

Application of multigraph sampling method in network traffic design of simulation model of integrated telecommunication and computer network

Slobodan Miletić^{1,}, Vladimir Mladenović², and Ivan Pokrajac¹*

¹Military technical institute, Communication and Information Department, 11000 Belgrade, Serbia

²Faculty of technical sciences in Čačak, University of Kragujevac, Information technology Department, 32000 Čačak, Serbia

Abstract. Designing an integrated telecommunications and computer network using a simulation model requires defining all the network and communication elements that make up the network and network communication. The accuracy, quality, and usability of the simulation results depend on the accuracy and the way of defining the network traffic as a timed event between the source and the destination in the network. In this paper, the method for a more accurate definition of network traffic required for the development of a simulation model using the software tool OPNET is proposed. Network traffic can be represented using a multigraph associated with an appropriate matrix of network traffic over time. Time definition of network traffic is enabled by applying the method of sampling multigraphs with the mathematical derivation of the corresponding statistical distribution function. Predicted or existing communication in the network describes using the derived function of network traffic distribution and precisely defines the OPNET simulation model.

1 Introduction

The efficiency of the decision-making process and basic actions in specific organizations such as industry, energy, banking and finance, traffic, emergency services, army, police, etc. are today based on major network telecommunications and computer communications between multiple participants in these processes. The nature of the said organization often requires the spatial distance of the participants in the communication with often specific working environmental conditions outside the building where the communication takes place. In addition, the conditions for the technological realization of such communication impose many limitations if the communication requires security, availability, reliability, and time efficiency. Such a form of communication requires the implementation of spatially distributed integrated telecommunications and computer networks (ITCN) with predetermined communication rules.

* Corresponding author: slobodan.miletic@vs.rs

Modification of the network architecture, technical-technological and topological optimization, and, if necessary, change of the type of communication and communication procedure are performed on the basis of the observed shortcomings. The accuracy of the network simulation results, which depends on the accuracy and precision of the description of the simulation model of the use of the dedicated software tool for simulation of use on the accuracy of the derived conclusions about the designed ITCN. The procedure of describing the simulation model of the network after defining all the factors of the network architecture and topology requires the most accurate definition of the planned or existing process of communication in the network as network traffic between network elements.

The way of defining the process for simulation of network traffic in the simulation model of the communication network is a procedure conditioned by the possibilities of the software tool and integrated methods. Defining network traffic can be complex, especially if the simulation of simultaneous distributions of several types of technologically different traffic is required. The procedure for defining network traffic in previous works [2-4] as well as in new papers [1, 5, 6, 7] is based on a statistical parametric description of similar or typical traffic as well as traffic measured and recorded in a similar network as in [1-6]. By applying the given methods for the realization of the simulation model, there is a possibility of wrong choice or description of network traffic and time of obtaining wrong simulation results. The description of network traffic in the OPNET simulation model is based on the selection and definition of statistical distribution functions that most closely describes network traffic events. In this paper, a method is proposed that applies estimates of multigraphs and associated matrices of communication interactions between network elements enables the derivation of distribution functions that more accurately describes the events of network traffic in the time interval.

2 Network traffic design of OPNET simulation model

Network traffic design methodology for the development of the OPNET simulation model ITCN given in [8] requests the realization of the network traffic matrix between sources and determination of communications in the network. The starting element for defining the network traffic matrix is a model of organizational structure that represents the elementary organization represents participants in communications (E_1, E_2, \dots, E_i), and shows their functional relations (F_1, F_2, \dots, F_n). Positions and mutual distances of participants in the space, where communications are committed, determine their physical and logical topological affiliation as elements of the network-central structure of ITCN and affiliation in subordinate network nodes. For the purpose of creating functional relationships, the type of necessary communications (voice, data, information) is defined, which are determined by information flows. For the realization and establishment of information flows in the ITCN, appropriate network application services are defined (S_1, S_2, \dots, S_n).

Operating procedures are applied within the network traffic matrix. in a given organizational structure, they define the times $\Delta T = [t_0, t_1, t_2, \dots, t_m]$ and the time schedule according to which the elements of the organization mutually establish a certain type of communication. The basic concept of mapping the organizational structure into a matrix of network traffic with relations and time sequence of communication between network elements, and mapping into the OPNET simulation model is illustrated in Figure 1. The application of the network traffic simulation method requires the definition of a time sequence in which network elements establish communication interaction and realize mutual communication and exchange of certain types of data.

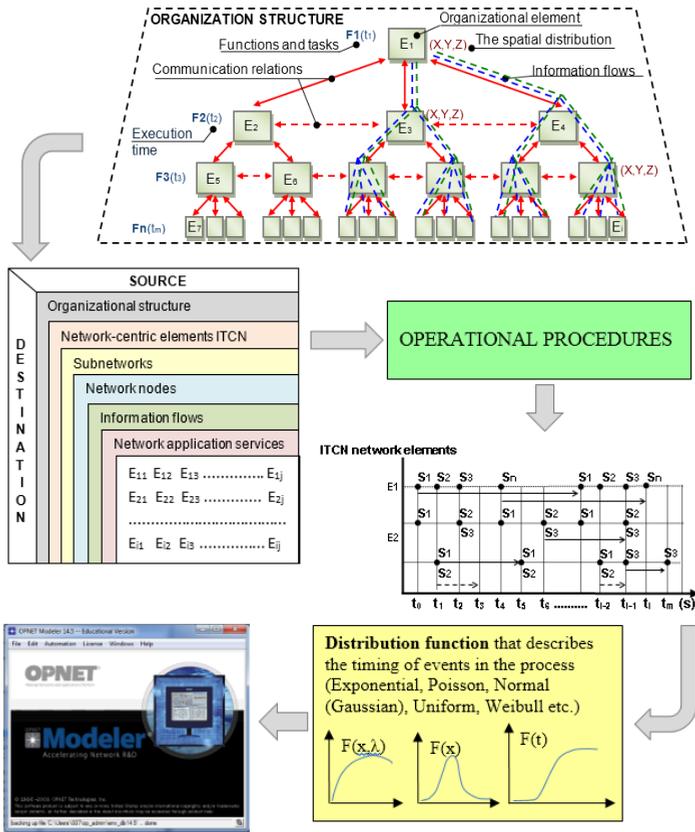


Fig. 1. The basic concept of forming a network traffic matrix for mapping into the OPNET simulation model of ITCN

Time moments of communication interactions between network elements in simulation time are determined by the time schedule of activation and duration of network services and shown by the timeline in Figure 2.

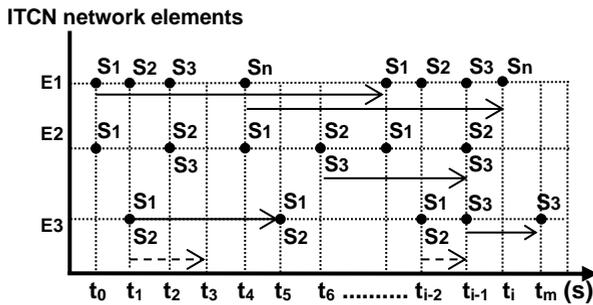


Fig. 2. Time-Line of activation and repetition of network elements communication interactions and network application services S_i

where is t_0 - activation time of services S_1 on the network elements E_1 and E_3 , t_1 - activation time of services S_1 and S_2 on the network elements E_3 and S_2 on the E_1 , t_i - the time by which activated required services S_n .

Communication relations, repeated communications, and time samples in the form of time schemes of activation, and duration of network application services are determined on the basis of operational procedures. The duration of communications is determined according to the amount of information for exchange and can be a stochastic or deterministic process.

2.1 Implementation of the network traffic matrix in the OPNET simulation model

For the realization of the network architecture and topological structure of ITCN in the simulation model, a network traffic matrix should be implemented. The definition of the connectivity matrix between the network elements of the ITCN is initially realized, and the predicted communication connections are involved with it. In the OPNET simulation model, time samples for establishing communications between network elements are described by selecting the statistical distribution function according to which time events are realized.

Communication between them is realized through one or more network services S_1, \dots, S_n and the duration of communication is also defined by the distribution function by the time when network element E_i establishes communication interaction with network element E_j by distribution function.

2.2 Selection of the distribution function

The selection of the distribution function should be done to describe realistic communication events from the operational procedures in order to OPNET's reality simulation model of ITCN presents network traffic analysis more realistic. The choice of the distribution function for application in the simulation model has a great influence on the reliability of the simulation of real network traffic, considering that the operational procedures are provided by time samples of a larger stochastic character. Defining the values of its parameters is necessary so as to really describe the process that is represented in addition to the choice of the distribution function.

The choice of distribution function can be based on the statistical study of communications in the network traffic acquisition according to [1-3] and [5, 6] or based on the approximation of a given type of network traffic (audio, messages, text, IP, VoIP, Video, HTTP, Web), ATM, etc.) with existing known distribution functions (Exponential, Poisson, Gaussian, Uniform, Weibull, etc.) according to [4, 9, 10, 11]. The way of selection of distribution function is arranged in the simulation model in the OPNET software according to [12] that is illustrated in Figure 3.

The problem of the real description of network traffic in the OPNET simulation model ITCN is complex in scope to require definitions of the distribution function for each predicted type of communication (network application services S_1 to S_n) between Internet network elements E_i and E_j .

The choice of an inappropriate distribution function or incorrect parameter definition results in a simulation of network traffic that does not correspond to the predicted ITCN network traffic. Such simulated results cannot be responsible for the analysis and optimization of communication networks. In order to reduce an error in the selection of distribution function, it is necessary to perform the appropriate distribution function on the basis of the network traffic matrix, which defines the communication interactions of network elements in the predicted time. By implementing the thus performed distribution function in OPNET simulation model, realistic descriptions and simulations of the predicted network traffic are provided.

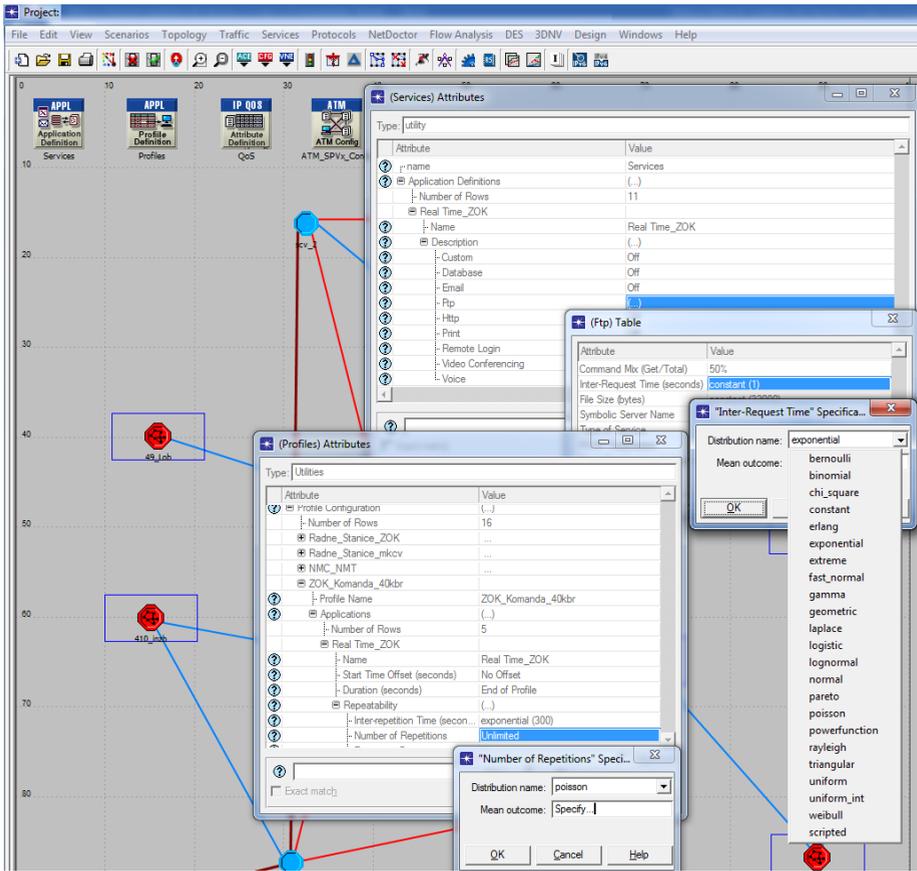


Fig. 3. Selection of the distribution function for activation and repetition time of the network application service in OPNET simulation model

3 Method for sampling multigraphs

The application of multigraph data sampling methods in [13] is one of the possible ways to derive the scheduling function approval from the network traffic matrix. Multigraphs can be expressed equivalently as symmetric integer neighborhood matrix with zero diagonal so that for sampling multigraphs we can equivalently sample neighborhood matrices. Sampling multigraphs is equivalent to sampling their associated matrices. The application of the multigraph sampling method in the design of network traffic of the OPNET simulation model ITCN enables the approval of scheduling functions to be defined more precisely. Within the ITCN design methodology, the method of sampling multigraphs can be applied to define the time distribution function of the realization of communication interactions between network elements of activation of network application services. Distribution functions are derived from the temporary matrices form of communication interactions of network elements associated with time changes to the multigraph.

3.1 Multigraph adjacency matrix of communication interactions of network elements

The time scheme of communication interactions is made between the network elements E_i and E_j , ($i \neq j$), by activating the network application services Srv at the predicted time $t = (t_0, t_1, .. t_{N-1}, t_N, .. t_m)$ interval $\Delta T = [t_0, t_m]$ based on the predicted order of communications given in Operational procedures. An example of the time order of activation and repeat of network application services S_1 to S_4 and the realization of communication interactions between network elements E_1 to E_8 in the form of a time scheme is shown in Figure 4. A given representation of a timeline is one of the possible forms of multigraph representation between multiple points. Each point represents one communication interaction between pairs of network elements (E_i, E_j) that occurs at time $t=(t_0, t_1, .. t_{N-1}, t_N, .. t_m)$. The color of the dot indicates the service S_1, S_2, S_3, S_4 with which the communication interaction is realized. The number of points between the two nodes represents the number of communication interactions.

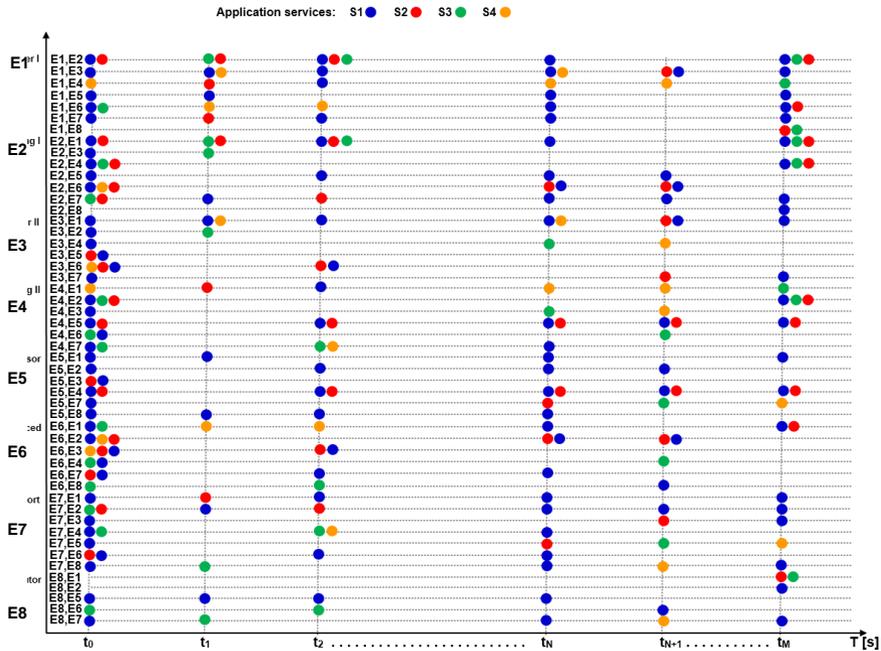


Fig. 4. Time-Line of activation and repetition of application services S_1 to S_4 and communication interactions between network elements E_1 to E_8

Additionally, in order to understand the above, Figure 5 illustrates a different graphical representation of the time scheme of communication interactions with Figure 4, where the network elements E_1 to E_8 in time moment $t=(t_0, t_1, .. t_{N-1}, t_N, .. t_m)$ (time plane) realize communication with other elements (E_i, E_j), ($i = j$), through one or more network application services S_1, S_2, S_3, S_4 . One network element may mediate one or more application services at one time to create multiple interactions with other network elements. For individual network products from E_1 to E_8 and for individual temporary assumptions $t = (t_0, t_1, .. t_{N-1}, t_N, .. t_m)$ a matrix of communication interactions is formed which are collected at that temporary moment.

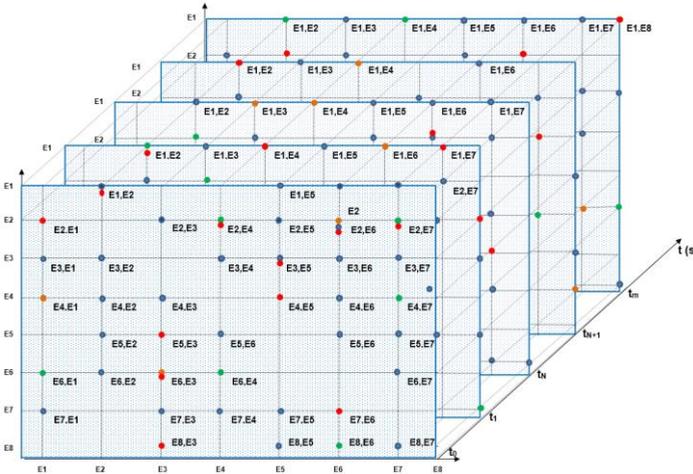
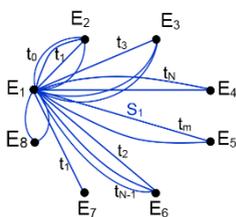


Fig. 5. Time-Plane of activation and repetition of application services S_1 to S_4 and communication interactions between network elements E_1 to E_8

Using the definitions of time schemes of communication interactions that represent a multigraph (as in Figure 6), the associated $n \times n$ matrix $TE_{i_Srv_t}$ with network elements $i = [1, n]$ is formed from communication interactions Network elements E_j ($i \neq j$), through the service S_{rv} , $rv=(1, 2,..)$ at time $t=(t_0, t_1, ..t_{N-1}, t_N, t_m)$. The creating of the matrix $T_{E_i_Srv_t}$ allows deriving the distribution function by the method of multigraph sampling, which defines the value of the evaluation of network traffic. This is realized by activating the application service S_1 between the network element E_1 and other network elements E_2 to E_8 at the temporary moment $t=(t_0, t_1, ..t_{N-1}, t_N, t_m)$.

By adding the matrix $T_{E_i_Srv_t}$, a matrix $T_{E_i_Srv_\Delta T}$ of total communication interactions through network products E_i is created through the service S_{rv} with other network elements in time intervals with the interval $\Delta T=[t_0, t_m]$. The matrix $T_{E_i_Srv_\Delta T}$ is reserved by a temporary variable multigraph which represents the change of communication interactions that function network products E_i with other network elements E_j , ($i \neq j$), via the service S_{rv} in the time interval $\Delta T=[t_0, t_m]$. For the example of the time scheme shown in Figure 4. the matrix $T_{E_1_S1_AT}$ is formed which represents all the communication interactions of the multigraph (example in Figure 6). They are created by network products E_1 with other network elements E_j , ($j = 2, 8$) by means of service S_1 in temporary moments of interval $\Delta T = [t_0, t_m]$, so:



$$T_{E_1 S_1 \Delta T} = \sum_{t=t_0}^{t_m} T_{E_1 S_1 t} \Rightarrow \begin{bmatrix} 0 & 11 & 9 & 7 & 4 & 7 & 5 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (2)$$

Fig. 6. Mapping multigraph in adjacency $n \times n$ matrix $TE_{1_Sv_t}$

Summarizing by all matrices $T_{E_i, S_{rv}, \Delta T}$, $i=[1, n]$, the summary matrix $T_{E:En, S_{rv}, \Delta T}$ of all communication interactions is obtained which develops between the network elements E_i and E_j , ($i=j$), using by the service S_{rv} in time interval $\Delta T=[t_0, t_m]$.

$$T_{E_i:En, S_{rv}, \Delta T} = \sum_{i=1}^n T_{E_i, S_{rv}, \Delta T} \tag{3}$$

thus the summation matrix $T_{E1:E8, S1, \Delta T}$ is obtained with Figures 4:

$$T_{E1:E8, S1, \Delta T} = \sum_{t=0}^8 T_{E1:S1, \Delta T} \Rightarrow \begin{bmatrix} 0 & 11 & 9 & 7 & 4 & 7 & 5 & 2 \\ 11 & 0 & 2 & 6 & 4 & 7 & 7 & 1 \\ 9 & 2 & 0 & 3 & 2 & 5 & 3 & 0 \\ 6 & 6 & 3 & 0 & 10 & 3 & 5 & 0 \\ 4 & 4 & 2 & 10 & 0 & 0 & 4 & 4 \\ 7 & 7 & 5 & 3 & 0 & 0 & 4 & 3 \\ 5 & 7 & 3 & 5 & 4 & 4 & 0 & 5 \\ 2 & 1 & 0 & 0 & 4 & 3 & 5 & 0 \end{bmatrix} \tag{4}$$

3.2 Function distribution of communication interactions of network elements

Multigraph sampling is the equivalent sampling of an associated matrix and the derivation of a distribution function $q(T)$ on the basis of which the matrix can be derived. The method for sampling $n \times n$ matrices is Sequential Significance Sampling (SIS) according to [13] which is a matrix sampled column by column and performs a function arranged using the asymptomatic approximations listed in [13] and [14]. Sequential importance sampling (SIS) method for sampling multigraphs with fixed degree based on an asymptotic approximation. This method samples column by column and performs best in cases where the graph is at least moderately sparse. The distribution function can be estimated using graphs and technical coefficients in the way given in [13] and [15] that achieves the approximation of multigraphs.

The matrix $T_{E_i}(t) = T_{E_i, S_{rv}, t}$ from the set of matrices $\Sigma T_{E_i, S_{rv}, t} = \Sigma T_{E_i}(t) = \Sigma T$, $t = (t_0, t_1, \dots, t_{N-1}, t_N, t_m)$ can be derived from the test distribution function $q(\cdot)$, where is $q(T_{E_i, S_{rv}, t}) > 0$ for all matrices $T_{E_i, S_{rv}, t} \in \Sigma T_{E_i, S_{rv}, t}$, and then the distribution estimate is:

$$E_q \left[\frac{1}{q(T_{E_i}(t))} \right] = \sum_T \frac{1}{q(T_{E_i}(t))} q(T_{E_i}(t)) = \left| \sum T \right| \tag{5}$$

$$\left| \sum T \right| = \frac{1}{n} \sum_{i=1}^n \frac{1}{q(T_{E_i}(t))} \tag{6}$$

To estimate the distribution function in the matrix function $\mu = E_p[f(T_{E_i}(t))]$ weights coefficients $\hat{\mu}$ are used as processes μ defined as in [13]. By sampling the matrix $T_{E_i}(t)$ of column by column (c_1, c_2, \dots, c_{n-1}) a special distribution function $q(\cdot)$ on the basis of which $q(T_{E_i}(t))$ should be determined can be represented as:

$$q(T_{E_i}(t) = (c_1, c_2, \dots, c_n)) = q(c_1)q(c_2|c_1) \dots q(c_n|c_{n-1}, \dots, c_1) \tag{7}$$

The sum of order d_i , the margin of order d of the matrix $n \times n$ are determined by the same procedure according to the given method in [13] and [14] for the matrix $T_{E_i}(t)$, then the updated margin of order $d^{(2)}, d^{(3)} \dots d^{(i)}$ matrices. The process of fitting the column of the matrix $T_{E_i}(t)$ is applied to the sampling of all columns c_1, c_2, \dots, c_n and the value of each

margin of the row is calculated, which is needed to express M - the total margin of the matrix [13]. The number of multigraphs $|\Sigma d|$ is calculated for the matrix $T_{Ei}(t)$ and the number of multigraphs corresponding to a separate submatrix is calculated for the submatrices formed by fitting the columns. The expression is derived on the basis of the asymptomatic approximations listed in [13] and [14]. The marginal distribution function of each column $p(c_i) \sim q(c_i)$ is determined by using the obtained expressions for each column (c_1, c_2, \dots, c_n) of the matrix $T_{Ei}(t)$, and represents the derived distribution $q(T_{Ei}(t))$.

$$p(c_1 = (0, \alpha_{21}, \dots, \alpha_{n1})) = \frac{|\Sigma d^{(2)}|}{|\Sigma d|} \tag{8}$$

$$p(c_2) = \frac{|\Sigma d^{(3)}|}{|\Sigma d|} \tag{9}$$

$$p(c_{n-1}) = \frac{|\Sigma d^{(n)}|}{|\Sigma d|} \tag{10}$$

By combining expressions [13, eq. (9) and (15)], the expression for $q(c_1)$ is derived:

$$q(c_1 = (0, \alpha_{21}, \dots, \alpha_{n1})) = \frac{1}{\prod_{i=1}^n (d_i - \alpha_{i1})!} e^{a(d^{(2)})} \tag{11}$$

The expressions for $q(c_1|c_2), \dots, q(c_n|c_{n-1}, \dots, c_1)$ and the calculation of $q(T_{Ei}(t))$ for each time moment $t = (t_0, t_1, \dots, t_{N-1}, t_N, t_m)$ should be performed in the same way. Derived distribution functions can be used to define the used calculated values from the derived expressions (Figure 7.) for the matrix $T_{Ei_Srv_ \Delta T}$, $i = [1, n]$.

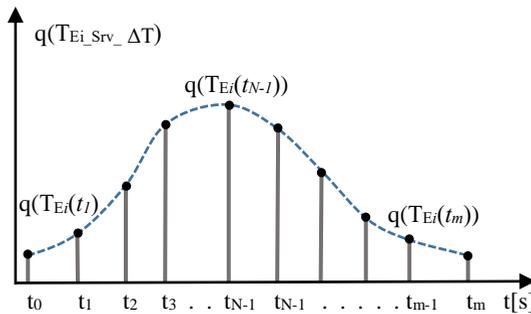


Fig. 7. Function distribution $q(T_{Ei_Srv_ \Delta T})$ for activation and repetition of application services S_{rv} and communication interactions between network element E_i

Matrix sampling efficiency estimates and the accuracy of the derived function distribute $q(T_{Ei}(t))$ in relation to the marginal distribution $p(T_{Ei}(t))$ can be determined by calculating standard error estimates and different cv^2 values by the procedure in [13]. The correction and adjustment of the values between the derived distribution functions $q(T_{Ei}(t))$ and the marginal arrangements $p(T_{Ei}(t))$ is realized by using the values of technical weights calculated by the procedure in [13] and [15].

4 Conclusion

In this paper, the method for a more accurate definition of network traffic required for the development of a simulation model using the software tool OPNET is proposed. The

statistical description of network traffic is more credible in relation to the annoying approximation and relative measurement of network traffic is shown in research and application of the SIS method of sampling multigraphs. Given the design process and the description of network traffic is tested with a simulation model that enables the use of the scheduling function to define network traffic. The advantage of applying multigraph sampling methods is that it allows a wider application so that communication events presented by the matrix can be statistically described and analyzed by simulation. Further research on the application of the multigraph sampling method will provide a statistical description of the change in the value of parameters in network traffic and its correlation with changes in the parametric connectivity and topology of the ITCN.

References

1. T. Tatarnikova, I. Sikarev, V. Karetnikov, A. Butsanets, E3S Web Conf. **244**, (2021)
2. I. Antoniuo, V.V. Ivanov, Valery V. Ivanov, P.V. Zrellov, *Statistical model of network traffic*, Physics of Particles and Nuclei, July (2004)
3. T. M. Chen, *Network traffic modeling*, Chapter in *The Handbook of Computer Networks*, Hossein Bidgoli ed., Wiley, (2007)
4. L. Leemis, *Input modeling techniques for discrete-event simulations*, Department of Mathematics, The College of William & Mary, Williamsburg, VA 23187–8795, U.S.A.
5. P. Dymora, M. Mazurek, D. Strzalka, *Computer network traffic analysis with the use of statistical self-similarity factor*, Annales UMCS Informatica AI XIII, 2 (2013) 69–81 DOI: 10.2478/v10065-012-0040-0, (2013)
6. M. Alsamar, G. Parisi, R. Clegg, N. Zakhleniuk, *On the distribution of traffic volumes in the Internet and its implications*, INFOCOM 2019, arXiv:1902.03853v1 [cs.NI], Feb (2019)
7. P. J. Sanchez, *Fundamentals of simulation modeling*, Winter Simulation Conference 2007, 1-4244-1306-0/07/ 2007 IEEE, (2007)
8. S. Miletić, M. Milošević, V. Mladenović, *Methodology for designing of tactical integrated telecommunication and computer networks for OPNET simulation*, Scientific Technical Review, 2020, **70,2**, 35-40, Belgrade, (2020)
9. R. De O. Schmidt, R. Sadre, A. Pras, *Gaussian traffic revisited*, 2013 IFIP Networking Conference, ISBN:978-3-901882-55-5, May (2013)
10. B. Chandrasekaran, *Survey of network traffic models*, https://www.cse.wustl.edu/~jain/cse567-06/ftp/traffic_models3/index.html
11. S. Manaseer, O. M. Al-Nahar, A. S. Hyassat, *Network traffic modeling*, International Journal of Recent Technology and Engineering (IJRTE), ISSN: 2277-3878, 7-5 (2019)
12. A. S. Sethi, V. Y. Hnatyshin, *The Practical OPNET User guide Computer network simulation*, CRC Press Taylor & Francis Group, ISBN: 978-1-4398-1206-8, (2012)
13. R. D. Eisinger, Y. Chen, *Sampling strategies for conditional inference on multigraphs*, Statistics and Its Interface, **11**, 649-656, (2018)
14. Y. Chen, P. Diaconis, S. P. Holmes, J. S. Liu, *Sequential Monte Carlo methods for statistical analysis of tables*, Journal of the American Statistical Association, **100**, **469**, Theory and Methods, DOI 10.1198/016214504000001303, March (2005)
15. A. Barrat, M. Barthelemy, R. P. Satorras, A. Vespignani, *The architecture of complex weighted networks*, Proc. Natl. Acad. Sci. USA **101**, 3747 (2004)