

Call Duration Characteristics based on Customers Location

Karolis Žvinys (*Doctoral Student, Vilnius Gediminas Technical University*),
Darius Guršnys (*Associate Professor, Vilnius Gediminas Technical University*)

Abstract – Nowadays a lot of different researches are performed based on call duration distributions (CDD) analysis. However, the majority of studies are linked with social relationships between the people. Therefore the scarcity of information, how the call duration is associated with a user's location, is appreciable. The goal of this paper is to reveal the ties between user's voice call duration and the location of call. For this reason we analyzed more than 5 million calls from real mobile network, which were made over the base stations located in rural areas, roads, small towns, business and entertainment centers, residential districts. According to these site types CDD's and characteristic features for call durations are given and discussed. Submitted analysis presents the users habits and behavior as a group (not an individual). The research showed that CDD's of customers being them in different locations are not equal. It has been found that users at entertainment, business centers are tend to talk much shortly, than people being at home. Even more CDD can be distorted strongly, when machinery calls are evaluated. Hence to apply a common CDD for a whole network it is not recommended. The study also deals with specific parameters of call duration for distinguished user groups, the influence of network technology for call duration is considered.

Keywords – Communications technology; Cellular networks; Communication system traffic; Telecommunication services.

I. INTRODUCTION

Mobile communications enlargement determines increased consumer's mobility and prevalence. Probably we will not find a man who does not use a mobile phone. Global trends show constantly increasing need of mobile data, however mobile calls remain the leading mobile network service even today [1]. Practically every mobile network user accesses this type of service, regardless of wherever it is. Usually it seems that the location of users should not have any impact on their habits to communicate. Nevertheless, the location of users is a very important factor that affects consumer's behavior using voice calls services [2]. Unfortunately, a clear dependency between user's location and call duration is not sufficiently examined until now. Generally, call duration as a key parameter is used, in assessing what is typical for mobile network voice service users [3] and [4]. This parameter helps to estimate the user's habits to communicate. Also it is very important for charging strategies, network optimization, performance and management analysis [5] and [6].

The purpose of this work is to represent the dependence of communication habits of mobile network users according to their location. This is made through analysis of customers call durations. The obtained results provide a much wider perception of the user's call duration distributions and at the

same time highlights the behavior of users located in a certain places. Besides data collected during the study showed that the machinery type voice calls in mobile networks can have a significant impact on the call duration distributions.

In order to accomplish the aim of work, further paper structure consists of comprehensive analysis of related works in Section II. The chapter provides a brief overview of the call duration analysis methods, CDD's. Also the relationship between the user's and the duration of voice call is discussed. Section III deals with detailed evaluation of CDD's related to specific user groups. Separate CDD's per individual group is introduced. Aside the influence of so called redirected calls to the distributions is shown. Remaining IV and V sections are specifically addressed to explore voice calls duration differences depending on the user's location or mobile network technology used for conversation. Finally, conclusions and other relevant facts are presented in Chapter VI.

II. RELATED WORKS

Many studies are prepared which examine the behavior of users and analyze the call duration distributions based on this parameter. Presented results evaluate the CDD's only per one user almost in all cases. Such call duration distributions are usually approximated using the lognormal, exponential, power-law or Erlang distributions [7], [8], [9] and [10].

Some of the researchers argue that the lognormal distribution of CDD corresponds to the most [10]. Another group believes that sometimes exponential distribution is more appropriate. The third group thinks that none of the mentioned distributions is accurate enough for CDD of individual user. As a result, they offer new models which are more suitable for assessment of CDD's. In [8], the authors present call duration distributions of individual users. For the analysis of these distributions TLAC (Truncated Lazy Contractor) model is applied, which is proposed by researchers. The obtained results claim, that TLAC model fits for more than 96 % of users data. Also the authors model the collective behavior of consumers based on the call durations and show that user's calling habits remain constant over time.

Meanwhile, Dong Y. et al investigate what is typical for call duration depending on the customer's social status, age, gender. They discovered that the stronger ties connect people the shorter calls they are tend to make. Also, authors found that opinion leaders, people with common neighbors prefer to talk shorter as well. While, for example, young females are prone to communicate much longer, especially in the evenings. Comparing men's and women's talking habits, it is

observed that women's are tending to make longer conversations. The latter habit is typical for youths too [9].

Another interesting observation is that a location having a large number of calls not necessarily has a large population or base of customers. It was noticed by a group of researchers from the Peking University in China, who have investigated urban population distribution based on mobile call data. In their study, they showed that call volume is more suitable in assessing consumer's population instead of traffic in Erlangs. According to the following measurement they received that the number of calls is directly proportional to the number of active mobile subscribers in a region. Distributing by regions, the results argue that early in the mornings the number of calls increased in rural areas sharply. While in urban areas the higher intensity is observed during the day time, due to business centers, and in the evenings or at nights, when people are going to the clubs [11] – [15].

Considering that what is discussed, it is clear that the researches regarding with the duration of customers mobile calls are not fully completed. Majority of studies based on consumer's call duration distributions analysis are linked with social relationships between the members of society or each individual is examined separately. Researches related to the user's call duration based on their location practically do not exist. The goal of this paper is to reveal the ties between user's voice call duration and the location, where the call was made. Our study includes the analysis of the users as a homogenous group, but not as individual user. This approach seeks to show that call duration characteristics for different groups of users differ. In addition CDD's of different consumer groups are not equal. Although they can well fit by the same function, but the parameters of such approximation are not identical.

We hope that this study will provide useful observations for those, who are interested in user's behavior based on mobile calls. The distributions of customers call duration can be helpful when network capacity issues are solved. Moreover knowing the probable behavior of consumer could facilitate tuning of parameters in mobile networks.

III. DATA OBTAINING

The data for this study was obtained from one of commercial operating telecommunications network in Lithuania. More than 5 million voice calls were collected. Those calls were collected from the base stations (BS) located at different places of Lithuania. The spread of BS's covered the territories of small towns, rural areas, roads, residential areas of the cities, entertainment and business centers. In that way the same user groups in this research were distinguished respectively. Collected call data consisted of a call start/end time, call duration and cell ID of the base station which was used for communication by the user's handset. The latter parameter allowed assigning the mobile calls to certain groups of users. In addition using the cell ID calls were separated into two groups: calls performed on 2G or 3G network.

The call duration distributions of mentioned user groups are shown in Fig. 1. They include all performed calls in 2G/3G base stations of selected areas.

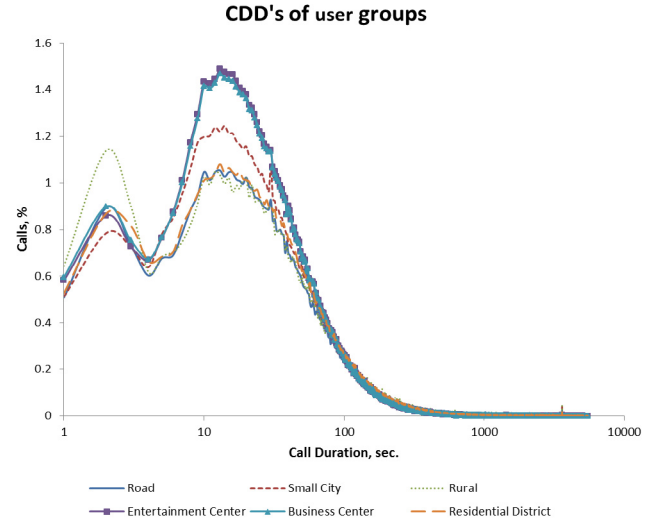


Fig. 1. Call duration distributions of different user groups.

Analyzing this type of distributions it is clear that we can not treat them as very simple because such a distribution could not be approximated by any well-known function. Hence investigating these distributions we are able to split them into two parts. As the first part of the distribution the part until its peak value could be considered while the second part is the tail of the distribution. This part of distribution could be described using the already well-known lognormal, power-law or others distribution functions. Mainly used is lognormal distribution function which probability density function is described as follows.

$$f(x | \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\ln x - \mu)^2}{2\sigma^2}\right\}; x > 0. \quad (1)$$

However, the density function is dependent, for example, on the location (μ) and scale parameters (σ) of distribution. These two parameters characterize the distribution and can be defined by the following formulas:

$$\mu = \ln\left(\frac{m^2}{\sqrt{v + m^2}}\right) \text{ and} \quad (2)$$

$$\sigma = \sqrt{\ln\left(1 + \frac{v}{m^2}\right)}, \quad (3)$$

where: x is call duration, but m and v are mean and variance.

This indicates that the tail of distribution could be well fitted with lognormal function, but the parameters will vary. This process is observed in the CDD's of the business and entertainment centers consumer groups, compared with others.

Going back to the first part of distribution, it must be said that it is more complex. It means that none of the aforementioned approximation functions is suitable for this part. This takes place because the beginning of distribution is

distorted. In Fig. 2 the graphs of call duration distributions are provided when lognormal function as approximation curve is used. It is obvious that whole analyzed user groups are characterized by a relatively strong distortion in the beginning of distribution. Considering that we separated two groups of CDD's, which are classified that urban and suburban areas in order to show, that the distortion is not dependent on location. First group in Fig. 2 (a) consists of CDD's from roads, rural areas, small cities and it is called suburban area. The other group consists of the data collected in densely populated areas and it includes residential districts, business and entertainment centers.

The results presented show that the lognormal distribution does not evaluate the first part of CDD completely. It only perfectly fits for the tail of users call duration distribution. So this proves that this type of distributions should be analyzed using more sophisticated approximation curves.

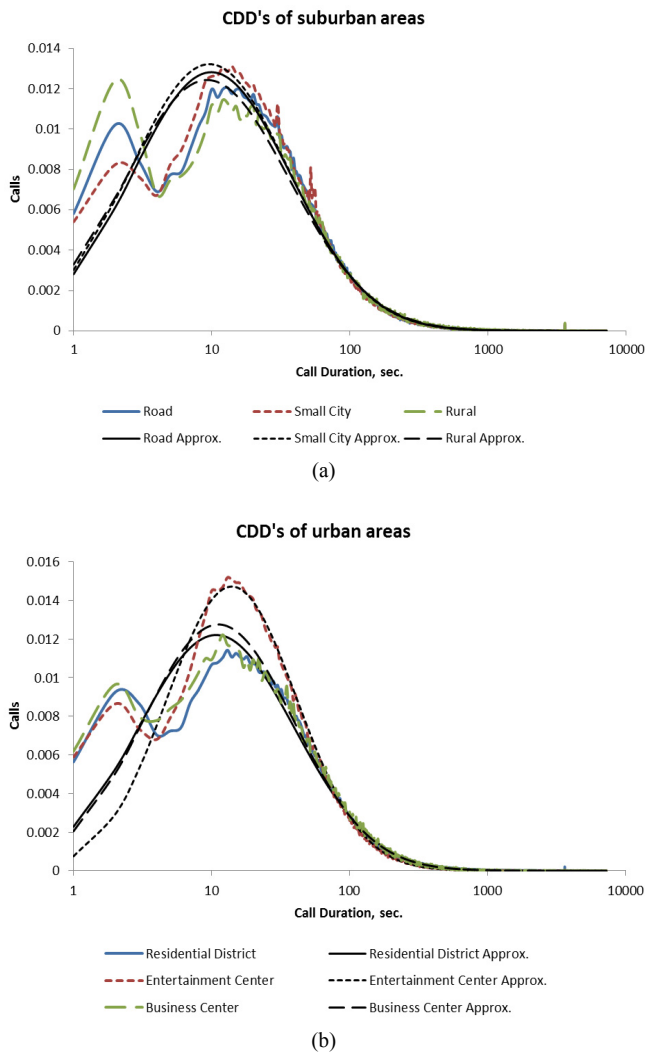


Fig. 2. (a) – CDD's approximation using lognormal distribution of customers from suburban areas. (b) – CDD's approximation using lognormal distribution of customers from urban areas.

Identified distortions of distributions have occurred due to redirected calls, which were included in the distributions. Practically, following calls are inherent for every mobile network. They are characterized by the fact that only takes place when the user is busy, unreachable or conditional forward is configured. Also short conversation time is typical for them. Therefore it is very important to evaluate these calls, because they represent a significant portion of all calls. The first part of the distribution including the redirected calls contains 11 – 15 % of calls. Redirected calls demand same network resources and create same load as ordinary voice calls. Usually similar type of studies does not evaluate such calls. They focus on the tail of the distributions. This is the reason why the data evaluating the beginning of distributions is limited.

As it is seen from Fig. 1, an impact of redirected calls is common to all groups of CDD's. Mostly these calls take about 2 – 3 seconds before the user hears the message and hangs up. That CDD is dependent on redirected calls could be seen from Fig. 3 (a), which shows the distribution without such calls.

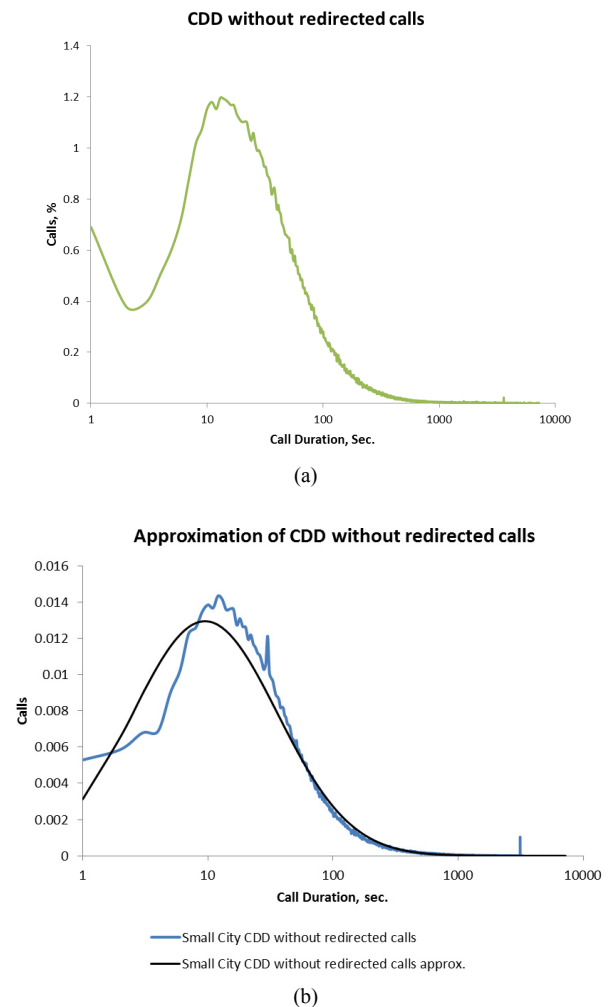


Fig. 3. (a) – Call duration distribution where redirected calls are eliminated. (b) – Approximation of CDD without redirected calls.

Submitted distribution depicts the CDD of the urban residential district user group. It contains 2G/3G voice calls performed on selected base stations in urban areas.

On the other hand it remains unclear whether eliminating redirected calls from CDD an applied approximation of lognormal distribution will be accurate. For this reason we selected a user group of small cities and removed more than 200 000 calls from it, which were identified as redirected calls. New CDD and its approximation using lognormal distribution of analyzed user group is shown in Fig. 3 (b).

According to the charts it is possible to state that the elimination of redirected calls from CDD helps to lognormal distribution to become more appropriate distribution for this type of data description. However, this cannot be said unambiguously, because initial part of the distribution does not match perfectly. This means that short calls in cellular networks also exist, but they consist of much smaller part compared with the whole amount of calls. Usually such calls are mentioned rarely. Therefore studies about the initial part of CDD estimation is not examined sufficiently

Summarizing, it is essential to estimate these changes when CDD's of users are analyzed. Moreover the distributions of various user groups are slightly different. Consequently common call duration distribution for a whole network cannot be applied.

IV. DIFFERENCES BETWEEN USERS BEHAVIOR

Previously mentioned, that the location of customers may definite influence over the use of voice services. Our analysis showed that the behavior of users being them in different locations is unequal. This diversity could be seen in Fig. 4.

Distinctly different consumer groups are concentrated in the entertainment and business centers. These two user groups are clearly different from the others, but they are very similar mutually. This is the reason why in Fig. 4 the CDF graphs practically overlap each other for these groups. Particularly, for them are typical large numbers of short calls. About 95 % of calls in business and entertainment centers are shorter than 6 minutes 50 seconds. Also the exclusivity of business and entertainment centers demonstrates the fact, that such user groups are tend to perform more short calls (<60 seconds) and their average call duration is much lower. This is visible in Fig. 5.

Respectively other user groups communicate longer. Especially it is noticeable in rural areas, small towns. Based on a submitted assessment the assumption comes up, that people being at home spend more time for mobile conversations. Thus, the claim that the people at their workplace spend the same amount of time for voice calls as in their homes is not always correct.

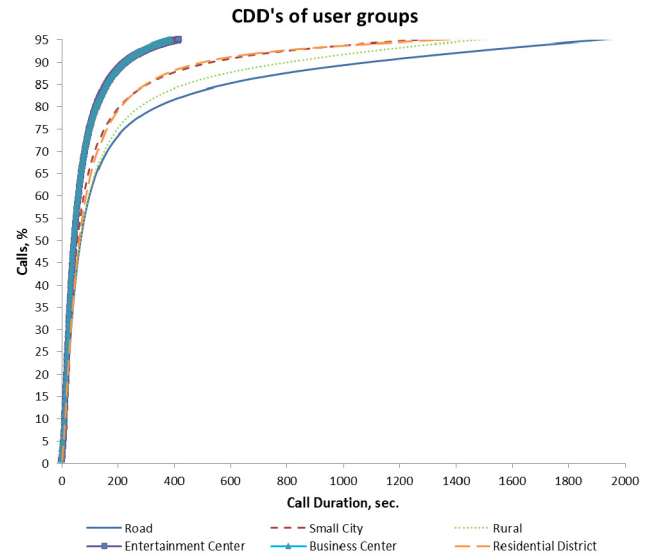
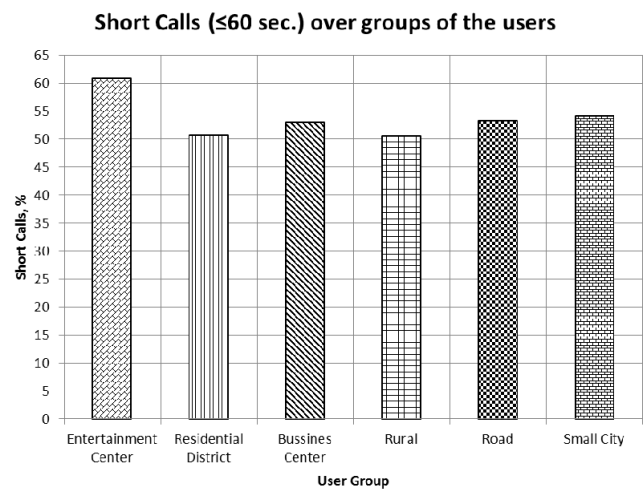
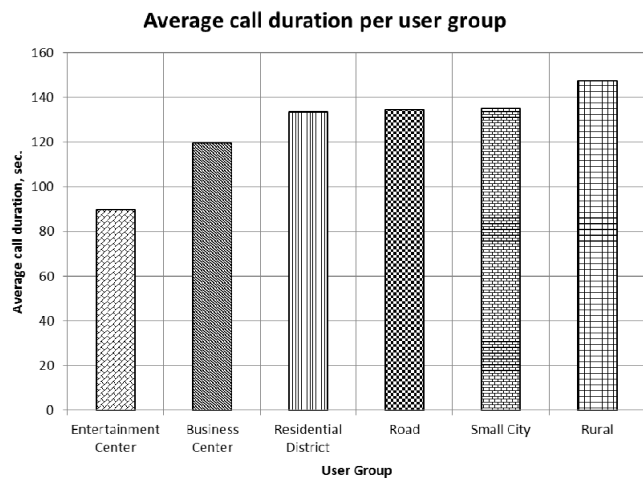


Fig. 4. Cumulative distribution functions of call durations for different user groups.



(a)



(b)

Fig. 5. (a) – percent of short calls (<=60 sec.) for a specific user group. (b) – average call duration for a specific user group.

Slightly different situation occurs to a road consumer group. On the one hand it looks that they are prone to talk longer. Otherwise considerable amount of short calls is specific to them. This happens due to the fact that at least one of base station sectors served small habitable zone. Considering the fact, it can be said that users, while they are driving cars are tend to talk less also. The same characteristics are typical for the long calls of this group as for the calls in rural areas.

V. CDD'S RELATED TO NETWORK TECHNOLOGY USED FOR VOICE CALLS

Another interesting aspect is how consumer's behavior depends on the network technology used for speech calls. This issue is becoming increasingly important because the prevalence of smart phone's increase with each passing day. Unfortunately, the opportunities to maintain unlimited capacity resources in telecommunications networks are limited. So the engineers of mobile networks are using various flow control mechanisms to improve network availability and service quality. The most significant are the voice call flow control algorithms. Normally these algorithms are based on the network load and resources availability requested by user before making a decision. The latter management feature is very important for controlling of voice call traffic between GSM and UMTS networks. It is triggered when a device applying to the network requests the RAB (Radio Access Bearer) only for voice call. Then a device is asking for resources in a 3G network it may be directly forced to make a call on a 2G network. This allows to control the traffic flow of users between the technologies and at the same time to guarantee the quality of other services. Of course, to enable this management, system should support 2G/3G communication technologies at same location. Our purpose was to investigate how the technology chosen for voice call impacts the user's behavior when this type of features is disabled. The call duration distributions of user groups related to network technology are given in Fig. 6.

Fig. 6 (a) depicts the CDD's of users who have performed the calls using only 2G network. The presented data show that there is no obvious difference between the joint and in Fig. 6 submitted distributions. All calls carried out in 2G network have no influence to make use of GSM technology mandatory. Only the natural causes, such as network coverage, the supported technology bands of mobile phones had influence for performing calls in GSM network.

In terms of the data in Fig. 6 (b) it is seen that situation with a CDD's of 3G calls is similar. As well as in the case of GSM network the calls on UMTS network were made without any additional corrections. Provided distributions represent the CDD's of 3G network for the following user groups: small cities, roads, entertainment centers and residential districts in urban territories. The reason is that the other two user groups simply did not have an access to 3G network. Comparing the obtained results with a previously discussed it is clear that the network technology used for voice call has no impact on the duration of conversation. CDD's highly depend only on the location of the customers.

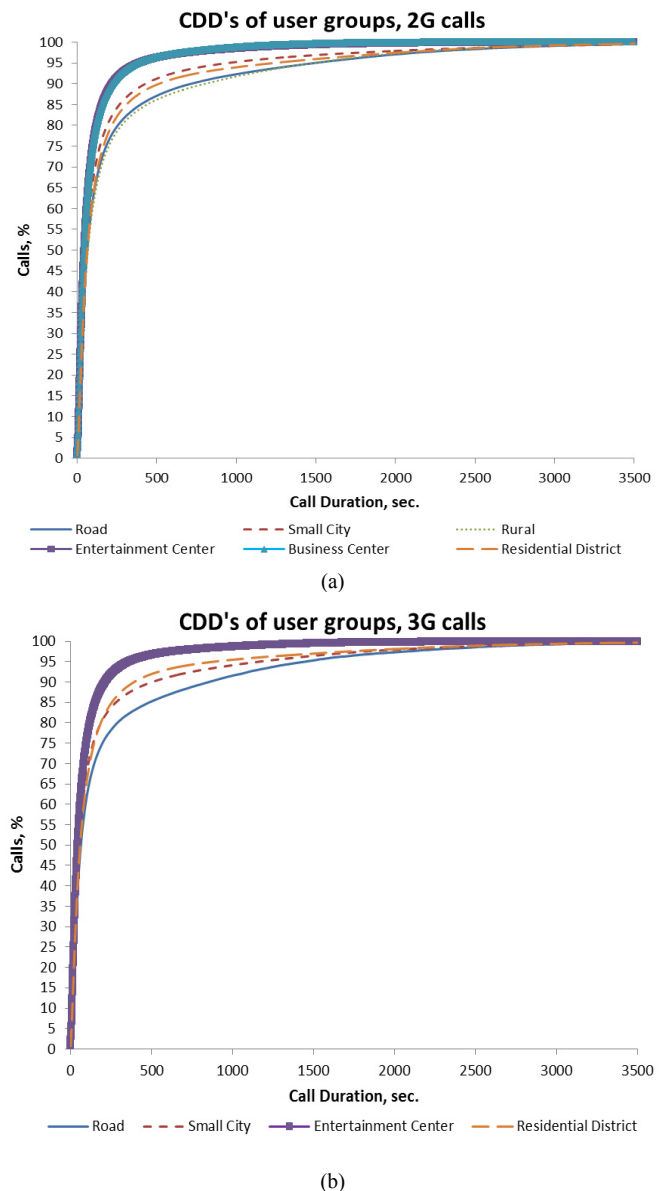


Fig. 6. (a) - Cumulative distribution functions of call durations for different user groups, when GSM network technology was used for voice call. (b) - Cumulative distribution functions of call durations for different user groups, when UMTS network technology was used for voice call.

The influence of a location to the distributions of calls performed accordingly in 2G/3G networks is shown in Fig. 7.

There is seen, that the larger is the evolved area the more calls are in 3G network. This is mainly typical for residential districts in urban areas and entertainment centers, which are also situated in cities. Whereas outside of major cities and in rural areas dominate a voice calls, which are made on 2G network. This situation occurs for several reasons. First of all, higher smart phone users penetration is observed in urban areas. Recently began evolution of smart phones penetrated into the business segment of cities primarily. Besides slightly higher financial opportunities of ordinary urban citizens, younger community impacted the increase of smart devices.

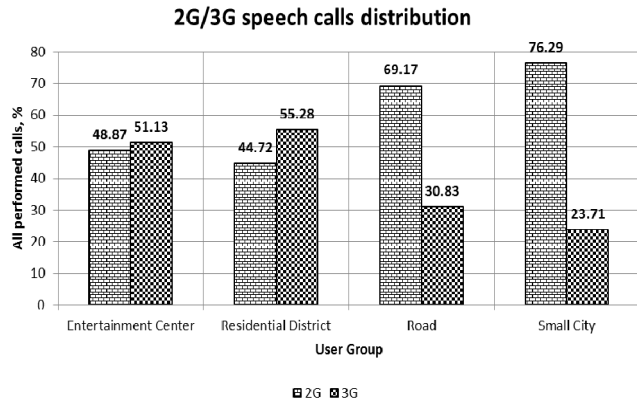


Fig. 7. GSM/UMTS voice call distributions according to network technology used for a call.

The other reason is that the citizens of rural areas are not so technologically advanced. Generally they do not feel the benefits of smart phones and they are not interested in new technologies too much. So it is not surprising that among these users the handsets supporting only 2G are still very common. The last one observation is that UMTS coverage holes in rural areas leads to increased number of GSM calls. It means that the peripheral of the country is not covered with 3G so good as urban areas.

The mentioned reasons affect the selection of technology which will be used to perform a voice call. However the technology used for speech call has no influence for duration of call. So this means that only the location of customer is important.

VI. CONCLUSIONS

The analysis showed clear differences between the call durations depending on the user's location. Created user groups led to a better understanding of what is typical for the majority of users, rather than individual user. This is very important, because a lot of active users are communicating in the mobile networks simultaneously. Presented users call duration distributions indicate the characteristics typical for certain user groups. The differences between them show that a general CDD for a whole network cannot be used, because it might be inaccurate. Also the CDD's strongly depend on the redirected calls. Assessing redirected calls the distributions become not so ordinary. Analysis of the distributions containing such calls showed that regular approximation curves for whole distribution are not appropriate. Eliminating redirected calls from CDD helps to lognormal distribution to become much more precise distribution for processing of call durations. For more accurate assessment of the distribution the composite distributions are required. This is one of the goals for further investigation.

In some cases, the influence of redirected calls to distributions can occur in different ways due to the differences between languages or message format of answering machine. This means that a user heard a musical message will understand that connection was not established much more

quickly. While voice format message requires more time for understanding that communication takes place amongst user and machine. It is also a compelling reason why the call duration distributions become more complicated. Unfortunately at this stage we had no opportunity to compare some of the received results with the other parties. Even so we have the reasons to believe that the consumers of the areas under analysis will not differ. This check is also our challenge for the future.

The research showed that the location of users is a significant factor, which impacts the user's communication time. User groups of entertainment and business centers are distinctly different comparing with others, because they are tend to perform short calls. Meanwhile users in rural areas or being at home are prone to interact much longer. This is an important factor in selecting certain settings of the base stations, in order to ensure adequate quality of service. In addition the study showed that the selected network technology on the base station has no influence for the call duration distributions. This has refuted the assertion that for the networks that support 3G HD voice feature longer voice calls is typical (this feature was enabled in the network). The user's location is much more important factor, when calls are performed.

Stated reasons and the results provided indicate that the habits and behavior of users are different. Therefore the planning and optimization of mobile networks is becoming more difficult. The engineers need to think not only about technical solutions, but they must assess the prospective customer's behavior depending on their locations also.

REFERENCES

- [1] ERICSSON, "HD Voice - It Speaks for Itself (ERICSSON White Paper)," 2011. [Online]. Available: <http://www.ericsson.com/res/docs/whitepapers/WP-HD-voice.pdf>. [Accessed: 10-Apr-2014].
- [2] T. Durga Laxmi, R. Baby Akila, K. S. Ravichandran, and B. Santhi, "Study of User Behavior Pattern in Mobile Environment," *Research Journal of Applied Sciences, Engineering and Technology*, vol. 4, no. 23, pp. 5021–5026, 2012.
- [3] V. B. Iversen, *Teletraffic Engineering (Handbook)*, Draft. Lyngby: Technical University of Denmark, COM Center, 2001, p. 321.
- [4] R. B. Cooper and D. P. Heyman, "Teletraffic Theory and Engineering," in *Encyclopedia of Telecommunications: Volume 16 - Subscriber Loop Signaling to Teletraffic Theory and Engineering*, F. E. Froehlich and A. Kent, Eds. CRC Press, 1998, p. 500.
- [5] A. Kivi, "Measuring Mobile User Behavior and Service Usage: Methods, Measurement Points, and Future Outlook," Helsinki University of Technology, Finland, 2007. [Online]. Available: <http://classic.marshall.usc.edu/assets/006/5570.pdf>. [Accessed: 11-Oct-2013].
- [6] G.-H. Tu, C. Peng, H. Wang, C.-Y. Li, and S. Lu, "How voice calls affect data in operational LTE networks," in *Proceedings of the 19th annual international conference on Mobile computing & networking - MobiCom '13*, 2013, p. 87.
- [7] V. B. Iversen, *Teletraffic Engineering and Network Planning*. Lyngby: Technical University of Denmark, 2010, p. 639.
- [8] P. O. S. Vaz de Melo, L. Akoglu, C. Faloutsos, and A. A. F. Loureiro, "Surprising Patterns for the Call Duration Distribution of Mobile Phone Users," in *Machine Learning and Knowledge Discovery in Databases (European Conference, ECML PKDD 2010, Barcelona, Spain, September 20-24, 2010, Proceedings, Part III)*, vol. 6323, Springer, 2010, pp. 354–369.

- [9] Y. Dong, J. Tang, T. Lou, B. Wu, and N. V. Chawla, "How Long Will She Call Me? Distribution, Social Theory and Duration Prediction," in in Machine Learning and Knowledge Discovery in Databases (European Conference, ECML PKDD 2013, Prague, Czech Republic, September 23-27, 2013, Proceedings, Part II), vol. 8189, Springer, 2013, pp. 16–31.
- [10] J. Guo, F. Liu, and Z. Zhu, "Estimate the Call Duration Distribution Parameters in GSM System Based on K-L Divergence Method," in 2007 International Conference on Wireless Communications, Networking and Mobile Computing, 2007, pp. 2988–2991.
- [11] C. Kang, Y. Liu, X. Ma, and L. Wu, "Towards Estimating Urban Population Distributions from Mobile Call Data," *Journal of Urban Technology*, vol. 19, no. 4, pp. 3–21, Oct. 2012.
- [12] S. Gao, "Human mobility, urban structure analysis, and spatial community detection from mobile phone data," University of California, USA, 2013. [Online]. Available: <http://www.slideshare.net/songgaogeo/song-gao-ucsbmobilephonedataanalysis>. [Accessed: 11-Oct-2013].
- [13] Center for Spatial Information Science (The University of Tokyo), "A Study on Urban Mobility and Dynamic Population Estimation by Using Aggregate Mobile Phone Sources," CSIS Discussion Paper No. 115. [Online]. Available: <http://www.csis.u-tokyo.ac.jp/dp/115.pdf>. [Accessed: 11-Oct-2013].
- [14] W. Loibl and J. Peters-Anders, "Mobile Phone Data as Source to Discover Spatial Activity and Motion Patterns," in in GI Forum 2012: Geovizualisation, Society and Learning, Berlin: VDE VERLAG GMBH, 2012, pp. 524–533.
- [15] C. Iovan, A.-M. Olteanu-Raimond, T. Couronné, and Z. Smoreda, "Moving and Calling: Mobile Phone Data Quality Measurements and Spatiotemporal Uncertainty in Human Mobility Studies," in in Geographic Information Science at the Heart of Europe, Springer, 2013, pp. 247–265.



Karolis Žvinys received the BSc in telecommunications engineering from Vilnius Gediminas Technical University in 2010, and the MSc at the same field in 2012. He is currently studying in Vilnius Gediminas Technical University as doctoral student in telecommunications engineering sector.

For last three years he is working for one of the major telecommunications operator, "Bite Lietuva", in Vilnius as a Radio Network Optimization Engineer. Previously working in "Teltonika" as a Test Engineer he was responsible for WiMax chipsets. His main research interests are in information and communication theory, mobility, resource management, fourth generation systems, network optimization.

Karolis is a student member of IEEE.

Postal Address: Naugarduko st. 41, Vilnius, Lithuania.

E-mail: karolis.zvinys@vgtu.lt



Darius Guršnys received the B.S., the M.S. and Ph. D. in electronic engineering from Vilnius Gediminas Technical University in 2001, 2003 and 2008, respectively. He is an Associate Professor in the Department of Telecommunication engineering at the Vilnius Gediminas Technical University.

D. Guršnys was the Head of Laboratory of Telecommunication engineering more than 5 years. Quality of service and the measurements of electromagnetic field intensity in the mobile communication networks is the main area of interests.

D. Guršnys is co-author of the Telecommunication theory workbook. Also he is associate Member of IEEE.

Postal address: Naugarduko st. 41, Vilnius, Lithuania.

E-mail: darius.gursnys@vgtu.lt