15

Modeling and Simulating the Self-Similar Network Traffic in Simulation Tool

Matjaž Fras[1], Jože Mohorko[2] and Žarko Čučej[2]
[1]Margento R&D, Maribor,
[2]University of Maribor, Faculty of Electrical Engineering
and Computer Science, Maribor,
Slovenia

1. Introduction

Telecommunication networks are growing very fast. The user's needs, in regards to new services and applications that have a higher bandwidth requirement, are becoming bigger every day. A telecommunication network requires early design, planning, maintenance, continuous development and updating, as demand increases. In that respect we are forced to incessantly evaluate the telecommunication network's efficiency by utilizing methods such as measurement, analysis modeling and simulations of these networks.

Measuring, analyses and the modeling of self-similar traffic has still been one of the main research challenges. Several studies have been carried-out over the last fifteen years on: analysis of network traffic on the Internet [30], [31], traffic measurements in the high speed networks [32], and also measurement in the next generation networks [33]. Also, a lot of research works exist, where attention had been given to analysis of the network traffic caused by different applications, such as P2P [34], [35], network games [36] and VoIP application Skype [37]. Analyses of the measured network traffic help us to understand the basic behavior of network traffic. Various have showed that traffic in contemporary communication networks is well described with a self-similar statistical traffic model, which is based on fractal theory [6]. The pioneers in this field are: Leland, Willinger, and many others [1], [5], [6]. They introduced the new network traffic description in 1994. New description appeared as an alternative to traditional models, as were Poisson and Markov, which were used as a good approximation for telephone networks (PSNT networks) when describing the process of call durations and time between calls [5], [20]. These models do not allow descriptions of bursts, which are distinctive in today's network traffic. Such bursts can be described by a self-similarity model [5], [6], because it shows bursts over a wide-range of time scales. This contrasts with the traditional traffic model (Poisson model), which became very smooth during the aggregation process. The measure of bursts and also self-similarity present the Hurst parameter [1]-[4], which is correlated with another very important property called long-range dependence [5]-[8]. This property is also manifested with heavy-tailed probability of density distributions [5], [6], such as Pareto [43] or Weibull [44]. So Pareto's and Weibull's heavy-tailed distributions became the most frequently used distributions to describe self-similar network traffic in communication networks.

During past years another aspect of network traffic studying has also appeared. In this case, the network traffic is researched from application or data source point of view, especially focused on statistics of file sizes and inter-arrival times between files [19]. These research works are very important for describing a relation between packet network traffic on lower ISO/OSI layers and data source network traffic on higher layers of ISO/OSI model. Based on the research of WWW network traffic, it has been shown that file sizes of such traffic are best described by Pareto distribution with shape parameter $a = 1$ [38]. That was also shown for the FTP traffic, where the shape parameter of Pareto distribution is in the range $0.9 < a < 1.1$ [20]. In [6], [39], and [40] it is shown that inter-arrival time of TCP connections are self-similar processes, which can be described by Weibull heavily tailed distribution.

With expansion of simulation tools, which are used for simulation of communication networks, the knowledge about simulating the network traffic also becomes very important. One of the important tasks in simulations is also knowledge about modeling and simulating of network traffic. Network traffic is usually modeled in simulation tools from an application point of view [42], [45]. It is usually supposed that the file size statistics and file inter-arrival times are known [39], [40]. Such kinds of traffic models are supported by most commercial telecommunication simulation tools such as the OPNET Modeler [10], [11], [24], used in our simulations and experiments. Consequently, for using the measured data of packet traffic, when modeling file statistics, it is necessary to transform packets' statistics into files' statistics [9, 10]. This transformation contains opposite operations in relation to the fragmentation and encapsulation process. Extensive research and investigation about traffic sources in contemporary networks show that this approach requires an in-depth analysis of packet's traffic (which needs specialized, very powerful and consequently, expensive instruments). This approach, in the case of encrypted packets and non-standard application protocols, is not completely possible. In such cases, capture of entire packets is also necessary, which can be problematic in contemporary high-speed networks. Another approach estimates distribution parameters of file data sources from measured packets' network traffic. For such approach, we have developed and tested different methods [42], [45]. Estimated distribution parameters are used for modeling of the measured network traffic for simulation purposes. Through the use of these methods we want to minimize discrepancies between the measured and simulated traffic in regards to an average bit rate and bursts, which are characteristic of self-similar traffic.

2. Network traffic

2.1 Packet network traffic measuring

The measuring and analyzing of real network traffic provide us with a very important knowledge about computer network states. In analyzing process, we need statistical mathematical tools. These tools are crucial for accuracy of a derived mathematical model, described by stochastic parameters for packet size and inter-arrival time [9]. Using this simulation model, we want to acquire information about telecommunication network's performances for:

- improvement of the current network,
- bottleneck searching,
- building and development of new network devices and protocols,

- and for ensuring quality of service (QoS) for real-time streaming multimedia applications.

Using this information, network administrators can make the network more efficient.

The simplest tools that measure and capture the packets of network traffic are packet sniffers. Packet sniffers, also known as protocol or network analyzers, are tools that monitor and capture network traffic with all content of network traffic. We can use sniffers to obtain the main information about network traffic, such as packet size, inter-arrival time and the type and structure of IP protocol. Sniffers have become very important and indispensable tools for network administrators. Figure 1 shows traffic captured by a packet sniffer.
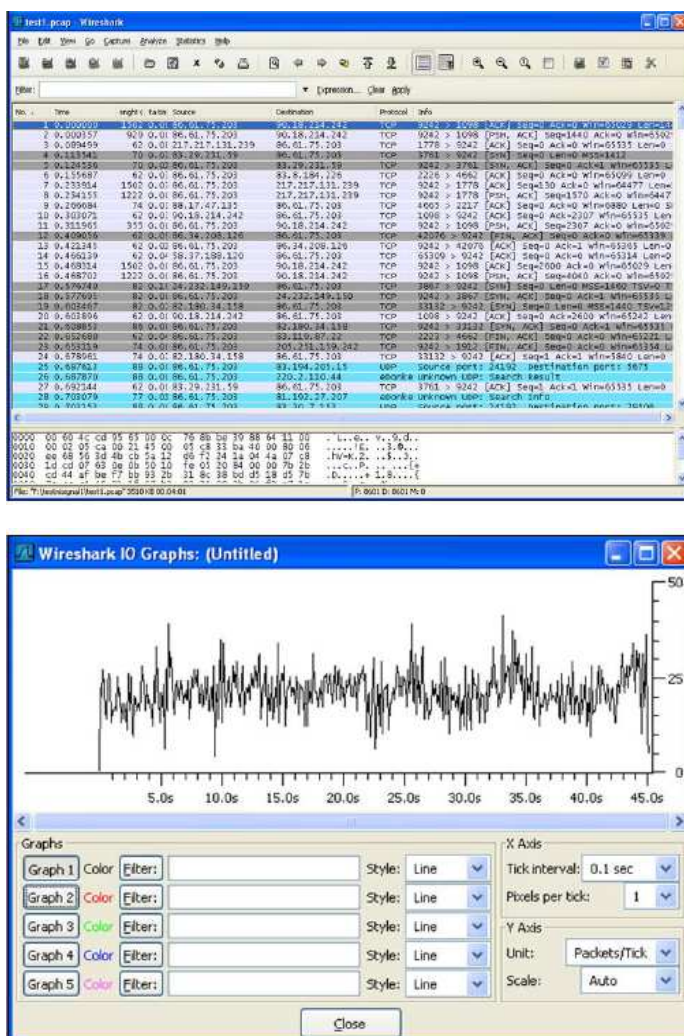

Fig. 1. User interface of WireShark sniffer during the network capturing.

Any sniffers are able to extract this data from the IP headers. Knowing them, it is then simple to calculate a length of IP PDU (Protocol Data Unit), which also contains a header of higher layer protocols. Using an in-depth header analysis, it is possible, in the similar way to the IP header, to calculate the lengths of all these headers.

An analytical description of network traffic does not exist, because we cannot predict the size and arrival time of the next packet. Therefore, we can only describe network traffic as a stochastic process. Hence, we have tried to describe these two stochastic processes (arrival time and packet size) with the use of Hurst parameter and probability distributions.

2.2 Self-similarity

In the 1990s, new descriptions and models of network's traffic were developed, which then replaced the traditional traffic models, such as Poisson and Markov [5], [20]. The Poisson process was widely used in the past, because it gave a good approximation of telephone network (PSNT networks), especially when describing times between each call and call durations. This model is usually described by exponential probability distribution, which is characterized by the parameter λ (number of events per second). However, these models do not allow for descriptions of bursts, which are distinctive in today's network traffic. Such
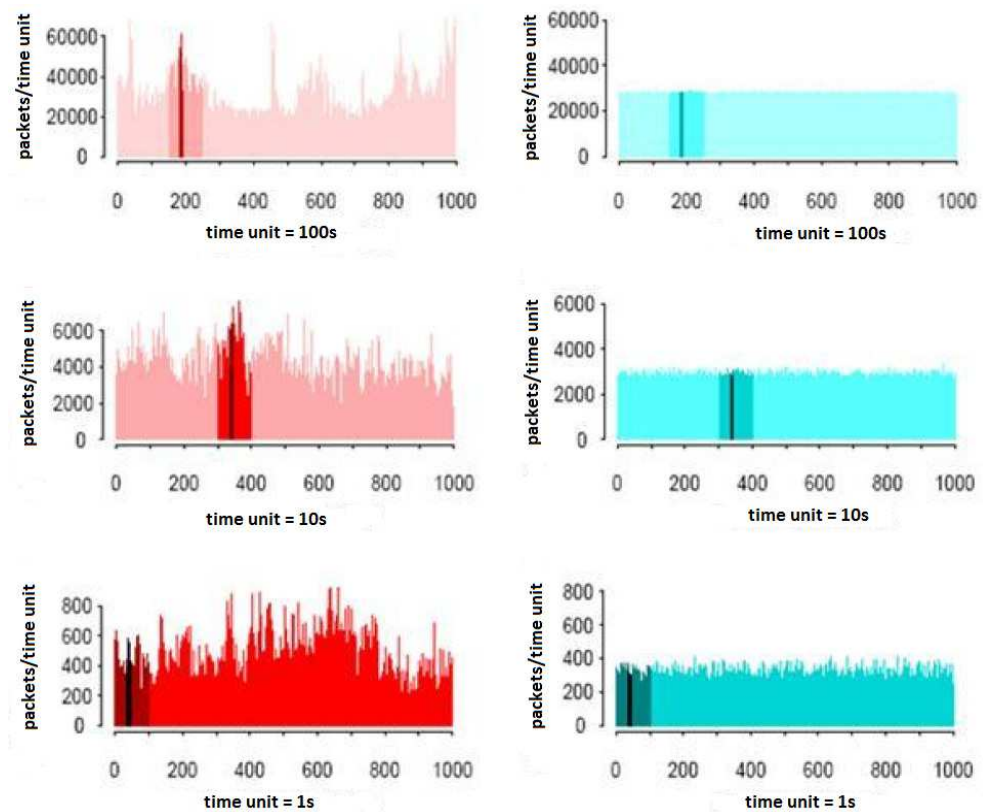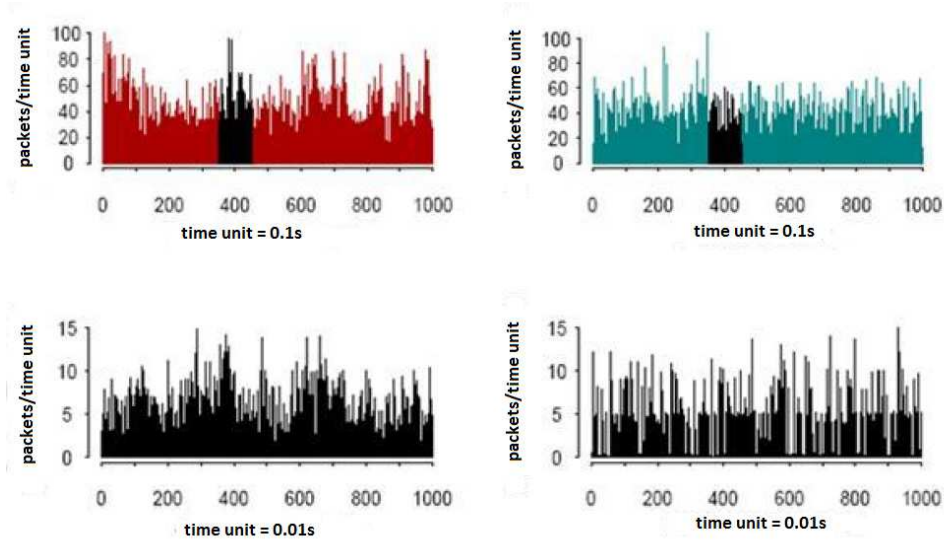
Fig. 2. Comparison of self-similar network traffic (left) and synthetic traffic created by Poisson model (right) on different time scales (100, 10, 1, 0.1 and 0.01s). Self similar traffic contains bursts on all time scales in contrast to the generated synthetic traffic, based on the Poisson model, which tends to average on longer time [1].

bursts can be described by a self-similarity model, because it shows bursts over a wide-range of time scales [1]-[4]. This contrasts the traditional traffic model (Poisson model), which becomes very smooth during the aggregation process.

2.3 Self-similarity

The definition of self-similarity is usually based on fractals for the standard stationary time series [5], [6], [21].

Let $X = (X_t, t = 0, 1, 2,…)$ be a covariance stationary stochastic process; that is a process with a constant mean, finite variance $\sigma^2 = E[(X_t - \mu)2]$, with auto-covariance function $\gamma(k) = E[(X_t - \mu)(X_t+k - \mu)]$, that depends only on k. Then the autocorrelation function $r(k)$ is:

$$r(k) = \frac{\gamma(k)}{\sigma^2} = \frac{E\big[(X_t - \mu)(X_{t+k} - \mu)\big]}{E\big[(X_t - \mu)^2\big]}, \qquad k = 0,1,2,… \tag{1}$$

Assume X has an autocorrelation function, which is asymptotically equal to:

$$r(k) \approx k^{-\beta} L_1(k), \quad k \to \infty, \quad 0 < \beta < 1, \tag{2}$$

where $L_1(k)$ slowly varies at infinity, that is $\lim_{t \to \infty}(L_1(tx) / L_1(t)) = 1$ for all x > 0. Such functions are for example $L_1(t) = const.$ and $L_1(t) = \log(t))$ [5], [6].

The measure of self-similarity is the Hurst parameter (H), which is in a relationship with the parameter β in equation (3).

$$H = 1 - \frac{\beta}{2} \tag{3}$$

Let's define the aggregation process for the time series [5], [6]:

For each m = 1, 2, 3, … let $X^{(m)} = (X_k^{(m)}, k = 1,2,..m)$ denote a new time series obtained by averaging the original series X over a non-overlapping block of size m. That is, for m=1, 2, 3, …, $X^{(m)}$ is given by:

$$X_k^{(m)} = \frac{1}{m}(X_{km-m+1} + ... + X_{km}), \qquad k = 1,2,3, ... \tag{4}$$

$X_k^{(m)}$ is the process with average mean and autocorrelation function $r^{(m)}(k)$ [6].

The process X is called an exactly second order with parameter H, which represents the measure of self-similarity if the corresponding aggregated $X^{(m)}$ has the same correlation structures as X and $\text{var}(X^{(m)}) = \sigma^2 m^{-\beta}$ for all m = 1, 2, … :

$$r^{(m)}(k) = r(k), \text{ for all } m = 1,2, ... \quad k = 1,2, ... \tag{5}$$

The process X is called an asymptotically second order with parameter $H = 1 - \beta/2$, if for all k it is large enough,

$$r^{(m)}(k) \to r(k), \qquad m \to \infty \tag{6}$$

It follows from definitions that the process is the second order self-similar in the exact or asymptotical sense, if their corresponding aggregated process $X^{(m)}$ is the same as X or becomes indistinguishable from X-at least with respect to their autocorrelation function. The most striking property in both cases, exact and asymptotical self-similar processes, is that their aggregated processes $X^{(m)}$ possess a no degenerate correlation structure as $m \to \infty$. This contrasts with the Poisson stochastic models, where their aggregated processes tend to second order pure noise as $m \to \infty$:

$$r^{(m)}(k) \to 0, \quad m \to \infty, \quad k = 0,1,2,... \tag{7}$$

Network traffic with bursts is self-similar, if it shows bursts over many time scales, or it can be also said over a wide-range of time scales. This contrasts with traditional models such as Poisson and Markov, where their aggregation processes become very smooth.

2.4 Long-range dependence

The self-similar process can also contain a property of long-range dependence [5]-[8]. Long range dependence describes the memory effect, where a current value strongly depends upon the past values, of a stochastic process, and it is characterized by its autocorrelation function. This property has a stochastic process, which satisfies relation (2), order with relation $r(k) = \gamma(k)/\sigma^2$.

For $0 < H < 1$, $H \neq 1/2$ it holds [6]

$$r(k) \approx H(2H-1)k^{-2H-2}, \qquad r \to \infty \tag{8}$$

For values $0.5 < H < 1$ autocorrelation function $r(k)$ behavior, in an asymptotic mean, as $ck^{-\beta}$ for values $0 < \beta < 1$, where c is constant $c > 0$, $\beta = 2 - 2H$, and we have:

$$\sum_{k=-\infty}^{\infty} r(k) = \infty . \tag{9}$$

The autocorrelation function decays hyperbolically, as the k increases, which means that autocorrelation function is non-summable. This is opposite to the property of short-range dependence (SRD), where the autocorrelation function decays exponentially and the equation (9) has a finite value. Short and long-range dependence have a common relationship with the value of the Hurst parameter of the self-similar process [6], [21]:

- $0 < H < 0.5 \rightarrow$ SRD - Short Range Dependence
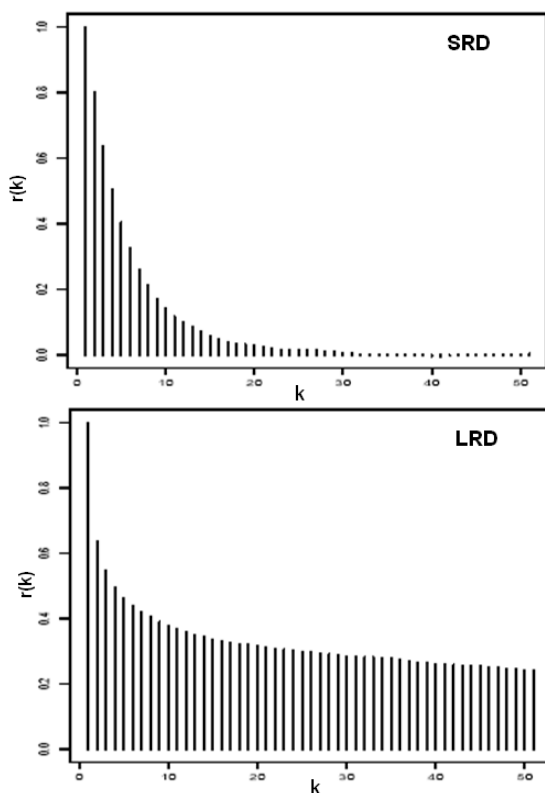- $0.5 < H < 1 \rightarrow$ LRD - Long Range Dependence

Fig. 3. Comparison between autocorrelation function of short range dependence process (left) and autocorrelation function of long range dependence process (right) [15].

2.5 Heavy-tailed distributions

Self-similar processes can be described by heavy-tailed distributions [5], [6], [9]. The main property of heavy-tailed distributions is that they decay hyperbolically, which is opposite to the light-tailed distribution, which decays exponentially. The simplest heavy-tailed distribution is Pareto. The probability density function of Pareto distribution is given by [43]:

$$p(x) = \frac{\alpha k^{\alpha}}{x^{\alpha+1}}, \quad k \le x, \quad \alpha, k > 0 \tag{10}$$

where parameter a represents the shape parameter, and k represents the local parameter of distribution (also a minimum possible positive value of the random variable x).
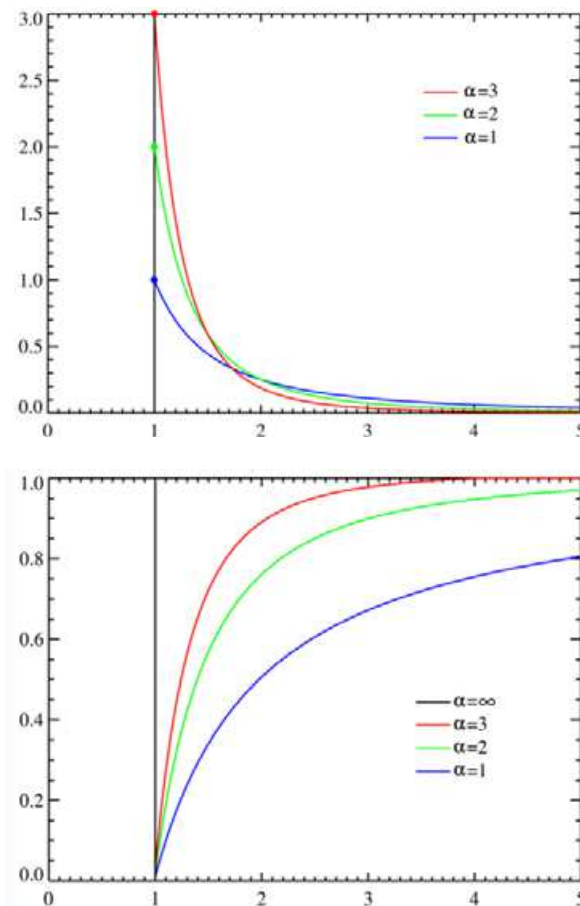

Fig. 4. Probability density function and cumulative distribution function of Pareto distribution for various shape parameters α and constant location parameter k = 1 [43].

Another very important heavy-tailed distribution is Weibull distribution, which is described by [44]:

$$p(x) = \frac{\alpha}{k} \cdot \left(\frac{x}{k}\right)^{\alpha-1} \cdot e^{-(\frac{x}{k})^{\alpha}}, \quad x \geq 0, \quad \alpha, k > 0 \tag{11}$$

where parameter a presents the shape parameter, and k presents the local parameter of distribution.
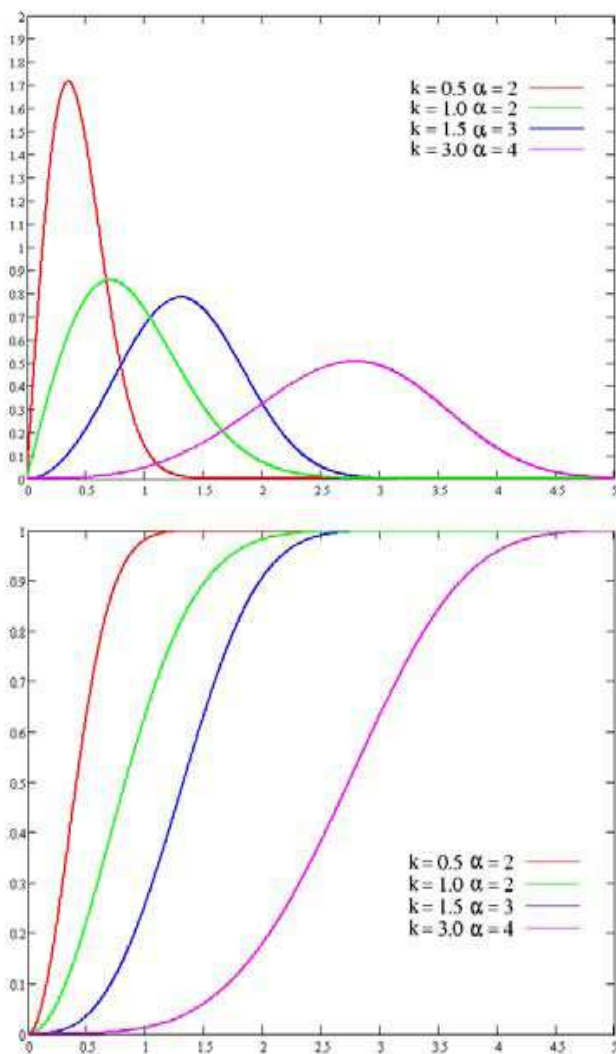

Fig. 5. Probability density function and cumulative distribution function of Weibull distribution for various shape parameters a and constant location parameter k [44].

2.6 Network traffic definitions

The network traffic can be observed on different layers of ISO/OSI model, for that reason we define different kinds of network traffics. The network traffic can be represented as a stochastic process, which can be interpreted as the traffic volume – measured in packets, bytes or bits per time unit, and it is consequent on data or packets, which are sent through the network in time unit. If we observe network traffic on the low level of ISO/OSI model, then define the packet network traffic [45] $Z_p[n]$:

Let define the packet network traffic $Z_p[n]$ as a stochastic process interpreted as the traffic volume, measured in packets per time unit. $Z_p[n]$ can be described as a composite of two stochastic processes:

$$Z_p[n] = X_p[n] \circ Y_p[n], \quad n \in \mathbb{R}.$$
(12)

where $X_p[n]$ represents packet size process and $Y_p[n]$ represents the packet inter-arrival time.

Packet-size process $X_p[n]$ is defined as a series of packet sizes l_{Pi} measured in bits (b) or bytes (B).

$$X_p[n] = \{l_{P1}, l_{P2}, ...l_{Pi}, ..., l_{Pn}\}, \quad 1 \le i \le n$$
(13)

where sizes of packets' l_{Pi} are limited by the shortest l_m and the longest l_{MTU} packet size (MTU - Maximum Transmission Unit).

$$l_m \le l_{Pi} \le l_{MTU}$$
(14)

Packet inter-arrival time process $Y_p[n]$ is defined as a series of times between packet arrivals t_{Pi} (time stamps).

$$Y_p[n] = \{t_{P2} - t_{P1}, ..., t_{Pi} - t_{Pi-i}, ..., t_{Pn} - t_{Pn-i}\}, \quad 1 \le i \le n$$
$$= \{\Delta t_{p1}, \Delta t_{p2}, ..., \Delta t_{pi}, ..., \Delta t_{pn-1}\}, \quad 1 \le i \le n$$
(15)

The measured network traffic is packet network traffic, which can be captured using special software program or hardware devices. For that reason, the measured network traffic is marked as $Z_{pm}[n]$. We also define modeled (simulated) network traffic as $Z_{ps}[n]$. We suppose, that the measured and modeled traffic is statistically equal, denoted by the symbol $\approx$,

$$Z_{pm}[n] \approx Z_{ps}[n]$$
(16)

if there are also statistical equalities between a packet size and inter-arrival time processes of measured, and modeled traffic.

$$X_{pm}[n] \approx X_{ps}[n]$$

and

$$Y_{pm}[n] \approx Y_{ps}[n]$$

(17)

Let's define network traffic on higher layers (application) of ISO/OSI model. Data source network traffic $Z_d[n]$ can be described as a composite of data source lengths $X_d[n]$ and data inter-arrival times $Y_d[n]$ processes:

$$Z_d[n] = X_d[n] \circ Y_d[n], \quad n \in \mathbb{R}$$

(18)

To provide statistical equality between packet network traffic $Z_p[n]$ and data sources network traffic $Z_d[n]$, we have performed a transformation between packet size process $X_p[n]$ and the process of data length $X_d[n]$ as well as transformation between packet inter-arrival time $Y_p[n]$ and data inter-arrival time $Y_d[n]$.

$$X_{pm}[n] \xrightleftharpoons{transformation} X_d[n]$$

(19)

$$Y_{pm}[n] \xrightleftharpoons{transformation} Y_d[n]$$

(20)

Transformation (19) and (20) allows estimation of packet traffic processes from data source traffic processes or vice verse.

3. Network traffic analysis and modeling

3.1 Hurst parameter estimations

Hurst's parameter represents the measure of self-similarity. There are several methods for estimating Hurst's parameter (H) [1]-[4] of stochastic self-similar processes. However, there are no criteria as to which method gives the best results. There are several different methods for estimating the Hurst parameter which can lead to diverse results [9], [10]. This is the reason why Hurst's parameter cannot be calculating but can be estimated. The most often used methods for Hurst's parameter estimation are [6], [8], [21]:

- Variance method is a graphical method, which is based on the property of slowly decaying variance. In a log-log scale plot, a sample variance versus a non-overlapping block of size m is drawn for each aggregation level. From the line with slope β we can estimate Hurst's parameter as a relationship, from equation (3).
- R/S method is also a graphical method. It is based on a range of partial sums regarding data series deviations from mean value, rescaled by its standard deviation. The slope in the log-log plot of the R/S statistic versus aggregated points is the estimation for Hurst's parameter.
- Periodogram method plots spectral density in a logarithm scale versus frequency (also in logarithm scale). The slope in periodogram allows the estimation of parameter H.

Figure 6 presents an example of test traffic and estimations of Hurst's parameter through different methods.
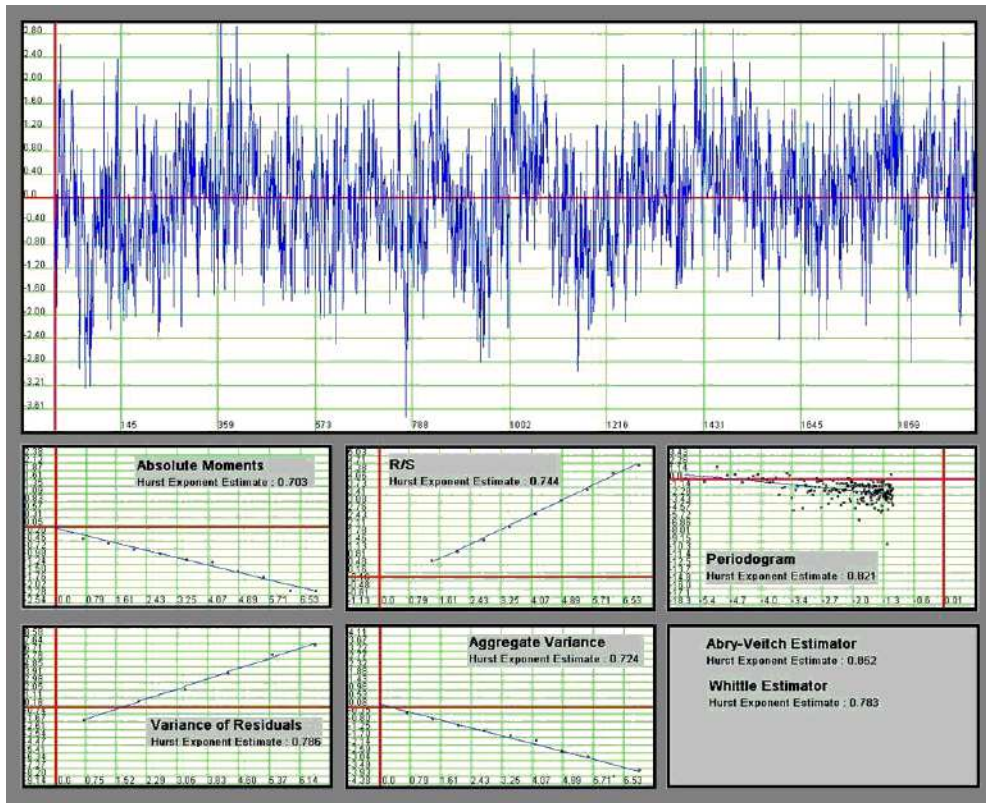
Fig. 6. Estimating parameter *H* for self-similar traffic (upper-left) with the variances method (lower left), R/S method (upper-right) and periodogram method (lower-right) using SELFIS tool [8].

3.2 Distribution parameter estimation for stochastic process of network traffic

Network traffic can be described by two stochastic processes, one for packet/data sizes and one for packet/data inter-arrival time. All processes are usually described by probability distributions. Self-similar process can be described by heavy tailed distributions. The main task for modeling the stochastic process with probability distribution is to choose the right distribution, which would be a good representation of our network traffic stochastic process. The statistic distribution parameters of data sources are then estimated by fitting tools [9], [25], [26] or other known methods, such as CCDF [6] or Hill estimator [17], [18]. Mathematical fitting tools are used (EasyFit), which allow us to automatically include the fit distribution of the stochastic process, and also estimate parameters of distribution from the captured traffic [9], [29].
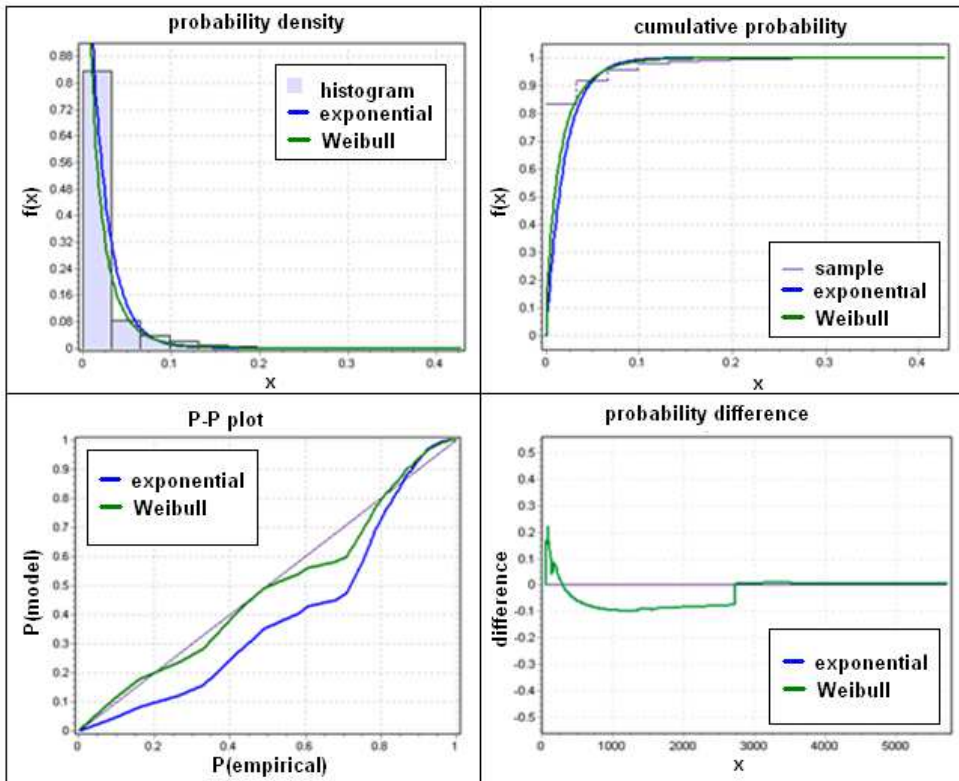
Fig. 7. For the stochastic process of inter-arrival time, distribution and estimate parameters of these distributions are chosen based on the histogram (upper left), and cumulative distribution function (upper right). Differences between empirical and theoretical distributions in P-P plot (lower left), and deferential distribution (lower right).

4. Simulation of network traffic in simulation tools

One of the very important tasks in simulation is modeling the real network parameters and network elements for simulation purposes. The main goal in successful modeling of network traffic is to minimize discrepancies between the measured simulations and by simulations statistically-modeled and generated traffic. This means, that both traffics are similar within the different criteria, such as bit and packet-rate, bursts (Hurst's parameter), variance, etc.

Network traffic simulations are usually based on modeling of data sources or applications. One of the most known simulation tools is OPNET Modeler [22], [23]. A simulation of network traffic in this tool is based on the "on/off" models [41] or more often used traffic generators. Difference between these manners is in a modeling manner. In the first case, the arrival process is described by Hurst's parameter (*H*) and the data length process is

described by probability density function (*pdf*). In the second case, processes of data length and data inter-arrival time are both described by *pdf*.

In OPNET Modeler, two standard node models appear [9]:

- Raw Packet Generator (RPG)
- IP station

Raw Packet Generator (RPG) is a traffic source model [16], [27] implemented specially to generate self-similar traffic, which is based on different fractal point processes (FPP) [41]. Self similar traffic is modeled with an arrival process, which is described by Hurst's parameter and the distribution probability for packet sizes. This arrival process can be based on many different parameters, such as Hurst parameter, average arrival rate, fractal onset time scale, source activity ratio and peak to mean ratio [16]. There are several different fractal point processes (FPP). In our case, we used the superposition of the fractal renewal process (Sub-FRP) model, which is defined as the superposition of *M* independent and probably identical renewal fractal processes. Each FRP stream is a point renewal processes and *M* numbers of independent sources compose the Sub-FRP model. Common inter-arrival probability density function *p(t)* of this process is:

$$p(t) = \begin{cases} \gamma A^{-1} e^{-\gamma t/A} & 0 \le t \le A \\ \gamma e^{-\gamma} A^{\gamma} t^{-(\gamma+1)} & t \ge A \end{cases} \tag{21}$$

where $1 < \gamma < 2$. Process FRP can be defined as Sup-FRP process, when the number of independent identical renewal processes *(M)* is equal to 1. A model Sub-FRP is described by three parameters: γ, A and M. γ represents the fractal exponent, A is the location parameter, and M is the number of sources. These three parameters are in relationship with three OPNET parameters. These parameters are Hurst's average arrival-rate λ, and fractal onset time-scale (FOTS). The relationships between these three parameters of Sub-FRP and parameters in OPNET model are:

$$H = (3 - \gamma) / 2$$

$$\lambda = M\gamma[1 + (\gamma-1)^{-1} e^{-\gamma}]^{-1} A^{-1} \tag{22}$$

$$T^{\alpha} = 2^{-1} \gamma^{-2} e^{-\gamma} (\gamma-1)^{-1} (2-\gamma)(3-\gamma)[1+(\gamma-1)e^{\gamma}]^2 A^{\alpha},$$

where $\gamma = 2 - \beta$. Hurst parameter H is defined by equation (3). In the Sub-FRP model from OPNET, we can set Hurst's parameter (H), average arrival-rate (λ) and fractal onset time-scale (FOTS) in seconds. The recommended value for the parameter FOTS in OPNET is 1 second.

The IP station [16] can contain an arbitrary number of independent simultaneous working-traffic generators. Each generator enables the use of heavy-tailed distributions, such as Pareto or Weibull, for the generation of a self-similar network traffic by two distributions, one for length of a data source process and another for data inter-arrival time process. In our research, a traffic generator contained in an Ethernet IP station model of the OPNET Modeler simulation tool is used, as shown in the Figure 8.
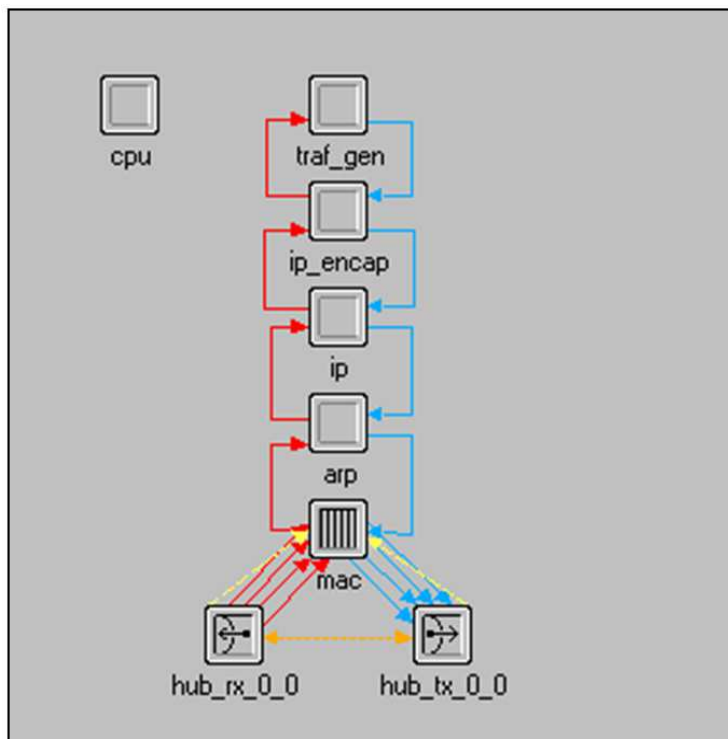
Fig. 8. Node model for used IP station in simulation.

In the IP station model, the traffic generator is placed above the IP encapsulation layer, which takes care of packets' formations and fragmentation. This is the process of segmentation of long data into the shorter packets, or vice versa, according to the RFC 793 [12]. Padding of the packet data payload with additional bits is also performed when data is shorter than a predefined minimal payload. Because the traffic is modeled, above IP level of the TCP/IP model, to the lengths of the generated data, 20 bytes of IP header are added. 18 bytes of information for MAC (14bytes) and CRC (4 bytes) are also further added. Structure of Ethernet frame used in the IP station model. Using this model, the applications' protocol does not impact the generated traffic. The model is suitable for the simulation cases, when we want to statistically model the network traffic, which can be caused by many arbitrary communications' applications. Using this approach, we can model such network traffic by single traffic source.

5. Estimation of simulation parameters of measured network traffic

The main problem of measured packet network traffic modeling is to estimate the parameter, which is needed for modeling measured network traffic in simulation tools. It has already been mentioned that the parameters of data source traffic processes are needed. We already described that transformation from packet network traffic $Z_p[n]$ to data source

network traffic $Z_d[n]$ is needed (section 2.6) [45]. There are many possibilities to make a transformation from $Z_p[n]$ to $Z_d[n]$, which allows estimation of parameters of data source network traffic processes. We investigated two algorithms [28]:

1.   algorithm with an in-depth analysis of all packet headers,
2.   algorithm with a coarse inspection of IP header only.

The main differences between them are complexity and the needed execution time. The first algorithm mimics a complete decapsulation process, and defragmentation in higher layers of the communication model. Any sniffers are able to extract this data from the IP header. Knowing them, it is then simple to calculate a length of IP PDU (Protocol Data Unit) which also contains a header of higher layer protocols. Through the use of an in-depth header analysis, it is possible, in the similar way as the IP header, to calculate the lengths of all these headers. Each packed IP header has four the so-called fragmentation fields that contain information about data fragmentation, which is shown on Figure 9.
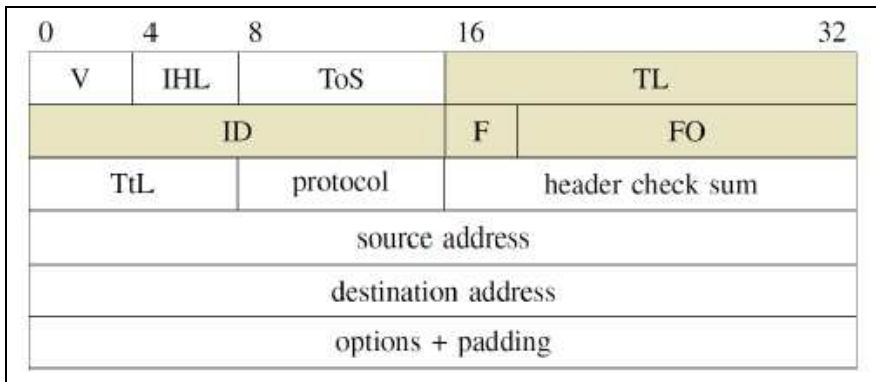

Fig. 9. IP header. Shadowed fields are used in the defragmentation process. Legend:
V: protocol version; IHL: Internet Header Length; ToS: Type of Service; TL: Total Length;
ID: Identification Data; F: Flags; FO: Fragment Offset; TTL: Time to Live.

Extensive research and investigation about traffic sources in contemporary networks show that this approach requires an in-depth analysis of packets (where need specialized, very powerful and consequently, expensive instruments), which in case of encrypted packets and non-standard application protocols, is not completely possible. In such cases, it is also necessary to capture the entire packets, which can be problematic in the high-speed networks. For these reasons, a simple algorithm has been developed, where only information of packets sizes, packet time stamps and IP addresses are needed.

The second algorithm skips decapsulation by considering the average lengths of packet headers and then uses only packet lengths and inter-arrival times. In the second case, the algorithm offers the estimation of data source network traffic, not the exact reconstructed data source traffic. The second algorithm represents the main part of method by mimic defragmentation process, which is described in detail in [45]. The main idea of mimic defragmentation process method is to compose data from the captured packet traffic, which is previously fragmented at the transmitter. The data source traffic estimation is

carried out by finding and summing fragmented packets' sequences without an in-depth analysis of packets. Fragmented sequence is defined as a sequence of l_{MTU} sized packets associated with the same source and destination addresses and terminated by packet shorter than l_{MTU}.

6. Simulation results

In real networks, we have captured packets of different network traffic through a Wireshark sniffer. The two different types of measured traffic are used for analysis, modeling and simulation purposes. These two test traffics are shown in Figure 10.
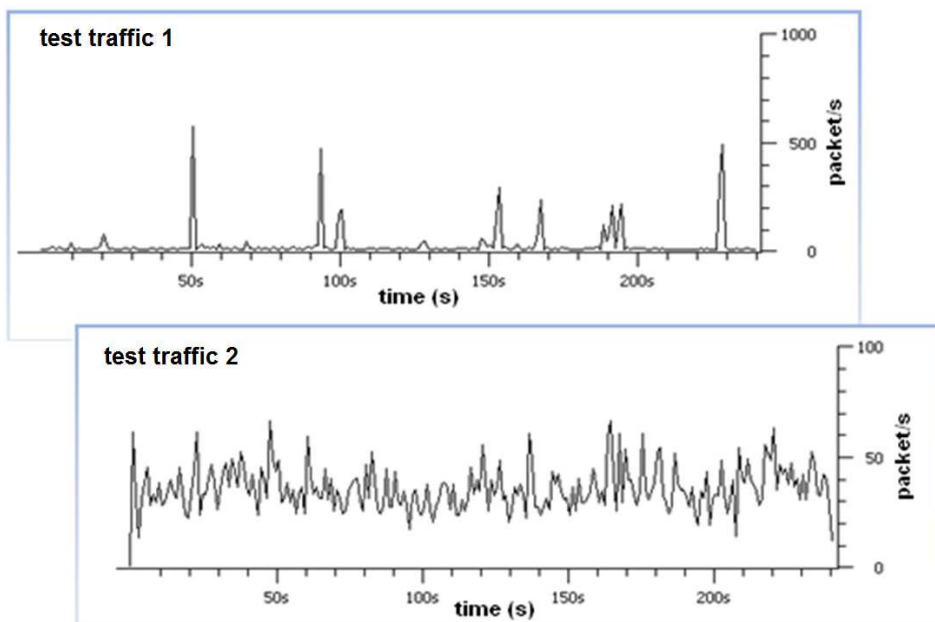

Fig. 10. Measured test traffic 1 and 2 captured by Wireshark sniffer.

measured test traffics	packet rate (p/s)	bit rate (kb/s)	variance method	R/S method	periodogram method
test traffic 1	24.02	108.90	0.630	0.723	0.843
test traffic 2	35.612	114.51	0.592	0.580	0.477

Table 1. The main properties of captured traffics. On the right side of the table the Hurst parameter is estimated using different methods for both test traffics.

For each of test traffics, the Hurst parameter has been estimated through different methods. The Hurst parameters for both cases are bigger than 0.5, so we can classify these test traffics

as a self-similar network traffic. Table 1 contains the estimated parameters *H* for both traffics, which are estimated by variance, R/S and periodogram methods. We also conducted tests about short and long-range dependence. In the case of the first test traffic, the autocorrelation function decayed hyperbolically, which means, that this traffic can have the property of a long-range dependence. For the second test traffic autocorrelation, function decayed exponentially towards 0. For this case, the sum of autocorrelations has finite results and, therefore, the test traffic 2 has the property of short-range dependence.

For both test traffics (test traffic 1 and test traffic 2) we estimate distribution and its parameters for data source traffic processes for simulation purpose. For that reason, we made an estimation of data source traffic from the captured packet traffic through the mimic defragmentation process method [45]. For both test traffics, the suitably heavy (Pareto or Weibull) and also light-tailed (exponential) distributions are chosen.

Based on the estimated distribution parameters for both measured test traffic (test traffic 1 and test traffic 2), we generated self-similar traffic in the OPNET simulation tool with two different station types – RPG and IP stations. We have created six different scenarios for each of test traffic. In the first two scenarios, the network traffic is generated by an RPG station, where a self-similarity is described by Hurst parameter. During the first scenario, we use heavy-tailed distribution for the data size process, while in the second a light-tailed distribution (exponential) is used. In the next four scenarios, network traffic is generated using the IP station, where we use different combination's distributions for the data size process and data inter-arrival time. One of the criterions, for successful modeling, is the difference between bit and packet-rates of the test traffic and modeled traffic in OPNET simulation tool. Besides the average values of bit and packet-rates, the more important criteria are also bursts' intensity within the network traffic. For each of test traffics (test traffic 1 and test traffic 2), the traffic which best represents the measured test traffic is chosen from six modeled traffics.

Test traffic 1 poses the property of long-range dependence, so there are a lot of bursts in the traffic. We model this measured-test traffic over six different scenarios. The results are shown in Figure 6 and Table 2. Table 2 shows the main properties of measured test traffic 1 and estimated distribution parameters which were used in OPNET simulation tool for simulating network traffic (the left side of Table 2). Table 2 (the right side) also shows main properties of simulated network traffics (six different scenarios) in OPNET simulation tool based on estimated distributions.

Table 2 shows modeling results for test traffic 1 over six different scenarios in OPNET simulation tool. There are estimated statistical parameters such as Hurst parameters and distributions used in models and simulation results using these models. Figure 11 shows all six modeled traffic traffics generated by OPNET, with estimated distributions and parameters from Table 2.

The best approximation for test traffic 1 is modeled traffic 5 from Table 2, which is described by Pareto distribution for data size process and Weibull distribution for data inter-arrival time. Figure 12 shows a comparison between the second test traffic and the modeled traffic 5 for bit rates. From all critera after comparison, we can say that the modeled traffic 5 is a good approximation of measured test traffic 1.
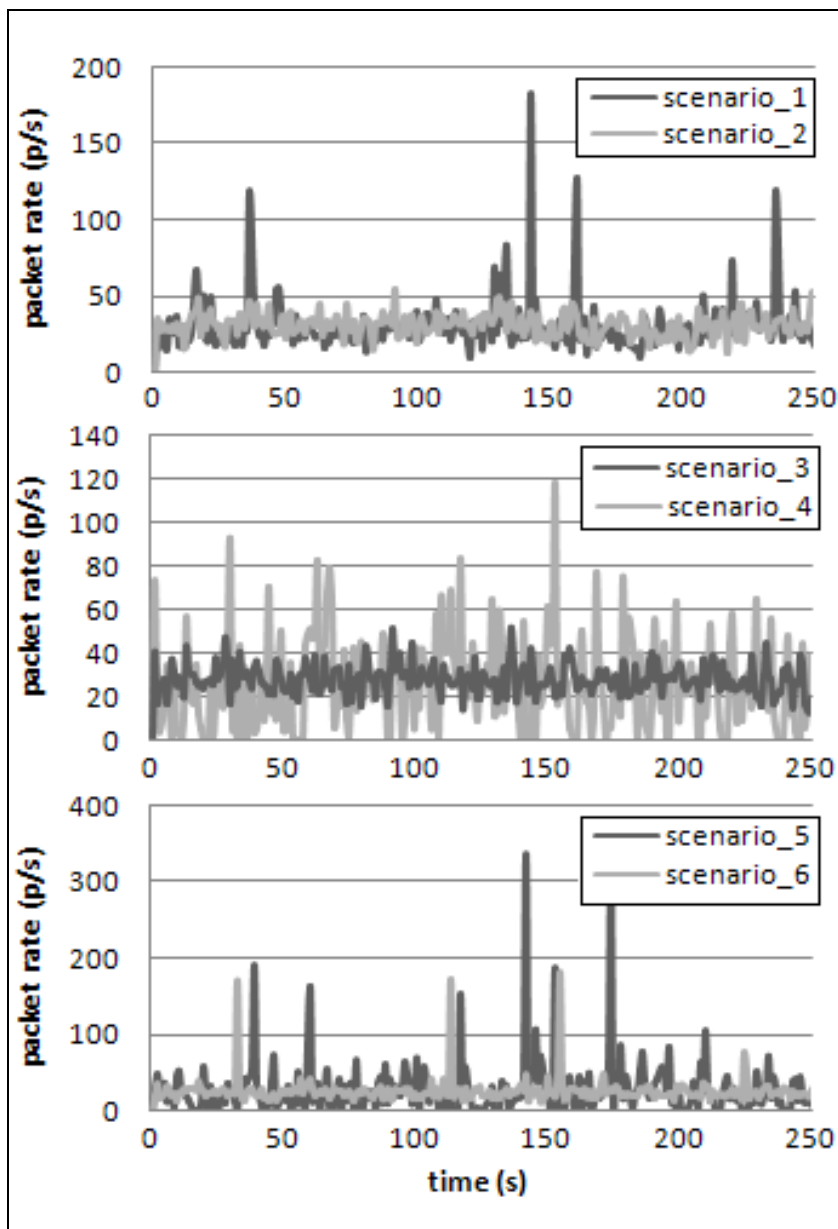
Fig. 11. Modeling measured test traffic 1 in OPNET simulation tool with six different estimated parameters from Table 2 (scenario 1 and 2 with RPG station, scenario 3, 4, 5, 6 with IP station).

parameters for modeling			parameters of measured and modeled traffic in OPNET		
traffic	data inter-arrival process	data size process	packet rate (p/s)	bite rate (kb/s)	H
measured test traffic 1	X	X	24	108.90	0.73
modeled 1	$H = 0.732$	Pareto $a = 0.9835$ $\beta = 432$	33.82	128.75	0.59
modeled 2	$H = 0.732$	exponential $\lambda = 7547.2$	29.18	181.44	0.59
modeled 3	exponential $\lambda = 0.0458$	exponential $\lambda = 933.4$	27.56	168.94	0.51
modeled 4	Weibull $a = 0.304$ $\beta = 0.00578$	exponential $\lambda = 933.4$	25.14	153.71	0.62
modeled 5	Weibull $a = 0.304$ $\beta = 0.00578$	Pareto $a = 0.9835$ $\beta = 34$	25.32	88.70	0.66
modeled 6	exponential $\lambda = 0.0458$	Pareto $a = 0.9835$ $\beta = 34$	26.63	81.30	0.55

Table 2. The left side of table shows the estimated distributions and parameters for measured test traffic 1 (six different distribution combinations). The right side of table shows main properties of modeled network traffic in OPNET simulation tool (six scenarios), where estimated distributions were used.
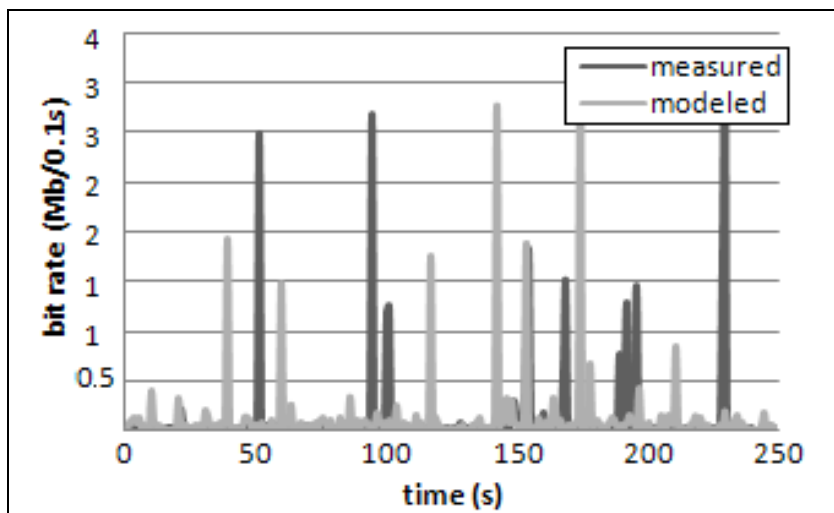

Fig. 12. Comparison between the modeled traffic 5 generated in OPNET simulation tool and the measured test traffic 1 in bits per second (kb/s).

Test traffic 2 is also modeled over six different scenarios, just like in the first case. Table 3 shows the main properties of measured test traffic 2 and estimated distribution parameters which were used in OPNET simulation tool for simulating network traffic (left side of Table 3). Table 3 (right side) also shows main properties of simulated network traffics (six different scenarios) in OPNET simulation tool.

As the best modeled traffic of test traffic 2 from all six cases (Table 3), we choose the case where simulated traffic is described by the exponential distribution for packet sizes and Weibull heavy-tailed distribution for inter-arrival time (modeled traffic 4). The bit-rate of this traffic is 33.27 (p/s) and packet-rate is 126.79 (kb/s), which are very close to the measured values. The Hurst parameter of the simulated traffic is 0.58, which is also close to the estimated values of the measured traffic. Figure 13 shows the comparison between the measured test traffic 2 and the best-modeled traffic (modeled traffic 4) for bit rates. From all critera after comparison, we can say that the simulated traffic is a good approximation of the measured traffic 2.

parameters for modeling			parameters of measured and modeled traffic in OPNET		
traffic	data inter-arrival process	data size process	packet rate (p/s)	bite rate (kb/s)	H
measured test traffic 2	X	X	35.61	114.51	0.55
modeled 1	$H = 0.55$	Pareto $a = 0.8373$ $\beta = 272$	49.46	231.98	0.62
modeled 2	$H = 0.55$	exponential $\lambda = 3619$	36.66	140.72	0.58
modeled 3	exponential $\lambda = 0.029$	exponential $\lambda = 452.48$	35.66	135.89	0.53
modeled 4	Weibull $a = 0.57$ $\beta = 0.01894$	exponential $\lambda = 452.48$	33.27	126.79	0.58
modeled 5	Weibull $a = 0.57$ $\beta = 0.01894$	Pareto $a = 0.8373$ $\beta = 34$	52.27	298.25	0.62
modeled 6	exponential $\lambda = 0.029$	Pareto $a = 0.8373$ $\beta = 34$	55.12	315.61	0.53

Table 3. The left side of table shows the estimated distributions and parameters for measured test traffic 2 (six different distribution combinations). The right side of table shows main properties of modeled network traffic in OPNET simulation tool (six scenarios), where estimated distributions and its parameters were used.
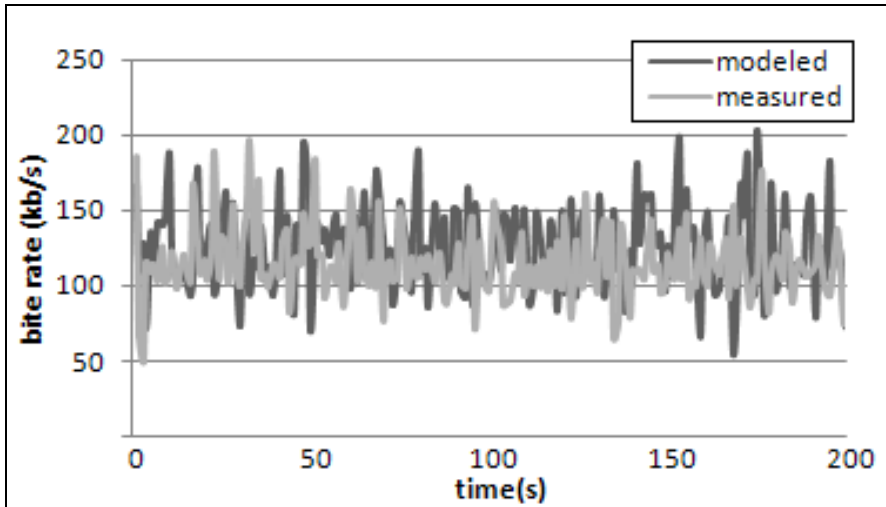
Fig. 13. Comparison between modeled traffic 4 generated in OPNET simulation tool and measured test traffic 2 in bits per second (kb/s).

7. Conclusion

In this chapter, we present our research in the area of measurements, modeling and simulations of the self-similar network traffic. Firstly, the state of the art method for modeling and simulating of self-similar network traffic is presented. We also describe a number of facts about self-similarity, long range dependences and probability, which are used to describe such stochastic processes. Described as well are the mechanism and models to simulate network traffic in the OPNET Modeler simulation tool. The main goal of our research is to simulate measured network traffic, where we tend to minimize discrepancies between the measured and the simulated network traffic in the sense of packet-rate, bit-rate, bursts intensity, and variances. One of the big challenges in our research work was to find appropriate method to estimate parameters of data source network traffic processes that are based on measured network packet's traffic. The estimated parameters are needed during the modeling of the measured network traffic in the simulation tool. For those reasons, we have developed different methods, which allow estimation of the parameters of data source network traffic processes, based on the measured network packet's traffic.

At the end of the chapter, all phases needed for simulating the measured network traffic in the OPNET simulation tool are presented. During the analysis phase we pay attention to the self-similar property, which has become the basic model for describing today's network traffic. In the network traffic theory, the properties of short and long-range dependence are directly prescribed by the values of estimated parameter H. In our network traffic analysis, we prove that network traffic (test traffic 2) can exist where Hurst parameter is bigger than 0.5, but this process does not have the property of a long-range dependence.

For the purpose of parameters estimation of data source network traffic processes, we have used a method that mimics packet defragmentation. Through the use of this method we

offer estimated parameters, used in simulations, where six traffics are simulated by different distributions for each of the measured test traffic. It can be seen from simulations that in the case of modeling self-similar traffic, short-range dependence is more appropriate for choosing exponential distribution to describe a packet-size process. The exponential distribution does not impact the extreme peaks in the modeled traffic. Pareto distribution is unsuitable for this purpose.

Heavy-tailed distributions, especially Pareto, are suitable for modeling a packet-size process of the measured network traffic, which are self-similar and also have the property of a long-range dependence (test traffic 1).

There are discrepancies between the measured and the modeled traffics in the sense of packet-rate, bit-rate, bursts intensity, and variances. With a method which mimics defragmentation, a good approximation of the measured network traffic is obtained. We cannot claim that this is the optimal method for all situations, because there are some limitations, although it shows good results through simulation in OPNET Modeler. We have noticed that estimating the shape-parameter of Pareto is very delicate, because a small deviation in the parameter causes large discrepancies regarding the network traffic's average values, which is one of the important criteria for traffic modeling.

8. Acknowledgment

9. 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] W. Willinger and V. Paxson, Where mathematics meets the Internet, Notices of the American Mathematical Society, 45(8): 961–970, 1998.
[3] K. Park, G. Kim and M. E. Crovella, On the Relationship Between File Sizes Transport Protocols, and Self-Similar Network Traffic, International Conference on Network Protocols, 171–180, Oct 1996.
[4] M. E. Crovella and A. Bestavros, Self-Similarity in World Wide Web Traffic Evidence and Possible Causes, IEEE/ACM Transactions on Networking, 1997.
[5] O. Sheluhin, S. Smolskiy and A. Osin, Self-Similar Processes in Telecommunications, John Wiley & Sons, 2007.
[6] K. Park and W. Willinger, Self-Similar Network Traffic and Performance Evaluation, John Wiley & Sons, 2000.
[7] T. Karagiannis, M. Molle and M. Faloutos, Understanding the limitations of estimation methods for long-range dependence, University of California.
[8] T. Karagiannis and M. Faloutos, Selfis: A tool for self-similarity and long range dependence analysis, University of California.

[9] M. Fras, J. Mohorko and Ž. Čucej, Estimating the parameters of measured self similar traffic for modeling in OPNET, IWSSIP Conference, 27.-30 June 2007, Maribor, Slovenia.

[10] J. Mohorko and M. Fras, Modeling of IRIS Replication Mechanism in a Tactical Communication network, using OPNET, Computer Networks, v 53, n 7, p 1125-36, 13 May 2009.

[11] J. Mohorko, M. Fras and Ž. Čucej: Modeling of IRIS replication mechanism in tactical communication network with OPNET, OPNETWORK 2007 - the eleventh annual OPNET technology Conference, August 27th-31st, Washington, D.C., 2007.

[12] RFC 793 - Transmission Control Protocol:. [Online]. Available: http://www.faqs.org/rfcs/rfc793.html

[13] M. Chakravarti, R. G. Laha and J. Roy, Handbook of Methods of Applied Statistics, Volume I, John Wiley and Sons, pp. 392-394, 1967.

[14] W. T. Eadie, D. Drijard, F. E. James, M. Roos and B. Sadoulet, Statistical Methods in Experimental Physics, Amsterdam, North-Holland, 269-271, 1971.

[15] A. Adas, Traffic Models in Broadband Telecommunication Networks, Communications Magazine, IEEE , vol 35/7, 82–89, 1997.

[16] J. Potemans, B. Van den Broeck, Y. Guan, J. Theunis, E. Van Lil and A. Van de Capelle, Implementation of an Advanced Traffic Model in OPNET Modeler, OPNETWORK 2003, Washington D.C., USA, 2003.

[17] B. Hill, A Simple Approach to Inference About tbc Tail of a Distribution, Annals of Statistics, Vol. 3, No. 5, 1975, pp.1163-1174.

[18] J. Judge, H. W. Beadle and J. Chicharo, Sampling HTTP response packets for prediction of web traffic volume statistics, IEEE Global Communications Conference (GLOBECOM'98), Sydney, Australia, Nov. 8-12, 1998.

[19] K. Park, G. Kim and M. E. Crovella, On the Relationship Between File Sizes Transport Protocols, and Self-Similar Network Traffic, International Conference on Network Protocols, 171–180, Oct 1996.

[20] V. Paxon and S. Floyd, Wide area traffic: the failure of Poisson modeling, IEEE/ACM Transactions on Networking, 3(3): 226–244, 1995.

[21] H. Yõlmaz, IP over DVB: Managment of self-similarity, Master of Science, Boğaziçi University, 2002.

[22] B. Vujičić, Modeling and Characterization of Traffic in Public Safety Wireless Networks, Master of Applied science, Simon Fraser University, Vancouver, 2006.

[23] M. Jiang, S. Hardy in Lj. Trajkovic, Simulating CDPD networks using OPNET, OPNETWORK 2000, Washington D.C., August 2000.

[24] J. Mohorko, M. Fras and Ž. Čučej, Modeling methods in OPNET simulations of tactical command and control information systems, IWSSIP Conference, 27.-30 June 2007, Maribor, Slovenia.

[25] A. M. Law and M. G. McComas, How the Expertfit distribution fitting software can make simulation models more valid, Proceedings of the 2001 Winter Simulation Conference.

[26] Free (demo) fitting tool EasyFit software [Online]. Available: www.mathwave.com/.

[27] F. Xue and S. J. Ben Yoo, On the Generation and Shaping Self-similar Traffic in Optical Packet-switched Networks, OPNETWORK 2002, Washington D.C., USA, 2002.

[28] Ž. Čučej and M.Fras, Data source statistics modeling based on measured packet traffic : a case study of protocol algorithm and analytical transformation approach, TELSIKS 2009, 9th International Conference on Telecommunications in Modern Satellite, Cable and Broadcasting Services, Serbia, Niš, 7-9 October, 2009.

[29] M. Fras, J. Mohorko and Ž. Čučej, Analysis, modeling and simulation of P2P file sharing traffic impact on networks' performances. Inf. MIDEM, 38(2):117–123, 2008.

[30] H. Abrahamsson, Traffic measurement and analysis, Swedish Institute of Computer Science, 1999.

[31] C. Williamson, Internet traffic measurement, IEEE internet computing, vol. 5, no. 6, pp. 70–74, 2001.

[32] P. Celeda, High-speed network traffic acquisition for agent systems, in Proc. IEEE/WIC/ACM International Conference on High-Speed Network Traffic Acquisition for Agent Systems, Intelligent Agent Technology, November 2-5, 2007, pp. 477–480.

[33] D. Pezaros, Network Traffic Measurement for the Next Generation Internet. Computing Department Lancaster University, 2005.

[34] D. Epema, J. Pouwelse, P. Garbacki and H. Sips, The bittorrent P2P filesharing system: Measurements and analysis. Peer-to-Peer Systems IV, 2005.

[35] S. Saroiu, P. K. Gummadi and S. D. Gribble, A Measurement Study of Peer-to-Peer File Sharing Systems, in Proc. of the Multimedia Computing and Networking (MMCN), January 2-5, San Jose, Ca, USA, 2002.

[36] E. Asensio, J. M. Orduna and P. Morillo, Analyzing the Network Traffic Requirements of Multiplayer Online Games, in Proc. 2nd International Conference on Advanced Engineering Computing and Applications in Sciences: ADVCOMP'08, 2008, pp. 229–234.

[37] Y. Yu, D. Liu, J. Li and C. Shen, Traffic Identification and Overlay Measurement of Skype, in Proc. International Conference on Computational Intelligence and Security, November 3-6, vol. 2, 2006, p. 1043 – 1048.

[38] M. E. Crovella and L. Lipsky, Long-lasting transient conditions in simulations with heavy-tailed workloads, in Proc. 1997 Winter Simulation Conference, December 7-10, vol. Atlanta, GA, USA, Edmonton, Canada, 1997.

[39] A. Feldmann, A. C. Gilbert, P. Huang and W. Willinger, Dynamics of IP traffic: a study of the role of variability and the impact of control, in Proc. Applications, technologies, architectures, and protocols for computer communication, August 30-September 03, Cambridge, Massachusetts, USA, 1999, pp. 301–313.

[40] C. Nuzman, I. Saniee, W. Sweldens and A. Weiss, A compound model for TCP connection arrivals for LAN and WAN applications, Computer Networks: The International Journal of Computer and Telecommunications Networking, vol. 40, no. 3, pp. 319–337, 2002.

[41] B. Ryu and S. Lowen. Fractal Traffic Model for Internet Simulation. In Proc. 5th IEEE Symposium on Computers and Communications (ISCC 2000), 2000.

[42] M. Fras, Methods for the statistical modeling of measured network traffic for simulation purposes, Ph.D. thesis, 2009, Maribor, Slovenia.

[43] http://en.wikipedia.org/wiki/Pareto_distribution.

[44] http://en.wikipedia.org/wiki/Weibull_distribution.

[45] M. Fras, J. Mohorko and Ž. Čučej, Modeling of captured network traffic by the mimic defragmentation process, Simulation: Transactions of The Society for Modeling and Simulation International, San Diego, USA, Published online 20 September 2010.

[46] M. Fras, J. Mohorko and Ž. Čučej, Modeling of measured self-similar network traffic in OPNET simulation tool, Inf. MIDEM, 40(3): 224-231, September 2010.
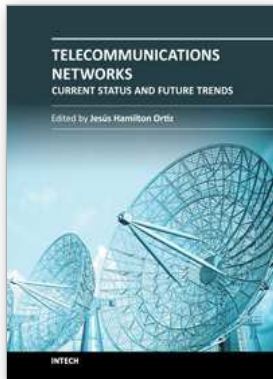
Telecommunications Networks - Current Status and Future Trends

Edited by Dr. Jesús Ortiz

This book guides readers through the basics of rapidly emerging networks to more advanced concepts and future expectations of Telecommunications Networks. It identifies and examines the most pressing research issues in Telecommunications and it contains chapters written by leading researchers, academics and industry professionals. Telecommunications Networks - Current Status and Future Trends covers surveys of recent publications that investigate key areas of interest such as: IMS, eTOM, 3G/4G, optimization problems, modeling, simulation, quality of service, etc. This book, that is suitable for both PhD and master students, is organized into six sections: New Generation Networks, Quality of Services, Sensor Networks, Telecommunications, Traffic Engineering and Routing.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

INTECH

open science | open minds