Journal of Electrical Engineering and Information Technologies, Vol. **6,** No. 1, pp. 49–57 (2021)
Article 185
Received: April 15, 2021
Accepted: April 18, 2021
In print: ISSN 2545–4250
On line: ISSN 2545–4269
UDC: 621.396.721.5:[004.624:004.738.5
https://doi.org/10.51466/JEEIT21061185045b
Original scientific paper

MOBILE DATA TRAFFIC ANALYSIS IN VIRTUAL PRIVATE CLOUD: A CASE-STUDY IN OPERATIONAL MOBILE NETWORK

Ana Bogdanovska[1], Vladimir Atanasovski[2]

[1]*A1 Makedonija DOOEL Skopje, Ploštad Presveta Bogorodica 1, 1000 Skopje, N. Macedonia*
[2]*Faculty of Electrical Engineering and Information Technologies,*
"Ss. Cyril and Methodius" University in Skopje, North Macedonia
ana_bogdanovska@yahoo.com

Abstract: Mobile data traffic analysis is quintessential for mobile telecom operators in order to improve the quality-of-service for their customers. A mobile operator is interested in the usage popularity of services as well as the temporal distribution of the service usage. This enables identification of the volume of incoming and outgoing traffic and aids the operators to design a network that can support expected user activity. The latest networking trends point in the direction of virtualized network resources, thus exhibiting more effective network usage. The design of virtual networks is strongly affected by the pattern of mobile data for certain services. Therefore, a prior and adequate knowledge of the mobile services' popularity and their usage can significantly improve the user experience on the network. This paper provides empirical measurements and analytical modelling of mobile data traffic using data from an operational nationwide mobile network. The network utilizes the latest trends, i.e. a virtual environment and a private cloud. The paper proposes estimation parameters for the measured traffic that match the distribution of real traffic with the lowest error.

Key words: cloud; virtualization; traffic analytics; log-normal distribution; normal distribution

АНАЛИЗА НА МОБИЛЕН ПОДАТОЧЕН СООБРАЌАЈ ВО ВИРТУЕЛЕН ПРИВАТЕН ОБЛАК: СЛУЧАЈ НА ОПЕРАТИВНА МОБИЛНА МРЕЖА

Апстракт: Анализата на сообраќајот на мобилните податоци е од есенцијално значење за мобилните телекомуникациски оператори со цел подобрување на квалитетот на услугата за нивните клиенти. Од интерес за мобилните оператори популарноста на користењето на услугите, како и временската дистрибуција на користењето на услугата. Ова овозможува идентификување на обемот на влезен и излезен сообраќај и им помага на операторите да дизајнираат мрежа што може да ја поддржи очекуваната активност на корисниците. Најновите мрежни трендови упатуваат на правецот на виртуелизирани мрежни ресурси, со што се прикажува поефикасно користење на мрежните ресурси. Дизајнот на виртуелни мрежи е под силно влијание на моделот на мобилниот сообраќај за одредени услуги. Затоа претходно и адекватно познавање на популарноста на мобилните услуги и нивната употреба може значително да го подобри корисничкото искуство на мрежата. Овој труд обезбедува емпириски мерења и аналитичко моделирање на мобилниот пакетски сообраќај со употреба на податоци од оперативна мобилна мрежа низ целата земја. Мрежата ги користи најновите трендови, т.е. виртуелно околина и приватен облак. Трудот предлага параметри за процена на измерениот сообраќај кои одговараат на распределбата на реалниот сообраќај со најмала грешка.

Клучни зборови: облак; виртуелизација; анализа на сообраќај; лог-нормална дистрибуција; нормална дистрибуција

1. INTRODUCTION

The increasing popularity of wireless technologies and wireless terminals has led to an explosive growth of mobile data traffic lately. It is driven by improved device capabilities, an increase in data-intensive content and more data throughput from subsequent generations of network technology. Ericsson's [1] forecasts suggest that global mobile data traffic will grow by a factor of 4.5 to reach 226EB per month in 2026.

Cloud-computing is one of the fastest growing segments in the telecommunications industry [2]. The main benefit of cloud-computing is that users can start using services with a minimum infrastructure installation time that is very different from the classic CAPEX model (one where the required processing resources are procured or leased from the data center owner or service providers through established processes). An inherently connected term with cloud-computing is the notion of virtualization [3]. In computing, it refers to the act of creating a virtual (rather than actual) version of computing resources on the same hardware, including, but not limited to, a virtual hardware platform, Operating System (OS), storage device or network resources.

The mobile networks evolution towards 5G inevitably requires implementation of the concepts of cloud-computing and virtualization [4]. This will allow mobile telecom operators to deploy more flexible networking architecture thus easily introducing and managing various novel services. On the other hand, the heterogeneous nature of different services imposes different network and Quality-of-Service (QoS) requirements. The process of design, operation and optimization of mobile networks is becoming more challenging as a result. A proper understanding of mobile data traffic, its patterns and behavior is crucial for day-to-day network operation.

This paper deals with empirical measurements of mobile data traffic in a nationwide operational mobile network that uses state-of-the-art technology (virtualized core).

The measurements are then used to analytically model the mobile data traffic using mathematical distributions with appropriate parameters. These models can be used for offline network planning, thus significantly aiding the process of understanding and predicting mobile data traffic patterns.

The paper is organized as follows. Section 2 gives an overview of the latest evolution technologies that are used in mobile packet core networks. Section 3 introduces target user scenarios for empirical measurements of mobile data traffic, whereas section 4 provides the actual measurements. Section 5 deals with mathematical modelling of mobile data traffic. Finally, section 6 concludes the paper.

2. CLOUD-COMPUTING AND VIRTUALIZATION

Cloud-computing can be defined as Internet-based computing that provides multiple scalable on-demand services through sharing or accessing computing resources. Those resources can be from private systems or third-party and users can access them locally or remotely.

2.1. *Different types of cloud*

There are four main cloud models. Public cloud is intended for open use by the general public. It is hosted or operated by a service provider who sells or offers a multitude of services to the public. Common uses are for application development and testing, file sharing memory and e-mail service. Private cloud is intended for the exclusive use by individual organization. This is a choice for companies that already have a data center and are developing IT infrastructure and have special security or performance needs. Community cloud includes collaboration and integration of IT infrastructure and resources from multiple organizations. Examples include universities collaborating in specific areas of research, or county or state police departments that share computer resources. Access to the community cloud is usually restricted to the organizations that are included in the community. Finally, hybrid cloud is a deployment model that merges two or more cloud service deployments, such as the private, public and community clouds.

2.2. *NFV architecture*

Network Function Virtualization (NFV) was developed by ETSI [5]. NFV provides a standardized model that is different from the commercial hardware model where each hardware element represents a function. This is reflected in network functions delivered through virtualized software on Commercial Off-The-Shelf (COTS) hardware called Virtualized Network Function (VNF). A Virtualized Network Function is a logical grouping of one or more software entities such as Virtual Machines (VMs). The result of virtualization is a network that is more agile and has the ability to better respond to the demand and dynamic needs of telecommunications traffic and services. Figure 1 shows the NFV architecture defined by ETSI [11].

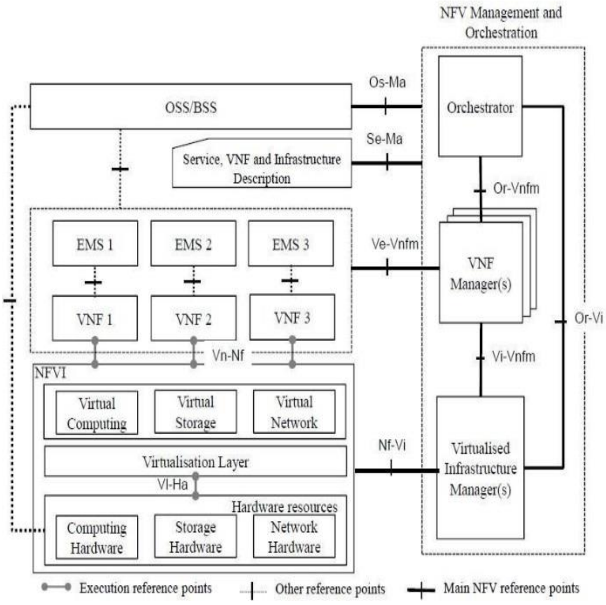

Fig. 1. ETSI NFV architecture.

The NFV architecture comprises three working domains. The first one is the NFVI or Network Function Virtualization Infrastructure. It is a basic platform that contains hardware resources as well as virtual instances that build the infrastructure on which VNFs are developed. The second one is VNF or Virtual Network Function, a software implementation of a network function that is adequate to run NFVI. Each VNF can be defined with one or more VMs. Each virtual network function is directly connected to the EMS Network Management System (EMS), which takes care of the functionality of each VNF that is connected to it. Finally, the third domain is NFV MANO (Management and Orchestration) that manages NFVI and VNF and is connected to other components including Physical Network Functions (PNF) and their management systems such as EMS and OSS / BSS.

Prior to the actual deployment of a cloud in the core, there is a need for a software platform capable of building and managing various clouds. The following subsection briefly introduces the most popular one, i.e. the OpenStack [6].

2.3. *OpenStack*

OpenStack is an open source project defined as a global collaboration of developers and technicians focused on developing an ubiquitous cloud-computing platform for public and private clouds. The technology consists of a series of interrelated projects that control pools of processing, storage, and networking resources throughout a data center – which users manage through a web-based dashboard, command-line tools, or a RESTful API.

The evolution of core networks of mobile operators converges towards virtualization. This proves to be cost-effective and flexible for service provisioning. The target network for performing empirical measurements in this paper has an already deployed virtual private cloud in the core. As of this writing, this implementation is still considered relatively new and not explored in sufficient details. The goal of the paper is to provide modelling of mobile data traffic of different types that are easily provisioned and managed in the virtual private cloud in the core. The following section will first address the user scenarios of interest for mobile data traffic analysis.

3. USER SCENARIOS

This section introduces two user scenarios for a nationwide mobile telecom operator that deploys a private cloud and offers mobile data services. The scenarios will be used for data traffic analytics through empirical measurements.

The first user scenario refers to the subscribers who use mobile data traffic. Namely, the total number of users who use the mobile service (voice and data) for this telecom operator is 1,050,000 users. Active subscribers to mobile Internet access are 680,000. The distribution of service usage is as follows:

- Web data traffic is used by 266,389 users on average per day (39% of users);
- Social networks data traffic is used by 153,916 users on average per one day (23% of the users);
- Messaging or chat traffic is used by 72,894 users on average per day (11% of the users).

The number of users that use a certain service is obtained as an average number of the users who have used that service during a period of two weeks (tests were run in summer 2020). It should be noted as this is a realistic scenario, subscribers can use multiple applications at one moment in parallel.

Web traffic includes the amount of traffic that users send and receive from websites. Traffic related to social networks includes Facebook, Instagram, SnapChat and TikTok. Chat traffic is analyzed for the applications Messenger and Whats App. Figure 2 shows average number of subscriber distribution per different Radio Access Technology (RAT) for the services under consideration.

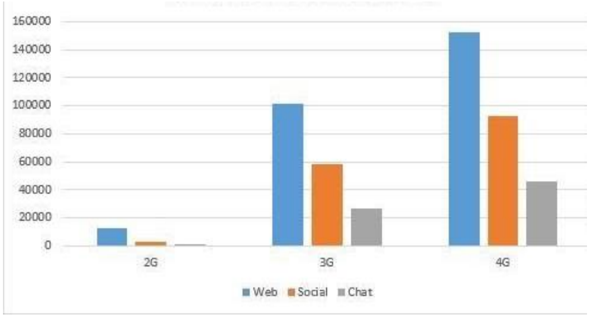

Fig. 2. Average number of subscribers per RAT.

The second user scenario refers to subscribers that are using Voice-over-LTE (VoLTE) service. VoLTE service is used by an average of 35 users in one day. It is important to emphasize that at the time the VoLTE analysis was performed, VoLTE is still considered a "test" service because not all mobile devices support it. Therefore, the VoLTE traffic is significantly lower than other types of traffic.

The targeted scenarios will be used for conducting mobile data traffic measurements. The following section provides insight into the empirical results.

4. EMPIRICAL MEASUREMENTS

The empirical measurements are performed for mobile data traffic in a time period of two weeks (during summer 2020). The analysis is made for each scenario separately, where the traffic consumed during working days is compared to the traffic consumed during the weekend in a period of 24 hours.

Figure 3 shows the average number of Simultaneously Attached Users (SAU) for the period of two weeks. The number of users is divided into 3 RATs, i.e. 2G, 3G and 4G. This figure gives a good insight into the distribution of network users per RAT that is important for the subsequent analysis.

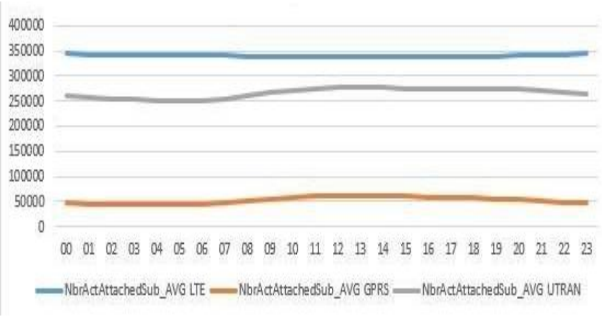

Fig. 3. Simultaneously attached users per RAT

Figures 4, 5, 6 and 7 show volumes of web traffic, social networks traffic, chat traffic and VoLTE traffic. It is easy to observe a typical user behavior with variations in traffic consumption during a 24-hour period and during weekends. Also, various traffic types exhibit peaks at different times and this is a crucial parameter for network resources planning and optimization.

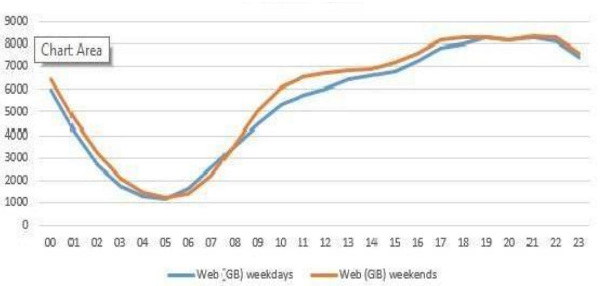

Fig. 4. Average volume of web traffic

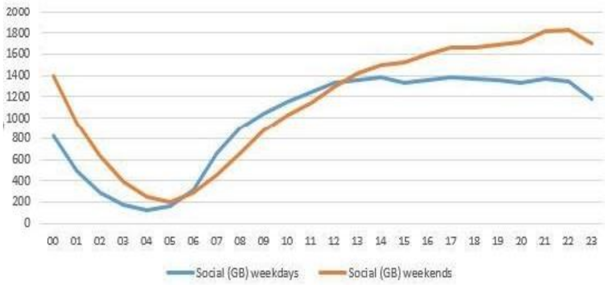

Fig. 5. Average volume of social networks traffic

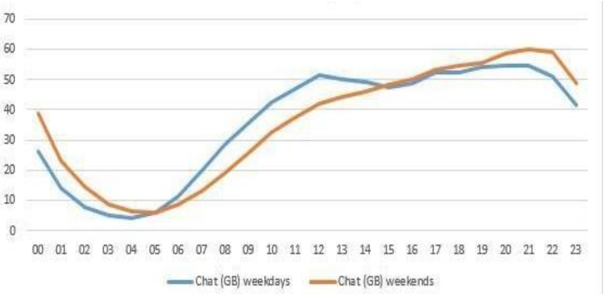

Fig. 6. Average volume of chat traffic

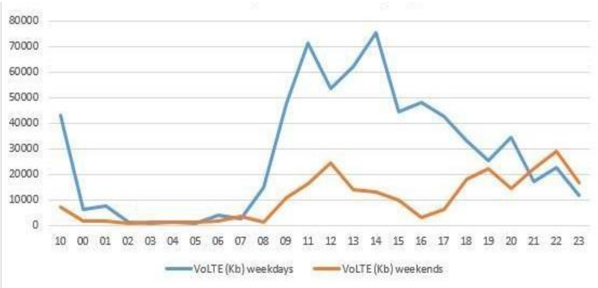

Fig. 7. Average volume of VoLTE traffic

The volume of traffic during weekends is higher for all traffic types. The difference occurs from 06:00 to 08:00, which is related to the beginning of the working day, and from 19:00 to 21:00, when the difference in traffic is negligible.

Generally, for all types of mobile data traffic it is evident that they exhibit the highest volume from 20:00 to 21:00 during working days. The situation differs for weekends where web and chat traffic exhibit peaks at 21:00, whereas social traffic reaches the maximum at 22:00.

5. MATHEMATICAL MODELLING OF MOBILE DATA TRAFFIC

Mobile data traffic analysis is a vital component for understanding network requirements and capabilities. Adequate traffic modelling is a basic condition for accurate capacity planning [7–10], [12–13]. The field still relies on the principles of Erlang and packet data modelling introduced several decades ago. The recent diversity of services makes the problem more complex as different packet data traffic types exhibit different patterns. This stems from the contextual nature of the different packet data services in use. It should be noted that there is no single traffic model, which can effectively capture the traffic characteristics of all types of networks under all possible circumstances.

The bursty nature of mobile data traffic results in periods with generation of high number of packets and periods with no encounters. Therefore, mobile data traffic can be best described by distributions that belong to long-tail distribution types such as Pareto, log-normal distribution, Weibull, etc. However, practical usage of these distributions is always limited by their complexity (usually reflected in the number of parameters they employ) and the level of accuracy they can sufficiently achieve.

This section introduces analytical modelling of mobile data traffic through curve fitting of appropriate distributions using the empirical measurements from the previous section. As mentioned, we limit to distributions that have lower computational complexity, but yield sufficiently accurate results for all types of mobile data traffic. As a result, the focus of the analysis is put on log-normal and normal distributions, which use 2 parameters, i.e. μ – mean value, and σ – standard deviation. These two distributions proved to provide sufficient results with the lowest Mean Square Error (MSE) with the empirical measurements in the real network.

5.1. *Curve fitting for PDF and CDF mModelling of mobile data traffic*

The analytical modelling relies on the accurate prediction of the Probability Density Function (PDF) and Cumulative Distribution Function (CDF) for all mobile data traffic types measured in section 4. For the analytical modelling we limit on two distributions which are simpler compared to other distributions and can only be modelled using two parameters: the mean and deviation (i.e. μ and σ respectively). Despite these two distributions there are more distributions which can be used for packed data traffic modelling, but due to their complexity are not used in this research leaving the possibility for future research. Still, the fitting shows that log-normal and normal distributions provide good approximations without much processing.

The PDF models the probability of occurrence of a certain traffic volume, whereas the CDF models the probability that a certain traffic volume with a given probability distribution will get a value less than or equal to x. In the following figures the *x*-axis value indicates the traffic volumes for each type of traffic in one day in GB, which are calculated as the average value of the traffic volumes per hour for two weeks. The traffic volume of each hour is the sum of the traffic during that hour. PDF and CDF are marked on the *y* axis respectively. The goal in the analysis is to fit curves using mathematical models with appropriate parameters that will most closely match the empirical PDF and CDF gained from the measurement campaign in the previous section. The curve fitting targets the minimal MSE. This section provides the results that led to the minimal MSE out of the many experimentations performed for the purposes of this research.

Figures 8 and 9 show the empirical vs. modelled PDF for mobile web traffic using log-normal and normal distribution, respectively. Figures 10 and 11 show the empirical vs. modelled CDF for mobile web traffic using log-normal and normal distribution, respectively. The *x*-axis value indicates the web traffic volume in one day (calculated as the average value of the traffic volumes per hour for two weeks). These figures give insight into the more suitable distribution for modelling mobile web traffic, which is the log-normal distribution.

Similarly, as with the previous case, Figures 12–15 show the empirical vs. modelled PDF and CDF for social networks traffic using log-normal and normal distribution. A key difference in this case is that both distributions can provide very similar results (low MSE) thus making the modelling of social networks traffic feasible with both of them.

Finally, Figures 16–19 show the empirical vs. modelled PDF and CDF for chat traffic using log-normal and normal distribution. The behavior of the chat traffic exhibits very similar characteristics as the mobile web traffic so the same conclusions apply.

This subsection presented the results of the modelling along with the estimation of the mean and the deviation for every mobile data traffic type (shown in the figures).

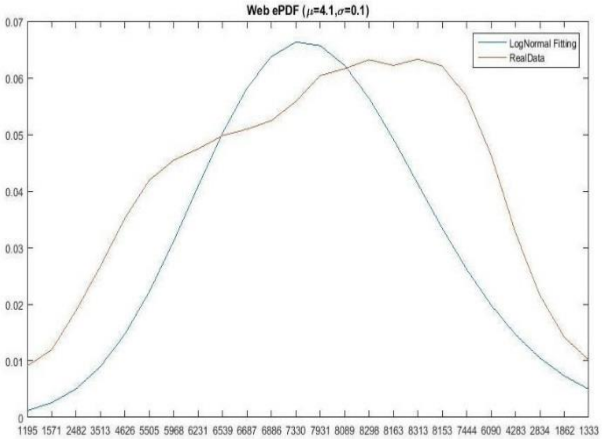

Fig. 8. PDF for web traffic using log-normal distribution

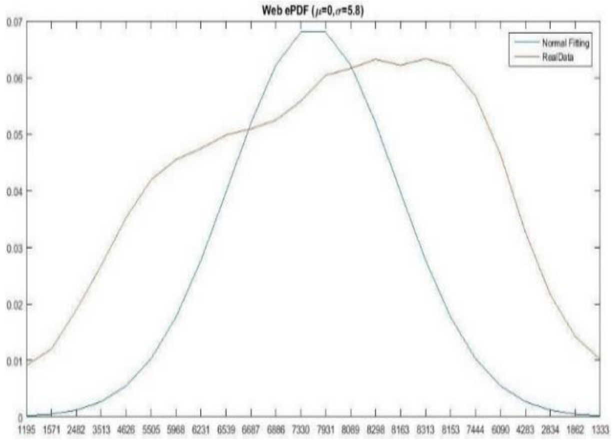

Fig. 9. PDF for web traffic using normal distribution

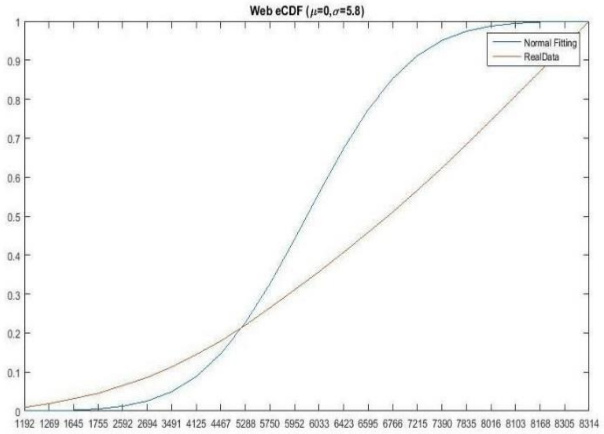

Fig. 10. CDF for web traffic using log-normal distribution

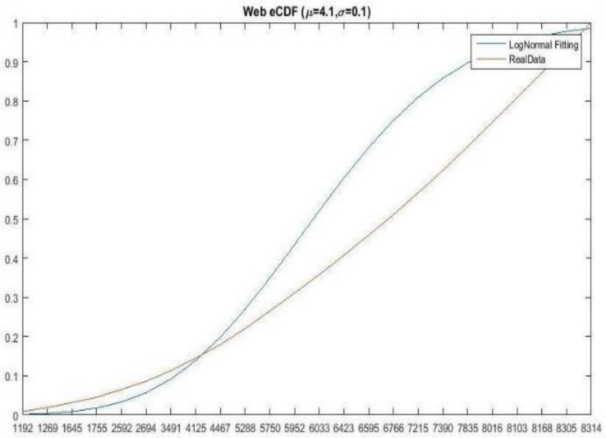

Fig. 11. CDF for web traffic using normal distribution

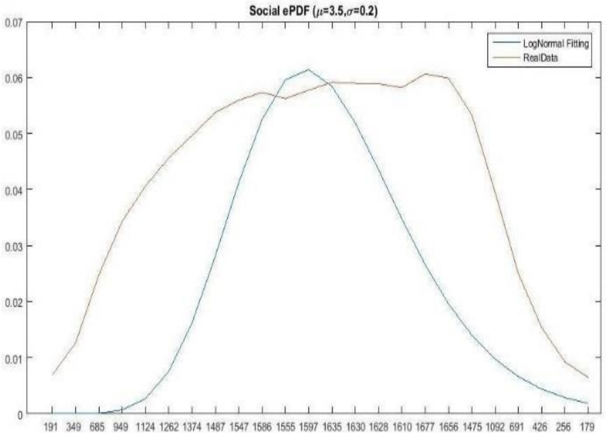

Fig. 12. PDF for social networks traffic using
log-normal distribution

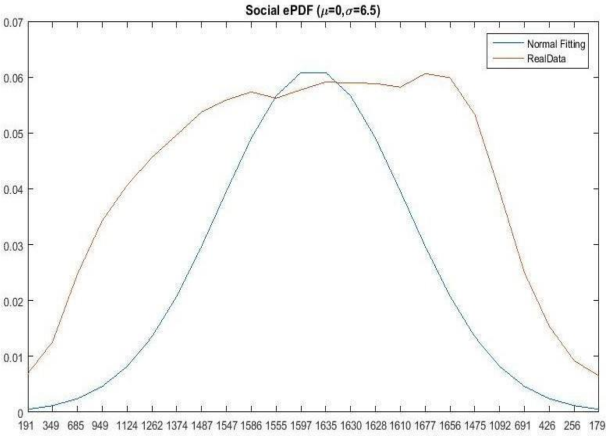

Fig. 13. PDF for social networks traffic using
normal distribution.

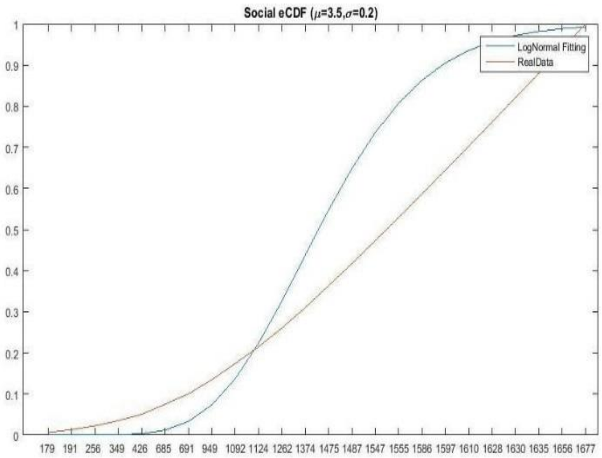

Fig. 14. CDF for social networks traffic using log-normal distribution

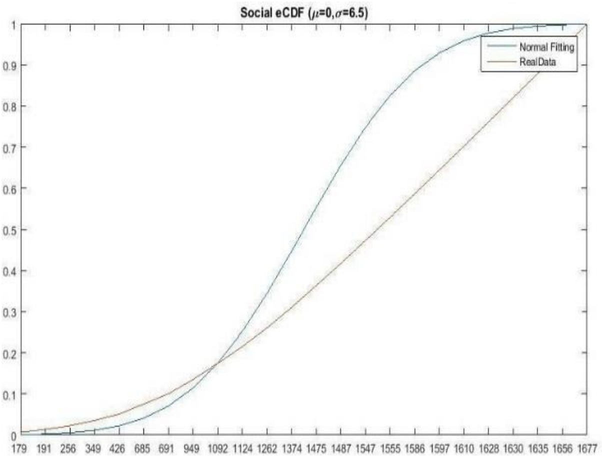

Fig. 15. CDF for social networks traffic using normal distribution.

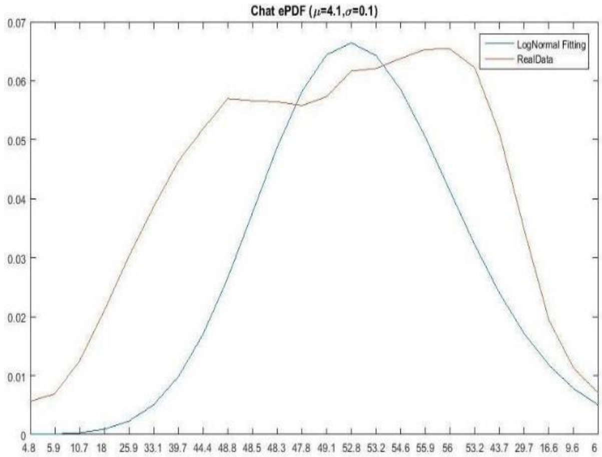

Fig. 16. PDF for chat traffic using log-normal distribution

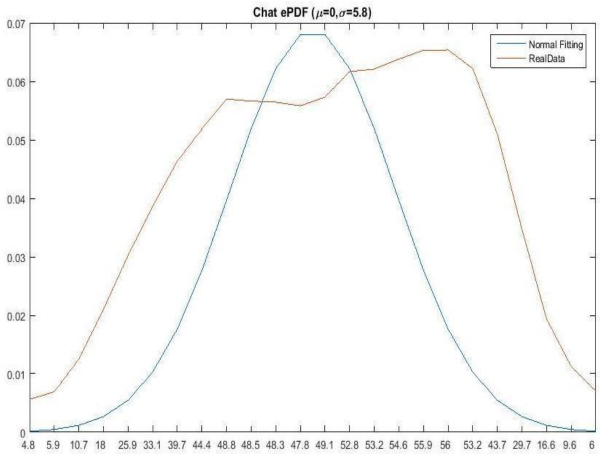

Fig. 17. PDF for chat traffic using normal distribution.

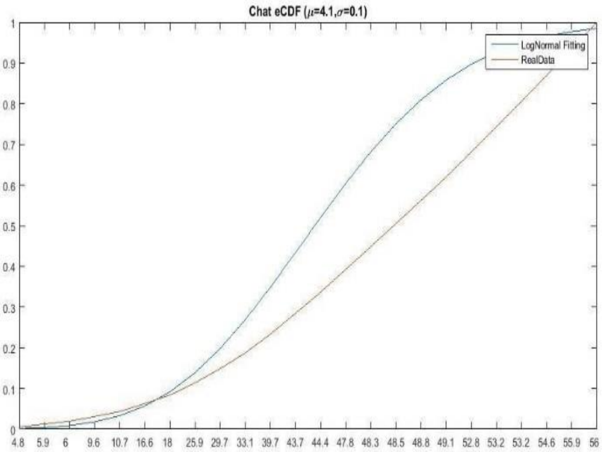

Fig. 18. CDF for chat traffic using log-normal distribution

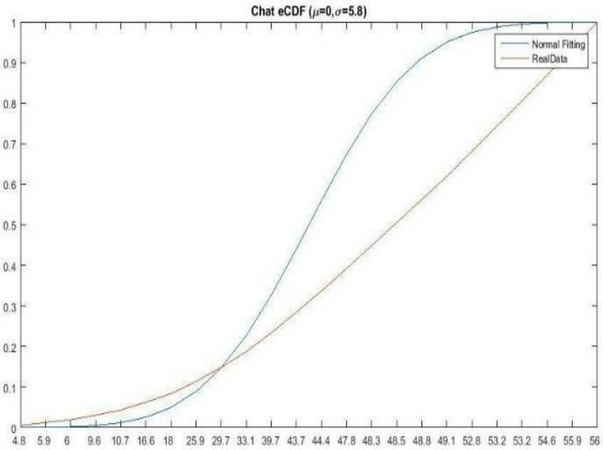

Fig. 19. CDF for chat traffic using normal distribution

5.2. *MSE Calculation*

The accuracy of the mathematical modelling in the previous subsection can be analyzed through the MSE, often used in error measurement statistics. MSE represents the error between the PDF / CDF of the proposed models with specific parameters for μ and σ and the PDF / CDF for real-time traffic distribution (measured). The MSE was calculated for every traffic type and both distributions. This value is always positive and values closer to zero are better. Table 1 summarizes obtained results for mean and deviation parameters that lead to the lowest possible MSE for every mobile data traffic type modelling.

Table 1

Estimation of μ and σ per distribution.

Traffic Type	Mathematical Model	Estimated parameters μ	σ	MSE
Web	log-normal PDF	4.1	0.1	0.000238
	normal PDF	0	5.8	0.000577
	log-normal CDF	4.1	0.1	0.0177
	normal CDF	0	5.8	0.0329
Social	log-normal PDF	3.5	0.2	0.00056
	normal PDF	0	6.5	0.000482
	log-normal CDF	3.5	0.2	0.0227
	normal CDF	0	6.5	0.0257
Chat	log-normal PDF	4.1	0.1	0.00039
	normal PDF	0	5.8	0.00061
	log-normal CDF	4.1	0.1	0.0196
	normal CDF	0	5.8	0.0345

The results shown in Table 1 clearly show that the MSE for web and chat traffic for both PDF and CDF is lower for log-normal distribution. On the other hand, in the case of social networks traffic, the MSE for the PDF is slightly lower in the normal distribution, while the MSE for CDF is slightly lower in the log-normal distribution.

As a result, it is evident that:

5.2.1. Web and chat traffic can be most accurately modeled with-lognormal distribution. The estimated parameters with which these two types of mobile data traffic can be mathematically modeled are $\mu = 4.1$ and $\sigma = 0.1$.

5,2.2. Social traffic is better modeled with normal distribution, with appropriate parameters: $\mu = 0$ and $\sigma = 6.5$.

It must be stressed that the modelling would be a continuous process with the increase of mobile data volume, increased service portfolio etc.

6. CONCLUSIONS

The virtualization of mobile packet core network is an ongoing process in operational mobile networks worldwide. For telecom operators and service providers, the return on investment related to the migration of network services in the cloud infrastructure comes in the form of automation of network operations.

This paper provides empirical measurements done for different mobile data traffic types in an operational nationwide mobile network with deployed virtual private cloud. The measurements show how and when mobile data users use the services. From the point of view of a telecom operator, the ability to know which are the most popular services and at what hours the users are most active is an opportunity to identify the amount of incoming and outgoing traffic and thus design a network that can support the expected activity.

Based on the results from empirical measurements, this paper offers mathematical modelling of measured traffic. We use log-normal and normal distributions to characterize the empirical results. These distributions provide sufficient compromise between accuracy and complexity and can aid the process of offline network planning and optimization.

Future work may include measurements in longer terms, experimentation with other mathematical distributions, etc.

REFERENCES

[1] ERICSSON: *Mobile Data Traffic Outlook*. Ericsson Mobility Report Data and Forecast. Available online at: https://www.ericsson.com/en/mobility-report/dataforecasts/mobile-traffic-forecast. Last retrieved 5.02.2021.

[2] Mageto, S. N., Balaji, N. V.: An overview of cloud computing and area of its application. *Karpagam JCS*, vol. **13,** (Sep. – Oct. 2019).

[3] Jain, N., Choudhary, S.: Overview of virtualization in cloud computing. *Symposium on Colossal Data Analysis and Networking (CDAN)*, India, 2016.
DOI: 10.1109/CDAN.2016.7570950.

[4] Choi, Y.-Ii., Kim, J.-H., Park, N.-I.: Revolutionary direction for 5G Mobile Core Network Architecture, *International Conference on Information and Communication Technology Convergence (ICTC)*, South Korea, 2016.
DOI: 10.1109/ICTC.2016.7763350.

[5] Adamuz-Hinojosa, O., Munoz, P., Ordonez-Lucena, J., Ramos-Munoz, J. J., Lopez-Soler, J. M.: Harmonizing 3GPP and NFV Description Models: Providing Customized RAN Slices in 5G Networks. *IEEE Vehicular Technology Magazine*, Vol. **14**, Issue 4 (December 2019).
DOI: 10.1109/MVT.2019.2936168.

[6] Tkachova, O., Jamal Salim, M., Raed Yahya, A.: An Analysis of SDN-OpenStack Integration. *Second International Scientific-Practical Conference Problems of Infocommu-*

nications Science and Technology (PIC S&T), Ukraine, 2015. DOI: 10.1109/INFOCOMMST.2015.7357 269.

[7] Wang, S., Zhang, X., Zhang, J., Feng, J., Wang, W., Xin, K.: An approach for spatial-temporal traffic modelling in mobile cellular networks. *27th International Teletraffic Congress*, Belgium, 2015. DOI: 10.1109/ITC.2015.31

[8] Svoboda, P.: *Measurement and modelling of internet traffic over 2.5 and 3G cellular core networks*, PhD Thesis Technical University in Vienna, 2008. Available on lineat: https://publik.tuwien.ac.at/files/PubDat_170755.pdf. Last retrieved at 5.02.2021.

[9] Hintze, D., Findling, R. D., Scholz, S., Mayrhofer, R.: Mobile device usage characteristics: The effect of context and form factor on locked and unlocked usage, *12th International Conference on Advances in Mobile Computing and Multimedia* (MOMM'14), Taiwan, 2014.

[10] Jin, L., Chen, Y., Wang, T., Hui, P., Vasilakos, A. V.: Understanding user behavior in online social networks: A survey, *IEEE Communications Magazine*, **59**, (9) (September 2013). DOI: 10.1109/MCOM.2013.6588663.

[11] https://www.etsi.org/deliver/etsi_gs/nfv/001_099/002/01.01.01_60/gs_nfv002v010101p.pdf

[12] Mohammed Alasmar, R., Clegg, N., Zakhleneiuk, G., Parisis: Internet traffic volumes are not gaussian – they are log-normal: An 18-year longitudinal study with implycations for modelling and prediction, *IEEE/ACM Transactions on Networking*, 2021.

[13] Amin S. Ibrahim, Khaled Y. Youssef, H,. Kamel, M., Abouelatta: *Traffic modelling of smart city internet of things architecture*, 2020,
DOI: 10.1049/iet-com.2019.12 52.