

Source Model of TCP Traffic in LTE Networks

Paweł Foremski, Michał Gorawski and Krzysztof Grochla

Abstract We propose a generator that represents traffic transmitted in mobile wireless networks. It is based on measurements of IP flows in a real, large-scale LTE network. The proposed tool generates TCP flow sizes and durations. We verify the generator by comparison with the traces available in the literature, and we propose application in discrete event simulators.

1 Introduction

The performance evaluation of wireless networks requires a proper representation of the traffic transmitted within the network. The realistic assessment of evaluated network architecture or networking mechanisms must be done under conditions representative for the analyzed scenario. In particular, the traffic transmitted through the network must have similar statistical properties as the traffic imposed in future working conditions. Therefore, a good modeling of traffic source is crucial for the performance evaluation of any network protocol.

The network traffic modeling problem has been analyzed for many years. The large number of models representing popular protocols and applications were proposed, like HTTP [2], video streaming [14], or online gaming [1]. As most of the traffic in the Internet is transmitted via HTTP, the novel models better representing this protocol are still being developed—see, e.g., [19]. The first traffic models were based on Poisson distribution, but the authors of [18] shown that it does not represent correctly the traffic characteristics. The Internet traffic has self-similar characteristics, thus a large number of more sophisticated methods were used to correctly represent the

P. Foremski (✉) · M. Gorawski · K. Grochla
Institute of Theoretical and Applied Informatics of the Polish Academy of Sciences,
Gliwice, Poland
e-mail: kgrochla@iitis.pl

M. Gorawski
e-mail: mgorawski@iitis.pl

P. Foremski
e-mail: pjf@iitis.pl

observed statistical properties: on/off processes, Hidden Markov Models (HMM), ARIMA processes, Wavelets, diffusion approximation and multifractals [3, 6, 10].

The wireless network traffic is currently rapidly growing: according to Cisco Visual Networking Index, the volume of data transmitted by mobile devices increased by more than 80 % in 2013 [17]. Mobile users use different applications and browse the web slightly differently than the users of regular PCs: according to [16], the largest fraction of traffic to mobile devices is multimedia, and the HTTP objects transmitted to mobile devices are on average larger than transmitted to other devices. A considerable part of the traffic on mobile devices is generated by the synchronization of mobile applications with the servers, such as, e.g., social networks, mail, or calendar applications. This type of traffic is not present on stationary computers, thus dedicated models representing the properties of mobile device traffic are needed for proper performance evaluation of wireless networking protocols and algorithms.

There are a few traffic source models dedicated for wireless networks evaluation. LiTGen [20] allows to reproduce accurately the traffic burstiness and internal properties over a wide range of timescales, but is limited to mail and P2P applications. Authors of [15] use Hidden Markov Models to represent different QoS classes of the network traffic, but the model is based on a small traffic trace gathered in laboratory from a WiFi network, thus is not representative per large-scale mobile network traffic. In [4], the authors propose a theoretical model of eNodeB traffic that considers 6 different parameters—for example number of subscribers, data and voice activity during the busy hour, and the bandwidth required for bearer sessions—but the model provides only rough estimates on the aggregated traffic volume of a typical eNodeB. To the best of our knowledge, we were not able to find a traffic source model based on large-scale measurements in the 4th generation wireless network, i.e., Long-Term Evolution (LTE) networks.

2 Proposed Model

In this work, we propose a traffic generator representative for modern mobile devices traffic in LTE networks. It is based on measurements of traffic in a large-scale wireless network. The proposed tool generates TCP flow sizes and durations. We implement our model in the OMNeT++ environment and verify it by comparison to traces available in the literature. The model is based on a real-world research data [12], and thus generates traffic that is similar to transmissions in real mobile devices.

2.1 Network Traffic Data

Traffic traces are the key requirement for modeling network traffic. For example, popular methods apply HMM to model the interpacket time gaps and the packet lengths basing on observations of real traffic [6]. However, obtaining adequate samples of real-world LTE traffic is practically hard, so in our work we build on the data presented

Table 1 Characteristics of the real-world LTE dataset [12]

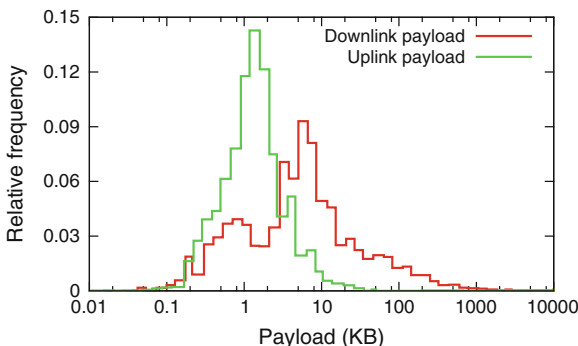
Capture start	Duration	Client IPs	Server IPs	Packets	Bytes	eNB count	Utilization	Flows (/5 min.)
October 12, 2012	240 h	379 K	719 K	3.8 G	2.9 T	22	29.5 Mbps	16.4 K

The column “Utilization” shows the average bandwidth used by all eNBs, while the “Flows” column shows the average number of traffic flows in 5-min time windows

in the recent paper by J. Huang et al., which gives an in-depth study of LTE performance [12]. We consider this work as representative and authoritative, as it presents the first study of a large real-world LTE packet trace. The traffic was collected from 22 eNBs in 2012 at a large metropolitan area in the U.S., and contains 2.9 TB of LTE traffic from over 300,000 users. The details on the dataset are presented in Table 1. For other papers that analyze wireless network traffic, see, e.g., [7, 8, 11, 13, 21]

The paper by Huang et al. characterizes several aspects related to the performance of LTE networks, but in our work we use the most fundamental data: the statistical characteristics of TCP session size and TCP session duration, given in Sect. 4.1 of [12]. As the authors point out, the TCP protocol dominates in LTE networks—carrying over 97 % of bytes and 95 % of flows—so we believe it is enough to model just the TCP protocol for a good approximation of the whole mixture of LTE traffic. Interestingly, for TCP, the pair of HTTP/HTTPS protocols is responsible for over 92 % of traffic flows.

In order to obtain the data required for our model, we digitized Figs. 2 and 3 from [12] and statistically processed the obtained data, as explained herein. After digitizing the Cumulative Distribution Functions (CDFs) of TCP session sizes and durations, we used the finite difference method to obtain their histograms and Probability Density Functions (PDFs), displayed in Figs. 1, 2, 3, and 4. Due to the range of the underlying data, density functions had to be presented in the logarithmic scale. This makes their interpretation harder, because integral of a PDF over its entire range must equal 1, but the corresponding histograms allow us to easily draw conclusions. The fit of the log-normal distribution is presented in Figs. 3 and 4 to roughly compare the data

**Fig. 1** Histograms of TCP flow sizes

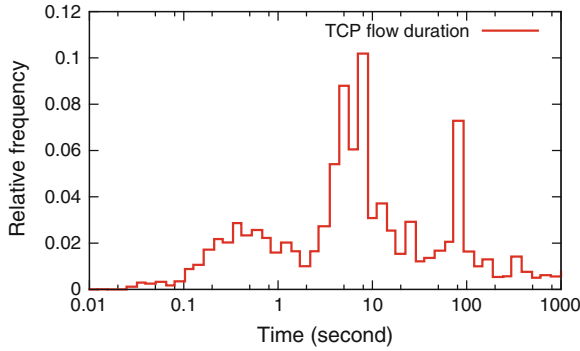


Fig. 2 Histogram of flow durations

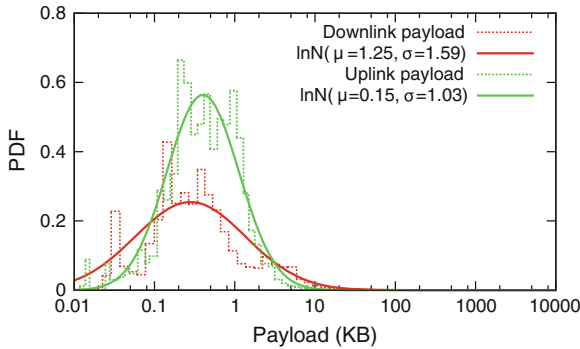


Fig. 3 Distribution densities of TCP flow sizes

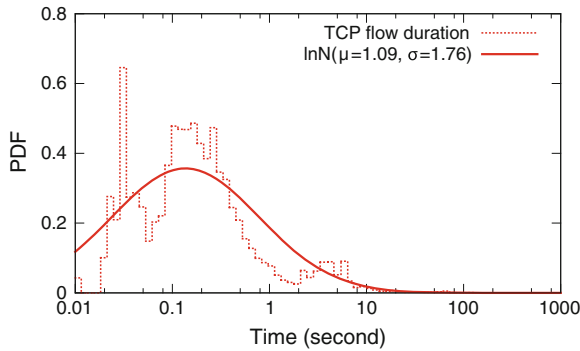


Fig. 4 Distribution density of flow durations

with a simple model. As visible in Fig. 4, the empirical distribution of flow duration cannot be explained using such a data model.

We chose to generate traffic flow data using inverse transform sampling on the digitized CDF plots. Each flow F is fully characterized with a tuple of $F = \langle S_u, S_d, D \rangle$,

where S_u and S_d describe the TCP payload size in the uplink and downlink direction (respectively), and D is the flow duration (i.e., the time from the first packet to the last packet). In order to linearly correlate $\log(S_d)$ and $\log(D)$ we used two correlated random variables X and Y for sampling the CDFs of S_d and D :

$$X \sim \mathcal{U}(0, 1), \tag{1}$$

$$Y = c \cdot X + (1 - c) \cdot Z, \tag{2}$$

$$Z \sim \mathcal{U}(0, 1), \tag{3}$$

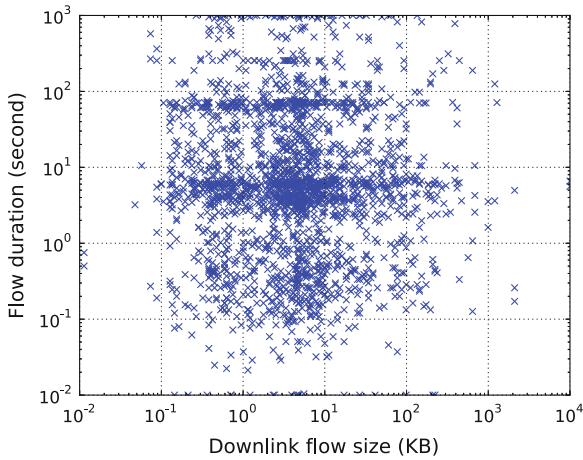


Fig. 5 Uncorrelated flow data

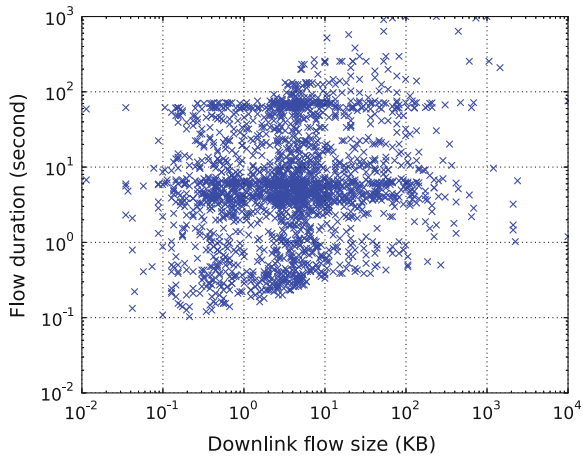


Fig. 6 Correlated flow data, $c = 0.16$

where $c \in [0, 1]$. The correlation of 0.196 reported in [12] was obtained for $c = 0.16$. See Figs. 5 and 6 for comparison between uncorrelated and correlated $\log(S_d)$ and $\log(D)$, for a sample of 2,500 pairs. For the uplink direction, we consider $\log(S_u)$ and $\log(D)$ as uncorrelated. Please refer to Chap. 7 and Example 2.1 in [5] for background on generating correlated random variates. We leave the task of describing F with a more sophisticated model for future work.

2.2 Network Traffic Model

To model the traffic from mobile devices, we used the OMNeT++/INET discrete event simulator, basing on the model of TCP connection described in [9]. To simulate uplink and downlink flows, we implemented two modules: a client (representing a mobile device) and a server (representing the Internet). Both modules were implemented as TCP applications (TCPApps) in the INET StandardHost module (Fig. 7). The StandardHost module enables simulating data flows in a TCP environment, along with a representation of all network layers. We assume no concurrent sessions, hence the communication is realized using a single pair of sockets. The sockets are closed after the session ends, and reopened for the next session: server socket is set to the ListenOnce state and binds to the client socket on connection request. Flow data is generated on the client side, where the correlated values described in Sect. 2.1 are

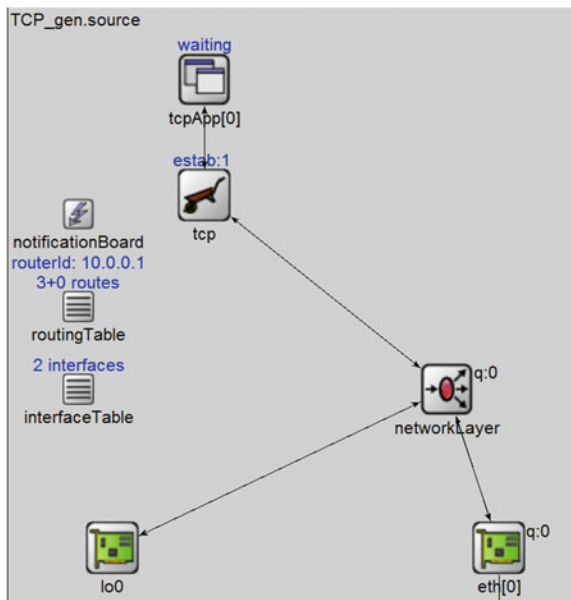


Fig. 7 The OMNeT++ implementation

drawn randomly. The packet transmission interval is set to 100 ms, and the packet size is constant for the whole session, so the total flow size matches the generated data.

The transmission is realized as follows. First, an uplink packet (named “Precast”) is sent. It includes the information about expected response size and session duration. When the server receives the first packet it starts its own independent transmission with an adequate packet size. Transmission ends when both sides finish their transfers, and the session is immediately closed. In case of no uplink data, the client sends a single Precast packet and waits for the server to finish. For session times below 100ms, the transmission is resolved by TCP modules; for session times between 100 and 200ms, transmission is trimmed down to 100ms. The interval between consecutive sessions is set by default to 1s, but it can be adjusted to any value or distribution, as needed. Note that this interval regulates the aggregated client bandwidth. Current implementation does not support simultaneous transmissions, which was left for future versions of our traffic generator.

3 Experimental Results

For experimental evaluation, we run our model implemented in OMNeT++ and collected statistics of 100,000 TCP sessions.

In Fig. 8, we compare the flow durations obtained from simulation with the expected flow data values, using a Quantile-Quantile plot (QQ-plot) of 1,000 samples. As visible, the distributions generally match each other, with an exception for the data in the range of 100–200 ms (on the OX-axis). This is expected, because our model presented in Sect. 2.2 divides data into 100 ms transfer windows, which limits

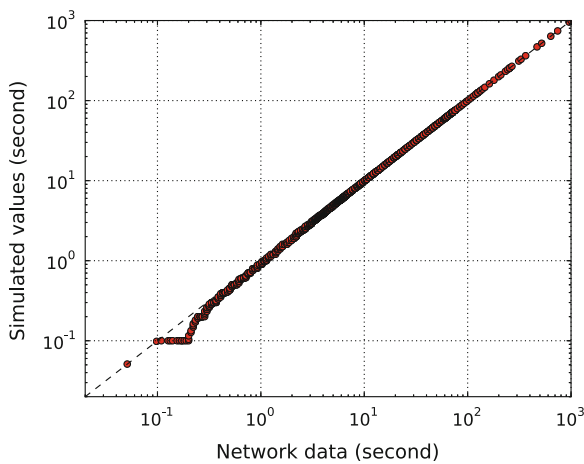


Fig. 8 QQ-plot of flow durations

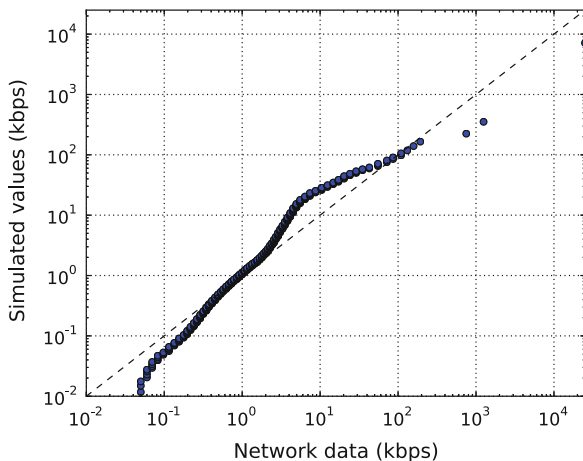


Fig. 9 Uplink flow rates (500 samples)

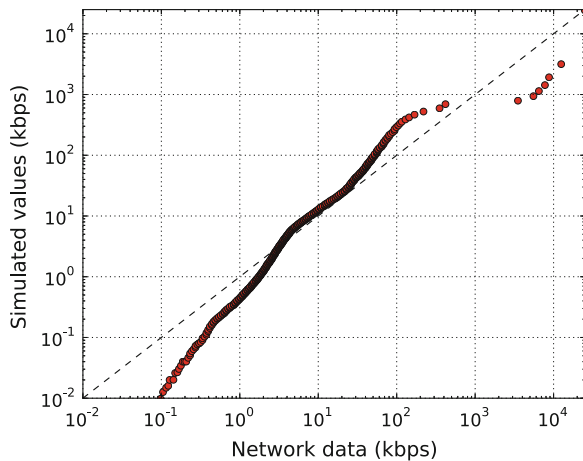


Fig. 10 Downlink flow rates (500 samples)

the granularity of flow durations observed in simulation. For larger flow durations this effect is negligible, hence not visible on the plot.

In Figs. 9, 10, 11, and 12 we compare the TCP flow rates obtained in OMNeT++ with the measured flow data shown in [12]. For both directions we see that the distributions are similar, with only minor deviations. Generally, slow flows (less than a few kilobits per second) and fast flows (more than a few megabits per second) tend to be transmitted with lower bit rates in the simulation environment. On the

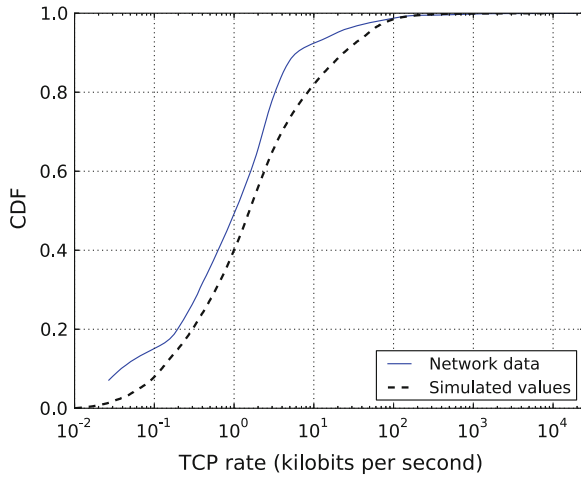


Fig. 11 CDF comparison of uplink flow rates

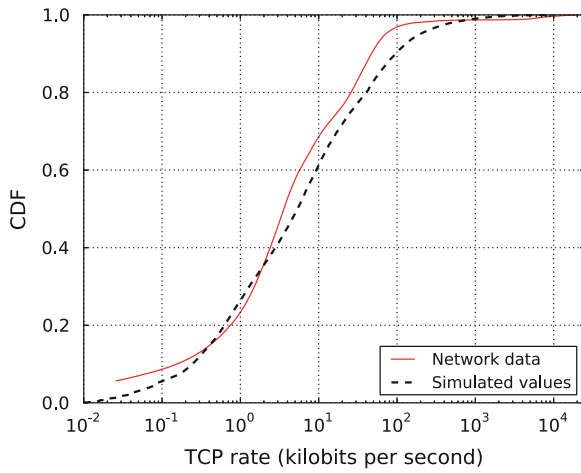


Fig. 12 CDF comparison of downlink flow rates

other hand, some flows in the middle are transmitted faster: for uplink the flows in the range of 5–100kbps and for downlink the flows in the range of 50–1,000kbps.

4 Conclusions

In this paper we proposed a model of TCP traffic that represents statistical properties of traffic in large-scale LTE networks, using the OMNeT++ discrete event simulation environment. The model is based on measurements of flow sizes and durations in a real network. The evaluation of the model confirms that it well matches with the traffic observed in the literature, providing similar distributions of transmission rates, for downlink and uplink directions.

The presented model can be used in various simulations as a source of TCP traffic generated by a mobile device working in an LTE network. We acknowledge that the model is based on the data collected in the U.S., whereas the behavior of the network users in other parts of the world may vary. However, as of this writing, there is no credible network data available, e.g., for European users, which may be an interesting avenue for future research.

The model can be downloaded at <https://projekty.iitis.pl/rezultaty-badan-2-en>.

Acknowledgments This work was funded by the Polish National Centre for Research and Development, under research grant nr LIDER/10/194/L-3/11/: project “Optimization and load balancing in next generation wireless networks”, <http://projekty.iitis.pl/zosb>.

References

1. M.S. Borella, Source models of network game traffic. *Comput. Commun.* **23**(4), 403–410 (2000)
2. H. Choi, J. Limb, A Behavioral Model of Web Traffic, in *International Conference of Networking Protocol 99 (ICNP 99)*, pp. 327–334 (1999)
3. T. Czachórski, K. Grochla, T. Nycz, F. Pekergin, A diffusion approximation model for wireless networks based on iee 802.11 standard. *Comput. Commun.* **33**, S86–S92 (2010)
4. D. Dababneh, M. St-Hilaire, C. Makaya, Traffic model for long term evolution networks, in *2013 International Conference on Selected Topics in Mobile and Wireless Networking (MoWNeT)*, pp. 13–18. IEEE (2013)
5. L. Devroye, *Non-Uniform Random Variate Generation* (Springer, New York, 1986)
6. J. Domańska, A. Domański, T. Czachórski, Internet Traffic Source Based on Hidden Markov Model, in *Smart Spaces and Next Generation Wired/Wireless Networking, Lecture Notes in Computer Science*, ed. by S. Balandin, Y. Koucheryavy, H. Hu. vol. 6869 (Springer, Heidelberg, 2011), pp. 395–404
7. H. Falaki, D. Lymberopoulos, R. Mahajan, S. Kandula, D. Estrin, A first look at traffic on smartphones, in *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement*, pp. 281–287. ACM (2010)
8. A. Gember, A. Anand, A. Akella, A comparative study of handheld and non-handheld traffic in campus Wi-Fi networks, in *Passive and Active Measurement*, pp. 173–183. Springer (2011)
9. K. Grochla, Simulation comparison of active queue management algorithms in tcp/ip networks. *Telecommunication Systems* **39**(2), pp. 131–136 (2008) doi:[10.1007/s11235-008-9117-x](https://doi.org/10.1007/s11235-008-9117-x)
10. H. Hlavacs, G. Kotsis, C. Steinkellner, Traffic source modeling. Tech. Heport Tr-99101 Inst. Appl. Comput. Sci. Inf. Syst. Univ. Vienna (1999)
11. J. Huang, F. Qian, A. Gerber, Z.M. Mao, S. Sen, O. Spatscheck, A close examination of performance and power characteristics of 4G LTE networks, in *Proceedings of the 10th international conference on Mobile systems, applications, and services*, pp. 225–238. ACM (2012)

12. J. Huang, F. Qian, Y. Guo, Y. Zhou, Q. Xu, Z.M. Mao, S. Sen, O. Spatscheck, An in-depth study of LTE: Effect of network protocol and application behavior on performance, in *Proceedings of the ACM SIGCOMM 2013 conference on SIGCOMM*, pp. 363–374. ACM (2013)
13. J. Huang, Q. Xu, B. Tiwana, Z.M. Mao, M. Zhang, P. Bahl, Anatomizing application performance differences on smartphones, in *Proceedings of the 8th international conference on Mobile systems, applications, and services*, pp. 165–178. ACM (2010)
14. M. Krunz, A. Makowski, A source model for VBR video traffic based on $M/G/\infty$ input processes, in *INFOCOM'98. Seventeenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, vol. 3, pp. 1441–1448. IEEE (1998)
15. S. Maheshwari, S. Mahapatra, C. Kumar, K. Vasu, A joint parametric prediction model for wireless internet traffic using Hidden Markov Model. *Wireless Netw.* **19**(6), 1171–1185 (2013) doi:[10.1007/s11276-012-0525-1](https://doi.org/10.1007/s11276-012-0525-1)
16. G. Maier, F. Schneider, A. Feldmann, A First Look at Mobile Hand-Held Device Traffic, in *Passive and Active Measurement*, vol. 6032, Lecture Notes in Computer Science, ed. by A. Krishnamurthy, B. Plattner (Springer, Berlin, 2010), pp. 161–170
17. C.W. Paper, Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2010–2015 (2011)
18. V. Paxson, S. Floyd, Wide area traffic: the failure of Poisson modeling. *IEEE/ACM Trans. Networking (ToN)* **3**(3), 226–244 (1995)
19. R. Pries, Z. Magyari, P. Tran-Gia, An HTTP web traffic model based on the top one million visited web pages. In: *Next Generation Internet (NGI), 2012 8th EURO-NGI Conference on*, pp. 133–139. IEEE (2012)
20. C. Rolland, J. Ridoux, B. Baynat, LiTGen, a Lightweight Traffic Generator: Application to P2P and Mail Wireless Traffic, in *Passive and Active Network Measurement*, vol. 4427, Lecture Notes in Computer Science, ed. by S. Uhlig, K. Papagiannaki, O. Bonaventure (Springer, Berlin Heidelberg, 2007), pp. 52–62
21. Q. Xu, J. Erman, A. Gerber, Z. Mao, J. Pang, S. Venkataraman, Identifying diverse usage behaviors of smartphone apps, in *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*, pp. 329–344. ACM (2011)