Journal of Physics: Conference Series

Service delays in strongly linked network communities

To cite this article: M I Bogachev *et al* 2019 *J. Phys.: Conf. Ser.* **1352** 012006

IOP Conf. Series: Journal of Physics: Conf. Series **1352** (2019) 012006 doi:10.1088/1742-6596/1352/1/012006

Service delays in strongly linked network communities

M I Bogachev[1], N S Pyko[1], S A Pyko[1], A N Vasenev[2] and A N Vasenev[3]

[1] Radio Systems Department, St. Petersburg Electrotechnical University,
5 Professor Popov street, 197376 St. Petersburg, Russia
[2] Department of Information Technologies, Ivanovo State University,
39 Ermak street, Ivanovo 153025 Russia
[3] University of Twente, 7522NB Enschede, the Netherlands

E-mail: rogex@yandex.com

Abstract. We analyze aggregated traffic dynamics obtained from strongly linked network communities. Our results based on two empirical data traces from university campus networks indicate that neglecting the statistical links between traffic patterns generated by individual network nodes leads to the drastic underestimation of both waiting and sojourn times. We also show that similar effects can be observed in simulated traffic patterns obtained by agent based modeling. Moreover, we suggest several indices that could be used to quantify the links between nodes and show their relation with the queuing system performance indicators.

1. Introduction

Modern public networks are typically characterized by intensive interactions between end users leading to the formation of strongly linked online communities. Such communities are often formed due to the users' interactions by means of various social network as well as professional online forums. Posting links that attract high attention in online resources lead to the abrupt traffic bursts at particular nodes hosting the information of interest. Additionally, localized communities are often formed by neighbors that are located closely to each other also in terms of network topology leading to similar effects in the local area networks (LAN) traffic patterns.

Conventional approach to the mathematical modeling of network traffic since the pioneering works of Erlang [1] relied upon the assumption that arrivals are independent of each other thus neglecting both auto- and cross-correlations in the users' access patterns. Providing reliable estimates for the entire era of telephones and early dial-up network communications, this approach remained being widely employed until the late 1980s. Over a long time particular model development have been focused mainly on the technical aspects of data transmission protocols such as packet data organization [2] leading to the modification like modulation of Poisson flow by a Markovian process [3] incorporating associated delays that were significant especially with the technologies of the time.

Although the emergence of erratic bursts in network traffic patterns as well as their association with the intensification of inter-user interactions dates back to the mid-1990s [4–6], in the last decade the role of the recent development of social networks and their impact on the users' activities that in turn determine traffic dynamics has been further elucidated [7–9]. Besides agent-based modeling, simulations based on phenomenological observations of long-term memory in the aggregated network traffic patterns have been developed [10–12].

For a better consideration of the variations in the users' activity patterns, in a series of recent publications [13–15] we considered a combination of the above approaches, including aggregated traffic modeling by Poisson flow of variable intensity determined by the number of active users in each time fragment based on the superstatistical approach [13], which was later superimposed by long-term memory in the users' activity patterns [14] with model parameters adjusted according to the universal rank-size traffic intensity distributions obtained from multiple empirical observations [15].

Our results indicate that the underestimation of the key performance indicators from queuing theory such as waiting and sojourn times could be reduced from 2–4 decades as provided by direct application of the conventional Kingman's formula [16] to less than one decade by considering appropriate models of both intensity distributions and long-term correlations in the users' access patterns under high network utilization conditions [13–15]. Although the above accuracy appears acceptable to a certain extent for such purposes as similar networks design where common hardware solutions provide one-decade scalability (100Mbps–1Gbps–10Gbps etc.), nevertheless more accurate estimations are required for other purposes such as the dynamic network resource control.

In the view of the above, we revisit our recent finding and suggest certain solutions that could lead to the generalization of our recently proposed network traffic models considering their keynote drawbacks. To our opinion, the first key drawback is that only autocorrelations in the traffic patterns have been simulated, this way only indirectly accounting for the cross-correlations in the access patterns induced by inter-user interactions. The second keynote aspect that only linear long-term memory models characterized by a single Hurst exponent have been considered, resulting in only partial assessment of the corresponding nonlinear effects achieved by the intrinsic linearization effects.

Therefore in this paper we suggest how the above limitations could be partially overcome by the direct assessment of the interactions between users. Moreover, in order to account also for the nonlinear relations between users' activity patterns, we suggest application of several metrics recently partially reviewed in [17] to the network traffic patterns. Considering two empirical data traces from university campus LANs that have been previously analyzed in [15], we show explicitly that individual access patterns from different network nodes indicate intrinsic cooperativity that also evolves over time. Based partially on our recent data [17], we show that several complementary indices used to quantify this cooperativity indicate characteristic phase transitions between rather autonomous and cooperative network activity patterns leading to considerably different dynamics of the aggregated traffic.

2. Data sources

We analyze two network traffic traces collected in-house at the downlinks between the local campus networks of St. Petersburg Electrotechnical University (LETI) and Ivanovo State University (IvSU) and their ISPs previously analyzed in [15]. The LETI dataset contained 10 complete daily traffic patterns from 17/03/2015 until 18/03/2015, from 16/04/2015 until 19/04/2015, and from 26/04/2016 until 29/04/2016 with typically 350 and 550 active internal IPs in 2015 and 2016, respectively. The IvSU dataset covered 15 consecutive days from 31/01/2017 until 14/02/2017 with about 250 active internal IPs. The total daily downlink traffic of these networks contained typically $2–4\times10^8$ packets. Traffic intensity was determined in non-overlapping windows of 1s duration resulting in 86400 data points per 24 hour cycle and evaluated in packets per second.

3. Analysis methods

In the following, we outline the key methods that we use to quantify cooperativity in the network traffic patterns induced by inter-user interactions. We consider two linear metrics based on conventional correlation analysis including the cross-correlation coefficients and cross-correlation times, respectively, as well as two nonlinear metrics based on the phase synchronization analysis and time delay stability estimates. For an extensive review of the latter methods as well as their performance analysis validated on simulated data series we refer to [17]. Due to the overwhelming amount of traffic directed to the busiest nodes in the network while least busy nodes being almost all the time unoccupied, we limited further analysis to the 100 busiest nodes for each daily access pattern. Each of the following indices has been estimated in non-overlapping windows of one hour duration. For interpretation simplicity all indices have been normalized such that they always belong to the [0;1] interval.

The cross-correlation analysis is based on the estimation of the averaged cross-correlation coefficients R between all possible pairwise combinations of users' access patterns. Additionally averaged correlation times T were determined by the integration of the absolute values of the mutual cosine similarity scores

$$CSS = \frac{\langle s_i, s_j \rangle}{\|s_i\| \cdot \|s_j\|} \tag{1}$$

obtained similarly for all possible pairwise combinations of users' access patterns.

The phase synchronization analysis is based on the assessment of the instantaneous phase differences between two arbitrary data series [18]. The instantaneous phases are determined using Hilbert transform which produces a complex dataset (sometimes called the analytical signal) by adding to the original real data series an imaginary part determined as

$$s_{\perp}(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{s(\tau)}{t-\tau} d\tau. \tag{2}$$

The instantaneous phase in then obtained as the argument of this analytical signal

$$\Phi(t) = \mathrm{arctg}\, \frac{s_{\perp}(t)}{s(t)}. \tag{3}$$

In long data series, phase differences are usually calculated in a gliding window, where neighboring points are treated as belonging to the same interval once the standard deviation of the phase differences is below a given threshold. To quantify the entire record, we have introduced the phase synchronization coefficient *Sync* that is determined as the fraction of windows with synchronous behavior [18]. Thus the method has at least two parameters, namely the gliding window size and the decision threshold.

The time delay stability estimate is based on the analysis of the relative shift of the maximum of the cross-correlation function of two studied series. We calculate the average delay in 50% overlapping windows of fixed length. The time delay stability episode is determined once within at least five consecutive windows the shift of the maximum of the cross-correlation function remains below a given threshold [19]. Like in the previous method, in order to estimate the time delay stability coefficient for an entire record, the TDS value is determined as the fraction of the time delay stability episodes in the total record duration.

4. Results

Based on the collected data series and using the above indicators, the daily traffic dynamics as well as the cooperativity patterns of the users' access traces has been assessed. The results indicate that, as expected, the traffic exhibits a characteristic daily cycle topping during afternoon and evening hours (see figures 1a and 2a), in general agreement with previous observations of barely stationary dynamics characterized by Hurst exponents close to one [15]. In contrast, the cooperativity indicators exhibit their characteristic maxima in the morning hours with moderate decay in the afternoon and evening towards the lowest levels observed around midnight (see figures 1b and 2b). Given that synchronization in arrival times leads to the considerable enhancement of delays, the traffic dynamics could be to a certain extent characterized by such derived metrics as the *effective cooperative traffic* obtained as a product of the traffic intensity by the appropriately normalized cooperativity metric (see figures 1c and 2c) or *vice versa* by the *equivalent non-cooperative traffic* indicating the amount of traffic that would lead to similar performance indicators in the case of non-cooperative users' access dynamics. It is remarkable that the *effective cooperative traffic* is characterized by rather limited daily variability.

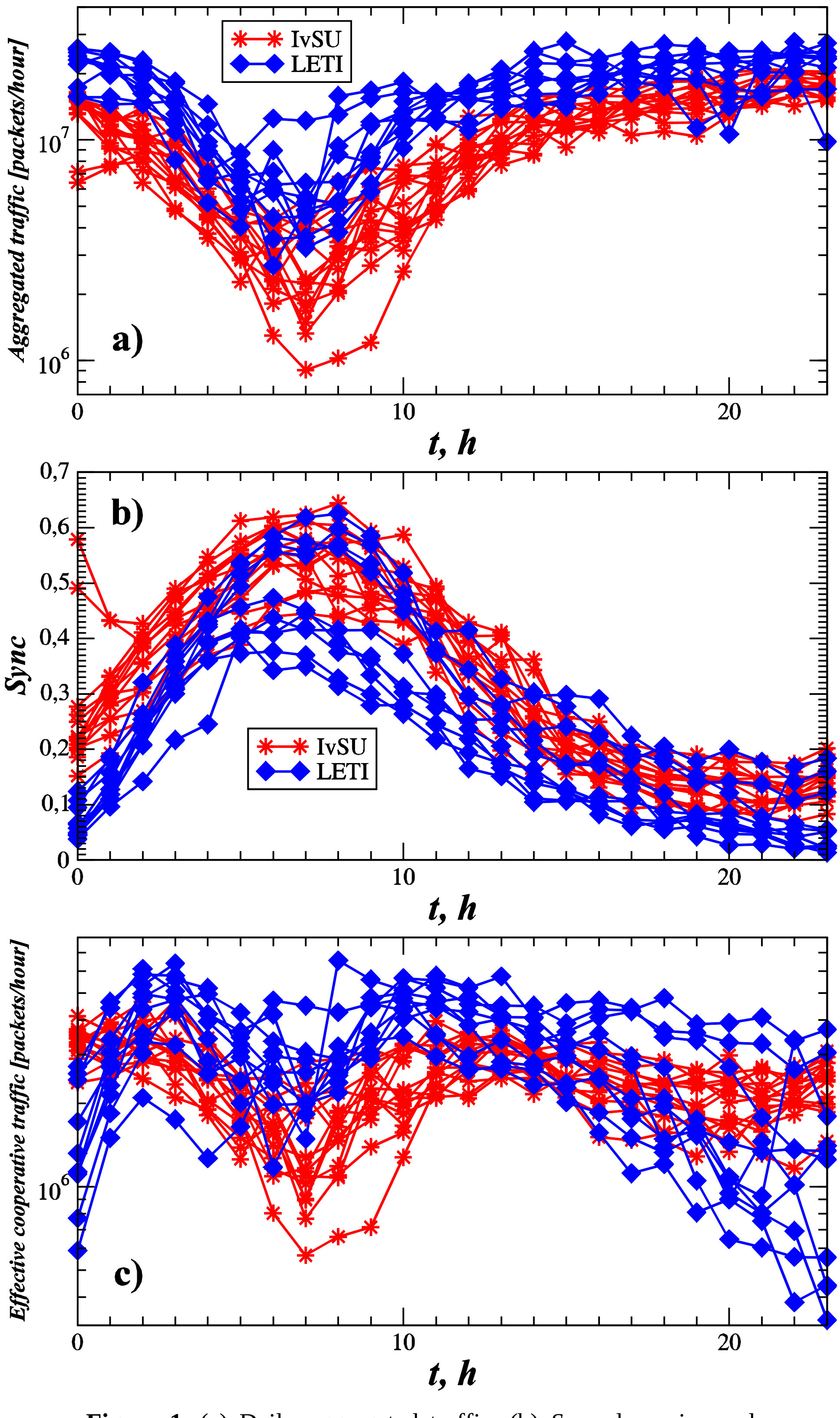

Figure 1. (a) Daily aggregated traffic, (b) *Sync* dynamics and (c) the *Sync*–based effective cooperative traffic estimate.

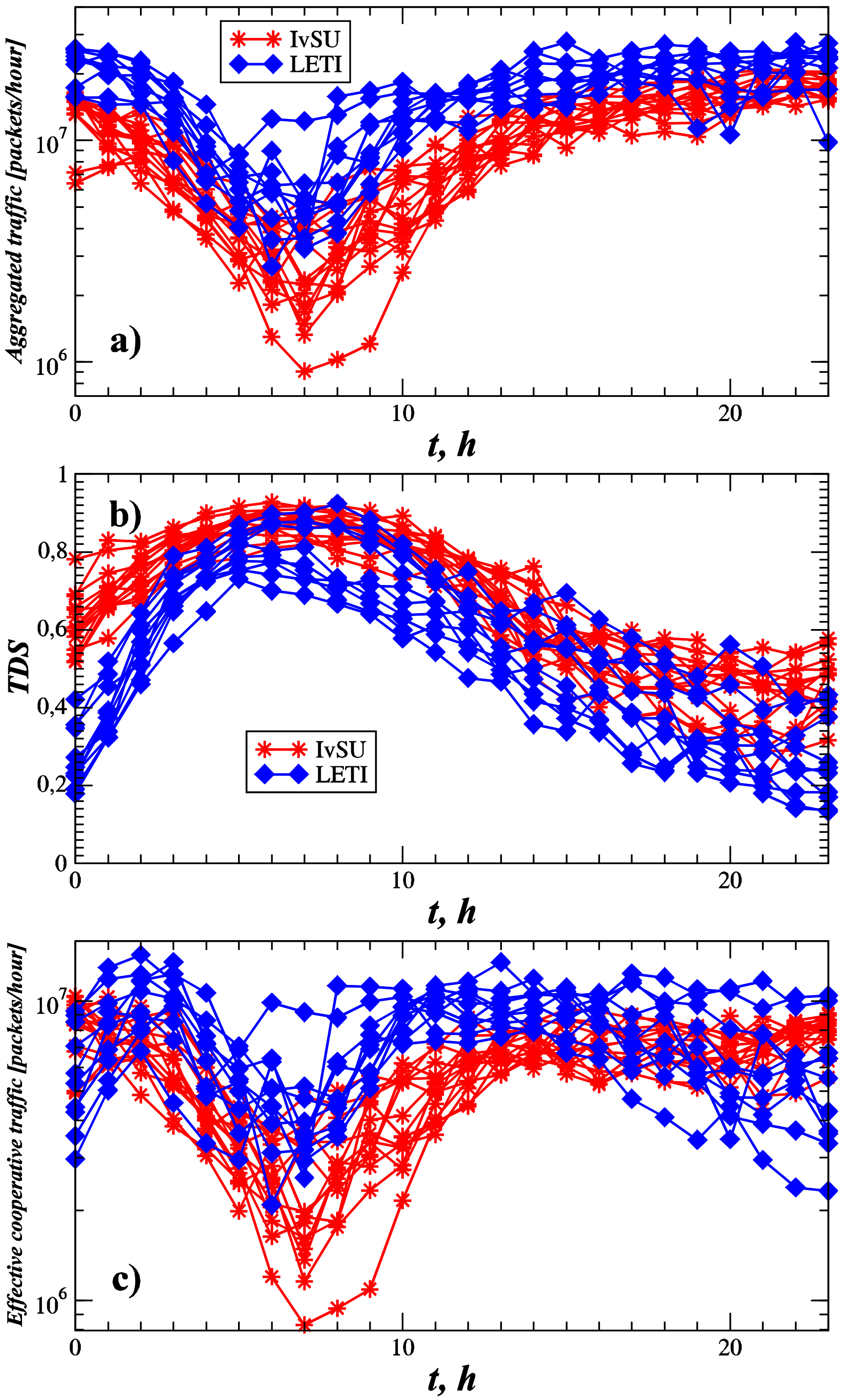

Figure 2. (a) Daily aggregated traffic, (b) *TDS* dynamics and (c) the *TDS*–based effective cooperative traffic estimate.

Also remarkably, both linear and nonlinear mutual behavior metrics provide nearly excellent reproducibility of the daily cycle as depicted in figures 3 and 4 indicating that it exhibits very characteristic and representative pattern while only slightly varying between the two studied LANs. Stability and reproducibility of the above indices in remarkable as it suggest a rather universal way of the cooperative access patterns simulation by generating multiple series with given mutual behavior indices.

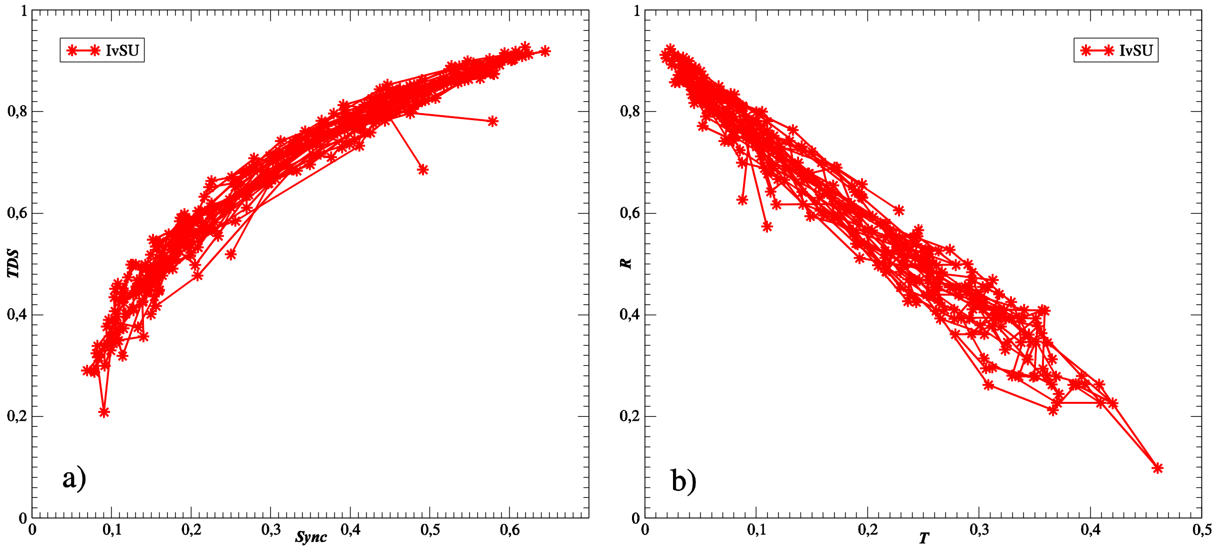

Figure 3. Cooperativity indices for the IvSU LAN traffic.

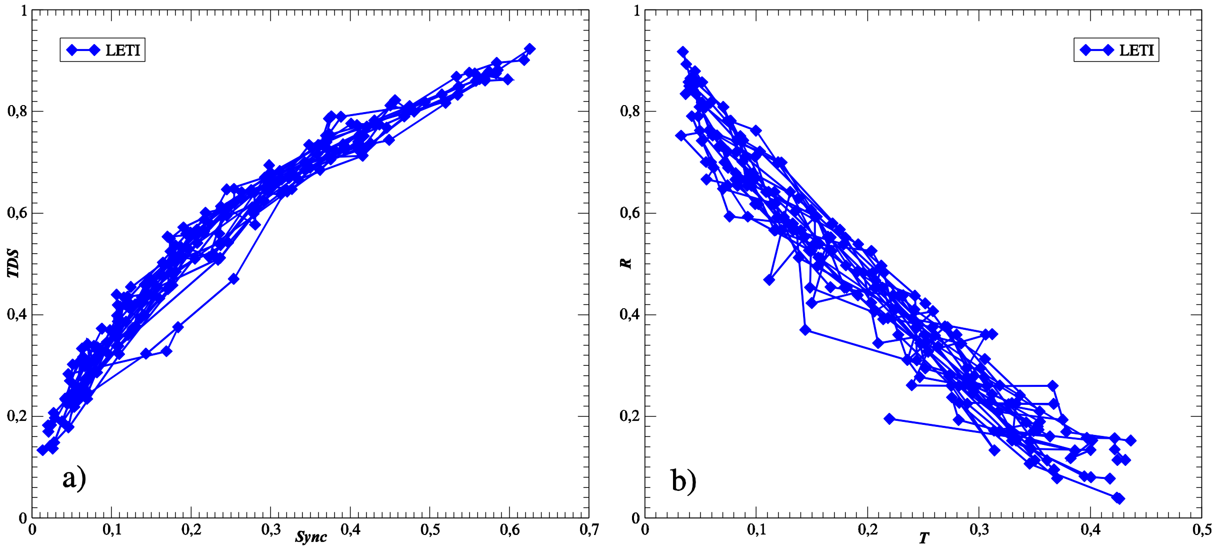

Figure 4. Cooperativity indices for the LETI LAN traffic.

In order to evaluate the effects of the cooperativity in the users' access patterns on the network performance characteristics, next simulations based on the queuing system theory principles has been performed. in terms of queuing theory. Either empirical or simulated packet arrival times have been applied to the downlink model based on the first-in-first-out (FIFO) queue with variable throughput such that average utilizations U between 0.1 and 0.98 have been achieved. Simulated packet arrivals were

IOP Conf. Series: Journal of Physics: Conf. Series **1352** (2019) 012006 doi:10.1088/1742-6596/1352/1/012006

obtained using the recently suggested superstatistical model by concatenation of 1s duration fragments of Poisson flows with variable intensity that was in turn determined by the model based on the rank-size statistics.

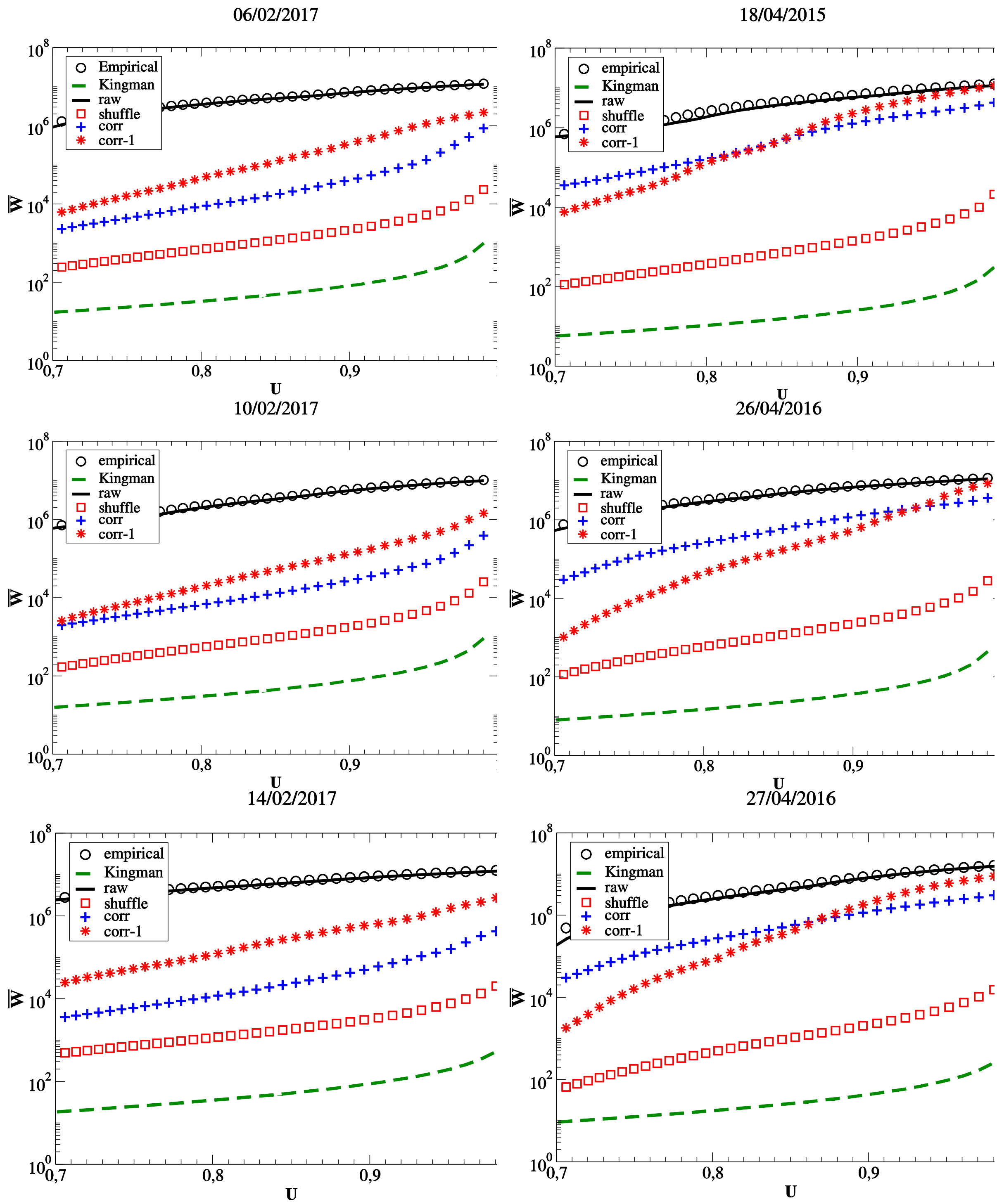

Figure 5. Queuing system performance simulation results.

Two variants of the superstatistical model have been considered, namely the one-dimensional model which considered only the aggregated traffic variations that implies corresponding long-term memory features as well as the two-dimensional model where individual access patterns from different network nodes have been simulated prior to their aggregation. In the latter scenario, inherent cooperativity effects could be to a certain extent taken into account by considering the empirical variability of the implied quantities. Simulation results were next compared against both empirical data and various simplified model variants such as series with shuffled inter-arrival times that accurately reproduced the distributions while contained no memory as well as the conventional approximation based on the Kingman's formula [16]. As the queuing system simulation followed standard methodology, for a more detailed discussion of the simulation procedure as well as its results interpretation strategy we refer to the corresponding sections in [13–15] and relevant references therein. Based on the simulation data the average times required to process a single packet known as sojourn times W given by the sum of the service time T_S and waiting time T_W as a function of the downlink utilization U have been evaluated (see figures 5).

5. Conclusion and outlook

Our results indicate that despite of the increased complexity of the two-dimensional model it outperformed the one-dimensional model only at extremely high utilizations U close to one. In contrast, at rather high but less extreme utilizations $U \approx 0.7\ldots0.9$ appearing more relevant in practical context no clear benefit could be observed.

Accordingly, additional information derived from a combination of nonlinear cooperativity indices might be helpful for further improvement the simulation accuracy. For that, simulation of data series with a given combination of auto- and cross-correlations plus certain combination of cooperativity indices is essential. One possible solution could be achieved by using multivariate nonlinear system identification tools. On the contrary, a reasonable alternative could be provided by finding the appropriate *equivalent non-cooperative traffic* indicating the corresponding access rates that would lead to comparable performance indicators under non-cooperative access pattern scenario.

Finally, as the proposed approach used rather universal features that are ubiquitous in nature we believe that similar principles could be employed for the simulation of other complex systems characterized by non-stationary dynamical and/or structural patterns that obey laws reasonably represented by superstatistical formalism (for a recent review, see e.g. [20] and references therein).

Acknowledgments

We would like to acknowledge the financial support of this work by the Ministry of Science and Higher Education of the Russian Federation in the framework of the basic state assignment to St. Petersburg Electrotechnical University (research project no. 2.5475.2017/6.7 to MIB).

References

[1] Erlang AK 1917 Solution of some problems in the theory of probabilities of significance in automatic telephone exchanges *Elektrotkeknikeren* **13** 5–13
[2] Jain R and Routhier S Packet trains – measurements and a new model for computer network traffic *IEEE journal on selected areas in Communications* **4 (6)** 986–995
[3] Heffes H and Lucantoni D 1986 A Markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance *IEEE Journal on selected areas in communications* **4 (6)** 856–868
[4] Leland W E, Taqqu M S, Willinger W and Wilson D V 1993 On the self-similar nature of Ethernet traffic *ACM SIGCOMM computer communication review* **23 (4)** 183–193
[5] Paxson V and Floyd S 1995 Wide area traffic: The failure of Poisson modeling *IEEE/ACM Transactions on Networking* **3 (3)** 226–244
[6] Feldmann A and Whitt W 1998 Fitting mixtures of exponentials to long-tail distributions to analyze network performance models *Performance evaluation* **31 (3-4)** 245–279

[7] Rybski D, Buldyrev S V, Havlin S, Liljeros F and Makse H A 2009 Scaling laws of human interaction activity *Proceedings of the National Academy of Sciences* **106 (31)** 12640–12645
[8] Rybski D, Buldyrev S V, Havlin S, Liljeros F and Makse H A 2011 Communication activity in social networks: Growth and correlations *The European Physical Journal B* **84 (1)** 147–159
[9] Rybski D, Buldyrev S V, Havlin S, Liljeros F and Makse H A 2012 Communication activity in a social network: Relation between long-term correlations and inter-event clustering *Scientific reports* **2** 560
[10] Riedi R H, Crouse M S, Ribeiro V J and Baraniuk R G 1999 A multifractal wavelet model with application to network traffic *IEEE transactions on Information Theory* **45 (3)** 992–1018
[11] Park K and Willinger W 2000 *Self-similar network traffic and performance evaluation* (NY: Wiley)
[12] Sheluhin O, Smolskiy S and Osin A 2007 *Self-Similar Processes in Telecommunications* (NY: Wiley)
[13] Tamazian A, Nguyen V D, Markelov O A and Bogachev M I 2016 Universal model for collective access patterns in the Internet traffic dynamics: A superstatistical approach *EPL (Europhysics Letters)* **115 (1)** 10008
[14] Markelov O, Duc V N and Bogachev M 2017 Statistical modeling of the Internet traffic dynamics: To which extent do we need long-term correlations? *Physica A: Statistical Mechanics and its Applications* **485** 48–60
[15] Nguyen V D, Markelov O A, Serdyuk A D, Vasenev A N and Bogachev M I 2018 Universal rank-size statistics in network traffic: Modeling collective access patterns by Zipf's law with long-term correlations *EPL (Europhysics Letters)* **123 (5)** 50001
[16] Kingman J F C 1961 The single server queue in heavy traffic *Mathematical Proceedings of the Cambridge Philosophical Society* **57 (4)** 902–904
[17] Pyko N S, Pyko S A, Markelov O A, Karimov A I, Butusov D N, Zolotukhin Y V, Uljanitski Y D and Bogachev M I 2018 Assessment of cooperativity in complex systems with non-periodical dynamics: Comparison of five mutual information metrics *Physica A: Statistical Mechanics and its Applications* **503** 1054–1072
[18] Bartsch R, Kantelhardt J W, Penzel T and Havlin S 2007 Experimental evidence for phase synchronization transitions in the human cardiorespiratory system *Physical Review Letters* **98 (5)** 054102
[19] Bashan A, Bartsch R P, Kantelhardt J W, Havlin S and Ivanov P C 2012 Network physiology reveals relations between network topology and physiological function *Nature communications* **3** 702
[20] Bogachev M I, Markelov O A, Kayumov A R and Bunde A 2017 Superstatistical model of bacterial DNA architecture *Scientific reports* **7** 43034