Analysis of User Activity in Wireless Local Area Network of Petrozavodsk State University

Aleksey Chuvak, Tatyana Surovtsova
Petrozavodsk State University
Petrozavodsk, Russian Federation
chuvakaleksei@yahoo.com, tsurovceva@petrsu.ru

Abstract—**Wireless networks are widely used for Internet access in public places, trading centers and educational institutions. The study of users' activity in these places can help solve problems of marketing, security, location, and monitoring of public transport. The article describes an experiment to measure the activity of users in wireless networks of Petrozavodsk State University because some educational and administrative processes are slowed because of a strict speed limit for all users. The authors developed a method to determine users' activity without the access to the service equipment and personal data, as well as a software complex to implement this method. Then they collected and analyzed data on the hours of the highest and lowliest activity, class schedule effect, and the dynamics of users' returns to the network. The information obtained was used to determine the allowed constraints of the Internet access speed for network users. A number of equations take into account the typical behavior of wireless network users, their activity and the capacity of external network channels. Computed rates for speed limits significantly exceed the established indicators in the network of Petrozavodsk University during hours of average and minimum network load. The results obtained can be used in corporate networks to dynamically calculate the permissible speed limit and to improve the quality of user service.**

I. INTRODUCTION

Wireless local networks, in particular Wi-Fi, are the most common technology to get access to information when mobility, ease of installation and use are required [1]. Therefore general public users with mobile devices may have access to wireless networks (WN). Such users usually login for short periods at specific times and often change their location [2], [3].

Analysis of users' behavior data may give necessary information to a wireless connection provider. For instance, such analytics helps to evaluate shop-windows customer appeal and advertizing campaigns effectiveness for a trading center [4]; to monitor students' attendance in class [5] or a specific event [6] for an educational institution. Such analytics is necessary for corporate security systems as well [7]. The obtained data may be used to optimize physical space [8]. Any suspicious activity in wireless networks must be recorded [9]. In addition, wireless networks could be used for objects localization in a room [10], [11], [12], [13]. Work with active mobile users could help to monitor vehicles routs [14]. Some studies indicated that individual people could be recognized via analysis of wireless network signal activity [15].

Petrozavodsk State University (PetrSU) provides Wi-Fi-based wireless connection to students, employees and guests. University WN is divided into several logic networks with various access modes: PSU-STUD for students, PSU-STAFF for employees and IT-GUEST for guests. Hundreds of people use the network every day therefore the necessary limitation of the maximum network capacity for each user results in inconveniences in education materials generating.

The purpose of this study was to collect data on WN users activity in PetrSU and to apply the obtained information for calculation of allowed rate constrains.

II. COLLECTION OF USERS' ACTIVITY DATA

Wireless networks in 13 buildings of PetrSU were accessed via more than one hundred access points Cisco Aironet 1602, 1040, 1240, each of them allowing connection to all logical WNs. A user who connected to any access point was allocated an IP-address from the integrated logical WN of IP-sub-network. Maximum number of users in each network was 1022.

The data receiving framework is shown on Fig. 1. There is a router in the network center, which is also a DHCP server. Is connected with a lot of communicators, which are wired to access points. Access points are also connected with mobile devices that get an IP address from the router using DHCP. Likewise the scanning device should be wired to the access point, get an IP address and then challenge all the sub-network using ARP protocol.
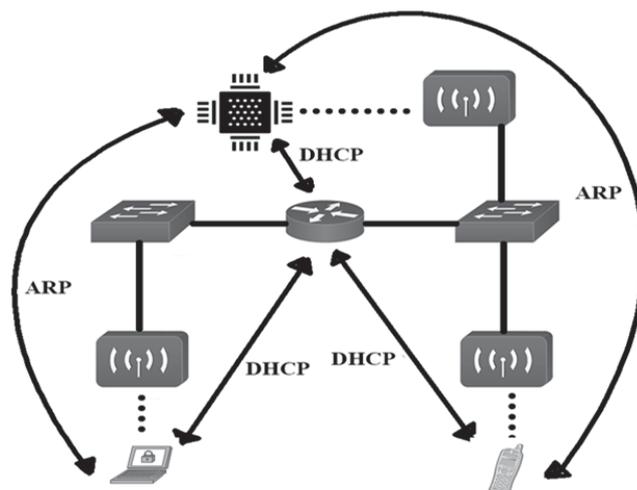

Fig. 1. Data receiving network

However, we should note that MAC-address does not always identify the user's device clearly. Modern facilities of the Apple (iPhones, iPads and MacBooks) change their MAC-address randomly with every connection. Nevertheless OUI remains the same. Thus when considering the number of unique users and returns of certain users in the university after some period of time, it is necessary to make an adjustment on the Apple devices with OUI.

Unfortunately, since all the computers are defined to the same IP-subnet, when using data collection method, it is impossible to determine to which exactly access point a user is connected. It means that the user's location cannot be identified without the user's permission or without administrative access to the network equipment.

Access to serving information on the network operation is restricted for security reasons. Otherwise it is possible to determine the number of active users through the protocol ARP (Address Resolution Protocol). By polling all possible IP-addresses in a sub-network with ARP-queries we might find out that the address is allocated to an active device (such device will give ARP-response) and determine MAC-address of the device for its identification. This technique is workable in case of connection to the studied IP-sub-network.

To collect data we developed a Python (Python features a dynamic type system and automatic memory management) application that uses utility Nmap [16] for network scanning using ARP- queries and automates computer switching between networks to simultaneous monitoring of several networks operation from one computer. The obtained data: date and time, number of active users, list of found MAC-addresses were stored in separate JSON- files.

Specifically for this purpose it was agreed to acquire a microcomputer "Orange Pi Lite" (Fig. 2). This microcomputer is a general-purpose computer which allows implementing arbitrary types of calculations. Likewise, it has a wireless access controller and an antenna and can be used for scanning of the wireless networks. The computer works on the Linux operating system and allows using such required tools as Nmap, Python and Network Manager. Due to portability of the computer it was possible to leave it in the university for the duration of an entire experiment including night time.

The experiment of data collection on user activity was performed from February 6 to March 26, 2017. For our research this time interval was enough. However, the accuracy can be increased with constant data collection.

A single-board computer Orange Pi Lite with Linux-based operating system used as a scanning device was installed in the main building of PetrSU for round-the-clock data collection. Approximately 8900 snapshots were made in each network with intervals from 5 to 10 minutes depending on the time of day.

In addition, certain patterns of the number of active users from time were defined. For this purpose we used regression analysis.
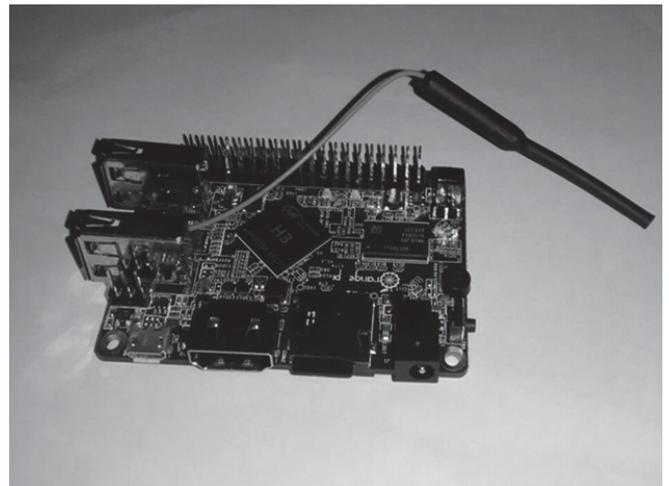

Fig. 2. Microcomputer «Orange Pi Lite»

The function that has been gained from regression analysis helps not only with identifying of the general nature of a dependence of user's activity on time, but also with the forecasting an exact number of active users at some point of time with a certain probability.

In connection with a supposed reliance nature it is necessary to use curvilinear regression methods. For that we used the least-square method and as a regression function we took polynomial $A(t) = b_0 + \sum_{i=1}^{k} b_i t^i + \varepsilon$ where A is the time-correlated function, t is time, b is the regression coefficient, k is polynomial order and ε is a random component.

Search of an approximate function was implemented using NumPy and SciPy libraries and Python, specifically the function of the least squares from a "scipy optimize" module.

The time of day was denoted by real numbers 0-24, which means that A(t) function was defined on the [0, 24] interval. The dimension of the polynomial to approximate function to the collected data has been selected for each network. The primary optimality indicators of an approximation are the optimality indicator as well as the appearance of the function graph in a comparison with the collected data.

The function graphs for the weekdays are shown in Fig. 3-5. These graphs illustrate indicators of the collected data and yield a curve from approximation. The vertical axis indicates the number of active users; the horizontal axis shows time periods.

III. ACTIVITY ANALYSIS

The obtained data demonstrated that the activity was similar in all networks on week-days. The activity growth started at 7:30 a.m. and reached a peak from 10:00 a.m. to 1:00 p.m. During lunch time (13:00 – 13:30) the activity fell down by 30%, then it returned to the previous level. After that the activity gradually decreased from 16:00 to 19:00. Fig. 6 shows dependency graph on time periods in PSU-STUD – network on a typical week-day 7 February 2017.
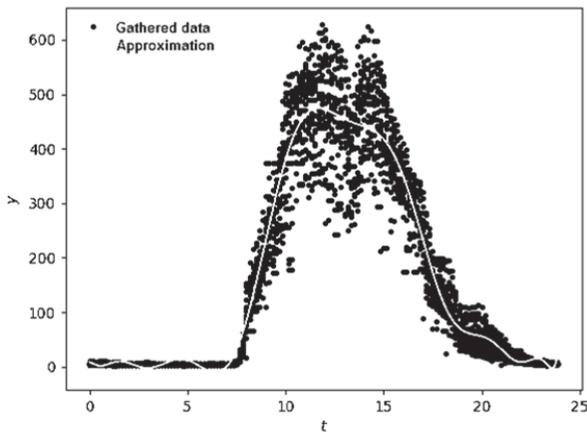
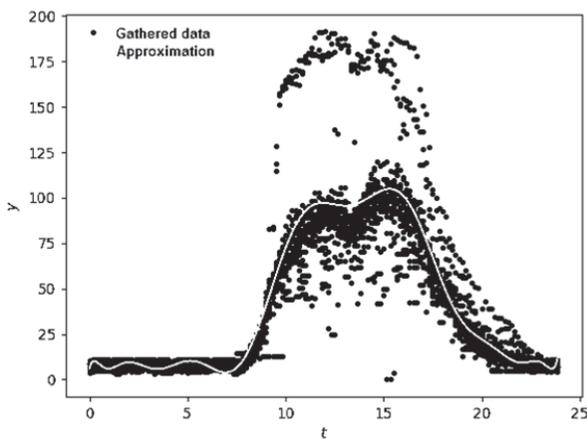Fig. 3. Approximate function graph of PSU-STUD activity (weekdays)

Fig. 4. Approximate function graph of PSU-STAFF activity (weekdays)
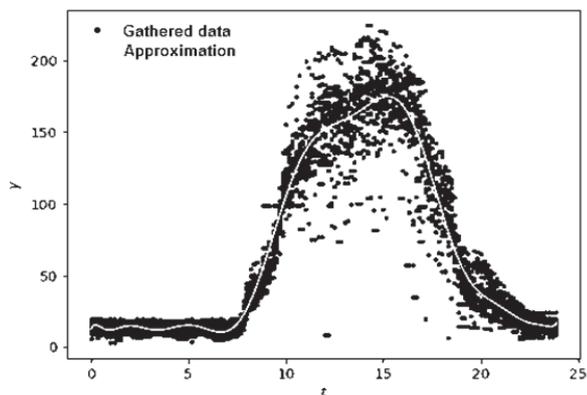

Fig. 5. Approximate function graph of IT-GUEST activity (weekdays)

Maximum number of active users on week-days was 450-650 people in PSU-STUD – network; 100-120 people in PSU-STAFF network and 160-230 people in IT-GUEST-network. The total maximum number of users was 926 for all three networks on the 22nd of March at 14:05. The activity indices were almost similar for neighboring week-days and for the first and second shifts.
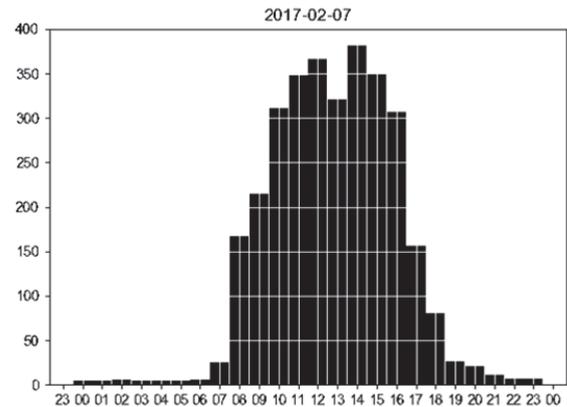

Fig. 6. Activity histogram in PSU-STUD on February, 7, 2017

Knowing MAC-addresses we were able to study the activity of individual users. There were 5435, 888 and 2182 devices in PSU-STUD, PSU-STAFF and IT-GUEST networks, respectively. Some users were active in several networks. The pair of networks PSU-STUD/PSU-STAFF was used by 116 people, PSU-STAFF/IT-GUEST was used by 152 people, and 31 users were active in all three networks.

Fig. 7 shows the graph of users continuous returns for PSU-STUD network. Week-days dates are shown in the horizontal direction and the percentage of the previous day users who were active on the following day are shown in the vertical direction. The graph shows the share of users who were active in the networks for several days in a row. The returns dynamics was similar for all networks and varied within 10-15%.

The obtained data demonstrate regular activity of users and smoothness of the education process on week-days as the user activity and the numbers of returns do not change.

IV. ACCESS SPEED LIMITATION FOR NETWORK USERS

The study of PetrSU users' activity levels revealed certain patterns, such as fluctuation of the number of users during the day and periods of time of minimum and maximum activity. Besides, the studies identified the functions describing users' dynamics throughout the day.

Access speed is limited for the users of PetrSU network because the external channel traffic for the university network is 150-170 Mbit/s and one or several users may not fully occupy the external channel thus hindering access for other users. Such speed limitation is the simplest one [17]. The speed is limited with server FreeBSD and network firewall ipfw. A limit of 1 Mbit/s is established for all network sites.

Actual speed is 400-500 Kbit/s during maximum load hours and 500-700 Kbit/s during average load hours (at 9 a.m. and 5 p.m.). The number of users who use a fixed connection is known, so we can allocate a given bandwidth for a wireless network. A limitation of 100 Mbit/s for a wireless network was used for calculations. All calculations can be performed for the dynamically changing channel capacity in the future.
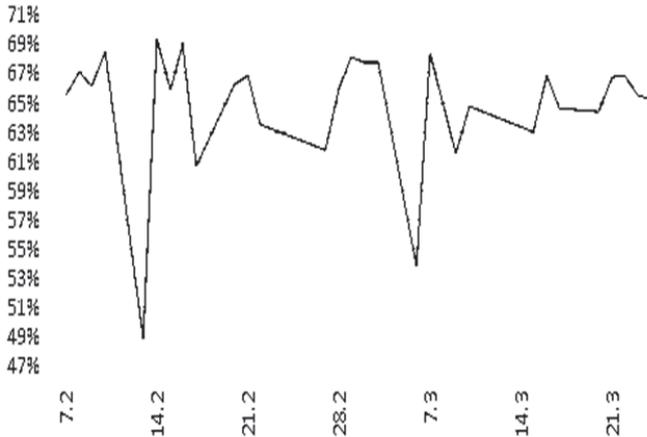

Fig. 7. Continuous returns for PSU-STUD

Evidently not every user in the network uses the entire available traffic at any moment. Researchers [18, 19, 20] demonstrated that network users could be divided into three groups according to the traffic usage: 17% of people use approximately 1-5 MB during the session (5.5 Kbit/s, in average), 73% use average traffic at the speed not more than 100 Kbit/s, and 10% use the maximum of the available traffic.

Based on the obtained data the following speed limit indicators can be identified:

- Permissible limitation based on the largest possible number of users in all the networks.
- Permissible limitation based on the largest known number of users in all the networks.
- Function calculating permissible limitation based on the function of dynamic of users' activity.
- Indicators similar to previous items, taking into account statistics of a traffic consumption.

Three of the biggest wireless networks in PetrSU are IT-GUEST, PSU-STUD and PSU-STAFF can be connected with up to 1022 devices simultaneously (see above). Thereby altogether it is possible to have 3066 devices connected simultaneously.

Assuming that users occupy all the data traffic capacity, then the permissible speed limitation per wireless user can be calculated using the formula below:

$$S = \frac{W}{N}$$

Where S stands for the speed limitation in Kbit/s, N is for the total number of users, W is the data capacity of an external Internet channel in Kbit/s. Applying the known data we result with a limitation of 32.62 Kbit/s per user.

Considering the fact that during the research the maximum number of users (3066) has not been reached, we can calculate the formula using a maximum number of active users discovered within the study: 926 users in total for three networks. In this case we result with a limitation of 108 Kbit/s.

Besides, our studies revealed that the maximum number of active users is reached at 11:00 a.m. and 14:00 p.m. During the rest of the day it can decrease dramatically. Assume that we have the number of users at a particular moment in each network, as well as the fact that speed limitation can be regulated dynamically; permissible limitation may be calculated via the formula below:

$$S(t) = \frac{W}{N_1(t) + N_2(t) + N_3(t)}$$

Where t is time, S(t) stands for a speed limitation depending on t in Kbit/s, $N_1(t)$, $N_2(t)$, $N_3(t)$ is the number of active users at a particular moment in the networks PSU-STUD, PSU-STAFF, IT-GUEST respectively.

Applying the data obtained above for the N function, we can get the desired result and calculate speed limitations for every user.

Let us calculate the channel width U in Kbit/s that is used by users of 1st and 2nd groups together at the total number N(t), depending on time:

$$U_{1,2}(t) = N(t) \times 0,17 \times 5,5 + N(t) \times 0,73 \times 100 \quad (1)$$
$$= N(t) \times 73,935$$

Let the maximum available width of the external channel be W then the channel width occupied by the users of the 3rd group will be:

$$U_3(t) = W - U_{1,2}(t) \quad (2)$$

At the specified division the used traffic-carrying capacity for the 3rd group of users can not be lower than that for the 2nd group of users. When the number of users reaches some limit the users of the 2nd and 3rd groups will work at the similar maximum speed. Let us calculate the maximum number of users when formulas (1) and (2) are valid:

$$W = N \times 73,935 + N \times 0,1 \times 100$$

For the number of users we will obtain:

$$N_{max} = \frac{W}{83,935}$$

If to take W as 100 Mbit/s then we will get $N_{max} = 1191$. If exceeded N_{max} for the 2nd and 3rd groups will be joined.

Up to this moment the speed is actually limited only for the 3rd group. Speed limitation S is derived from the channel bandwidth remainder when $U_{1,2}$ is taken into account:

$$S(t) = \frac{U_3(t)}{N(t) \times 0{,}1} = \frac{W}{N(t) \times 0{,}1} - 739.35 \qquad (3)$$

In the PetrSU network $N(t) = N_1(t) + N_2(t) + N_3(t)$, i. e. this is the sum of all network users. Formula (3) allows for calculation of the speed allowed constraint in Kbit/s in the wireless network dependence on time taking into account the number of users and their typical behavior. According to these calculations it is possible to set the constraint of 17.89 Mbit/s for the average number of 53.67 users at 7:30 a.m. on weekdays, the constraint of 1.46 Mbit/s for 454.68 users at 17:00, thus significantly exceeding the experimental data.

It should be noted that the optimal size of user groups and the traffic they use is individual for each network and their calculation requires further research. In view of the fact, we take the formula which considers sizes of users group and channel widths used by them:

$$S(t) = \frac{W - (N(t) * P_1 * C_1 + N(t) * P_2 * C_2)}{N(t) * P_3} \qquad (4)$$

In some cases there could be defined more than 3 strongly pronounced user groups. In this way it is necessary to generalize formula 6 for usage with the multiple groups:

$$S(t) = \frac{W - N(t) * \sum_{i=1}^{n-1} P_i C_i}{N(t) * P_n} \qquad (5)$$

In this formula we expect availability of n user groups, for group №"n" we expect an occupation of an entire bandwidth. P_i is the size of the group in percentage, C_i is the average throughput of the users from group i. The formula below helps calculating the maximum amount of users, for whom formula 5 makes sense:

$$N_{max} = \frac{W}{\sum_{i=1}^{n-1} P_i C_i} \qquad (6)$$

If the number of users exceeds N_{max}, then the last two groups unite:

$$S(t) = \frac{W - N(t) * \sum_{i=1}^{n-2} P_i C_i}{N(t) * (P_{n-1} + P_n)} \qquad (7)$$

For formula 7 the maximum number of users is calculated using the formula below:

$$N_{max} = \frac{W}{\sum_{i=1}^{n-2} P_i C_i} \qquad (8)$$

Compiling formulas 7 and 8 we get:

$$N_{max}^k = \frac{W}{\sum_{i=1}^{n-k} P_i C_i} \qquad (9)$$

$$S^k(t) = \frac{W - N(t) * \sum_{i=1}^{n-k} P_i C_i}{N(t) * \sum_{i=n-k+1}^{n} P_i} \qquad (10)$$

Formula 10 allows calculating the speed limitation that will limit the speed of k groups. Formula 9 allows calculating the maximum total number for that k which let formula 10 to make sense.

Thus formula 4 is a particular case of formula 10 where n = 3 and k = 1.

Grouping of the users and usage of generalized formulas 9 and 10 will make us reach the maximum efficient speed limitation and provide them with the best network experience.

The speed limitation materializes with the server "FreeBSD", firewall ipfw and "dummynet" subsystems. Using the given formulas and collected information we can set a new limit at the specific periods of time using the following command in the command interface:

```
ipfw pipe P config bw SKbit/s
```

Here P is the number of the virtual channel for a traffic control (pipe). S is the set speed limitation in Kbit/s. Limitation can be calculated by formula 3. As a data input we can use multinomial formulas that were calculated earlier. In order to restrict sub-network where users are set with a limit, the following channel should be created:

```
ipfw add pipe P all from A.B.C.D/M to any
```

Here A, B, C, D is a sub-net address; M is a sub-net mask in the format CIDR.

Constant update of the data input allows us to get the most accurate calculation of the speed limitation and provide users with the best network experience at any point of time.

V. CONCLUSION

Nowadays wireless public networks are widespread. Access can be provided in the shopping malls, railway stations, public and educational institutions, etc. Studying the behavior and activity of users of organizations may provide them with crucial information, which is important to ensure the best quality service, load balancing and other tasks.

Research of users' activity in a wireless network may result in more effective management of the organization providing access to this network. Therefore general public users with mobile devices may have access to wireless networks (WN).

The experiment on users' activity in the networks of Petrozavodsk State University demonstrated distinct regularities in users' behavior: activity peaks between 11:00 and 13:00 and 14:00-15:00, activity decay during lunch time, regular activity of 70% users on neighbor days. Basing on the obtained data we proposed a method of speed allowed constraint that allows constraint calculation depending on users' activity and their typical behavior. The developed application and corresponding setting of PetrSU server for speed limitation will allow us to correct the constraint according to the current network load and analysis of users

behavior data may give necessary information to a wireless connection provider.

REFERENCES

[1] Gudin M., Fedorov M. Nekotorye tendentsii razvitiia tekhnologii Wi-Fi, WiMax, deistvuiushchikh na chastote 2, 4 GGts [Some tendencies of developing of Wi-Fi, WiMax working at frequency 2,4 Gz]. *Besprovodnye tekhnologii.* – 2006. – no. 2. – pp. 4-6.

[2] Manweiler J. et al. Predicting length of stay at wifi hotspots. *INFOCOM, 2013 Proceedings IEEE.* – IEEE, 2013. – pp. 3102-3110.

[3] Zola E., Barcelo-Arroyo F. A comparative analysis of the user behavior in academic WiFi networks. *Proceedings of the 6th ACM workshop on Performance monitoring and measurement of heterogeneous wireless and wired networks.* – ACM, 2011. – pp. 59-66.

[4] Ivanov D. A., Surovtsova T. G., Tiakhti E. A. Analiz aktivnosti v besprovodnykh setiakh kak innovatsionnyi metod izucheniia povedeniia pokupatelei v torgovykh tsentrakh [Analysis of activity in wireless networks as an innovative method of studying of behavior of trade center]. *Nauchno-tekhnicheskie vedomosti Sankt-Peterburgskogo gosudarstvennogo politekhnicheskogo universiteta.* – 2013. – no 183-2. – pp. 271-275.

[5] Khrabrov D. E. Metodika kontrolia poseshchaemosti studentov i lokal'nogo pozitsionirovaniia na osnovanii Wi-Fi seti [Methods of control of student's attendance and location based on Wi-Fi network]. *Informatsionnye tekhnologii i sistemy 2015 (ITS 2015): materialy mezhdunarodnoi nauchnoi konferentsii*, 2015. – pp. 236-237.

[6] Marchenkov S. A., Korzun D. G. Presence Detection in SmartRoom: Experimental Performance Evaluation. *Proc. 15th Conf. Open Innovations Framework Program FRUCT.* – ITMO Univeristy, 2014. – pp. 221-222.

[7] Kamaikin A. G., Osipov I. E., Shumarin O. E. Korporativnye Wi-Fi [Enterprise Wi-Fi networks]. *Tekhnologii i sredstva sviazi.* – 2006. – no. 1.

[8] Eldaw M. H. S., Levene M., Roussos G. Presence analytics: Discovering meaningful patterns about human presence using wlan digital imprints. *Proceedings of the International Conference on Internet of things and Cloud Computing.* – ACM, 2016. – pp. 53.

[9] Loiko M. V., Ovchinnikov A. L. Setevye analizatory besprovodnykh setei. Bezopasnost' dostupa k lichnoi informatsii cherez seti Wi-Fi [Network analyzers of wireless networks. Safety of access to personal data over Wi-Fi network]. *Vostochno-Evropeiskii zhurnal peredovykh tekhnologii.* – 2012. – vol 3. no. 9 (57).

[10] Malodushev S.V., Rogov A.A. Opredelenie lokatsii v korporativnykh Wi-Fi setiakh [Definition of location in enterprise Wi-Fi networks]. *Vestnik IuUrGU. Seriia «Matematicheskoe modelirovanie i programmirovanie».* 2016. vol. 9, no 1. pp. 92-104.

[11] Voronov R. V., Malodushev S. V. Dinamicheskoe sozdanie kart urovnia WiFi-signalov dlia sistem lokal'nogo pozitsionirovaniia [Dynamic creation of maps of Wi-Fi signal level for location systems] *Sistemy i sredstva informatiki.* – 2014. no 9, pp. 37-41.

[12] Lantsev N. A. Pozitsionirovanie v besprovodnykh setiakh Wi-Fi [Location in wireless networks of Wi-Fi]. *VESTNIK VEGU.* – 2010. – no. 2. – pp. 92-94.

[13] Shin B. J. et al. Indoor WiFi positioning system for Android-based smartphone. *Information and Communication Technology Convergence (ICTC), 2010 International Conference on.* – IEEE, 2010. – pp. 319-320.

[14] Neznanov I. V., Namiot D. E. Kontrol' transportnykh marshrutov s pomoshch'iu mobil'nykh telefonov. [Control of transport routes with use of mobile phones] *International Journal of Open Information Technologies.* – 2015. – vol. 3. – pp. 8.

[15] Wang W. et al. Understanding and modeling of wifi signal based human activity recognition. *Proceedings of the 21st annual international conference on mobile computing and networking.* – ACM, 2015. – pp. 65-76.

[16] Lyon G. F. Nmap network scanning: The official Nmap project guide to network discovery and security scanning. – Insecure, 2009.Marcon M. et al. The local and global effects of traffic shaping in the internet. *Communication Systems and Networks (COMSNETS), 2011 Third International Conference on.* – IEEE, 2011. – pp. 1-10.

[17] Marcon M. et al. The local and global effects of traffic shaping in the internet. *Communication Systems and Networks (COMSNETS), 2011 Third International Conference on.* – IEEE, 2011. – pp. 1-10.

[18] A. Balachandran, G. M. Voelker, P. Bahl, P. V. Rangan. Characterizing User Behavior and Network Performance in a Public Wireless LAN. *ACM SIGMETRICS Performance Evaluation Review.* – ACM, 2002. – vol. 30. – no. 1. – pp. 195-205.

[19] Mirylenka K. et al. Characterizing home device usage from wireless traffic time series. *19th International Conference on Extending Database Technology (EDBT).* – 2016.

[20] Fukuda K., Nagami K. A measurement of mobile traffic offloading. *International Conference on Passive and Active Network Measurement.* – Springer Berlin Heidelberg, 2013. – pp. 73-82.