cells per slot on the figure, the value of real traffic is about 10 Gbit/sec, it is obvious that the communication link at this point was loaded close to 100%, and the flow control started to limit traffic (see [10] for network equipment description). Our model cannot process such situation yet.

TABLE I. HURST PARAMETER VALUES

Trace	**Hurst parameter**		
	Real traffic	***Modified M/G/∞***	***Basic M/G/∞***
Trace 1	0.66	0.75	0.74
Trace 2	0.71	0.73	0.70
Trace 3	0.70	0.71	0.72
Trace 4	0.75	0.73	0.74
Trace 5	0.89	0.84	0.84

As we can see, a multimodal behavior of the traffic probability distribution often takes place in WAN traffic during the last two decades, but traffic rates grew significantly over the same period of time.

Hurst parameter values for all traces are presented in a table 1. As we can see, self-similar and correlation behavior are captured by our M/G/∞ Input models good enough.

IV. CONCLUSION

The set of experimental results produced by SelfSimMGI OMNET package for WAN traffic analysis and simulation was presented. In the package the modified M/G/∞ Input model with easily tractable parameters is used. Processing of real traces by this package reveals interesting phenomenon of a multimodal "saw-like" form of the probability distribution of the modern WAN traffic. The reason of such behavior is still unclear and requires deeper research and understanding.

REFERENCES

[1] W. Leland, M. Taqqu, W. Willinger and D. Wilson "On the self-similar nature of Ethernet traffic (extended version)", IEEE/ACM Trans. On Networking, Vol. 2, No.1, February 1994, pp.1-15.

[2] Self-similar Network Traffic and Performance Evaluation, K. Park, W. Willinger., Ed. Wiley // John Wiley & Sons Inc. – New-York, 2000. – 556 p

[3] Xiaolong Jin and Geyong Min. "A heuristic flow-decomposition approach for generalized processor sharing under self-similar traffic",*J. Comput. Syst. Sci.* 74, 6 (September 2008), 1055-1066.

[4] Candelaria Sansores Perez, Luis Rizo Dominguez, and Julio Cesar Ramirez Pacheco. "A simulation scenario for performance studies under self-similar traffic." In *Proceedings of the 8th WSEAS international conference on Data networks, communications, computers* (DNCOCO'09), Stevens Point, Wisconsin, USA, 185-189.

[5] Anatoly M. Galkin, Olga A. Simonina, and Gennady G. Yanovsky. "Multiservice IP network qos parameters estimation in presence of self-similar traffic", In *Proceedings of the 6th international conference on Next Generation Teletraffic and Wired/Wireless Advanced Networking* (NEW2AN'06). Springer-Verlag, Berlin, Heidelberg

[6] G. Horn, A. Kvalbein, J. Blomskold and E. Nilsen. "An empirical comparison of generators for self similar simulated traffic", *Perform. Eval.* 64, 2 (February 2007), 162-190.

[7] M. M. Krunz and A. M. Makowski. "Modeling video traffic using M/G/∞ input processes: a compromise between Markovian and LRD models",*IEEE J.Sel. A. Commun.* 16, 5 (September 2006), 733-748.

[8] A. Privalov and A. Blagov. "Some Models Parameters Calculation for Simulation of Network Traffic Marginal Distribution and Self-similarity" ECMS. 23rd European conference of modeling and simulation. Madrid, 2009. pp. 51-60

[9] A. Privalov and A. Blagov "Approximative models of data network self-similar traffic." // LAP Lambert Academic Publishing - Germany, 2013, - 120 p *(in Russian*) ISBN 978-3-659-47353-1

[10] http://www.caida.org The Cooperative Association for Internet Data Analysis (CAIDA).

[11] http://ita.ee.lbl.gov The Internet Traffic Archive sponsored by ACM SIGCOMM.

[12] http://omnetpp.org/index.php/models/catalog Model catalog on OMNET++ community cite.