A METHOD TO INCREASE THE THROUGHPUT CAPACITY OF A MOBILE TELECOMMUNICATIONS NETWORKS

N. Bogomolova[1], O. Denisieva[1], A. Gobzemis[2]

[1] *Moscow Technical University of Communications and Computer Science,*

[2] *Riga Technical University*

[1] *Moscow, Russia*

[2] *Riga, Latvia*

gobzemis@edi.lv

The construction of a mobile network is a project that involves major costs, whereas the use of approximated models may lead to errors in calculation of the needed equipment, which in the end leads to addition costs of network construction.

Authors of the given paper offer to take account of and single out different categories of users during the design. Each of such groups will have its own profile and parameters created by the information and signal load and a specific range of additional services (AS). With the help of imitation modeling the authors have performed qualitative assessment of function of mobile networks. The work contains recommendations of introduction of differentiated servicing of different categories of users, which will eventually result in the general increase of the throughput capacity of the network.

1. INTRODUCTION

At the present moment we see the increase of the number of users that are ready to bear additional costs to receive higher quality communication services. We may single out three modern approaches to ensuring quality of service in communication networks [1]:

- users are provided with a guaranteed network resource which is sufficient to transfer the requested amount of information in a given period of time; guarantees in this case are set with the help of probability indicators;
- users are offered a multiple-level service where service set is differently for calls by different groups of users;
- certain network resources are reserved for users depending on the type of transferable information.

The first approach is traditional and indeed until now probability indicators have been used most often in the design. The given paper is focused on the second approach, i.e. authors recommend to take into account various categories of users upon the design and exploitation of the network by offering them different levels of service and at the same time also reducing the mutual impact of users in these different categories. The third approach in the opinion of the authors is a rather prospective one in case the respective network resources are available.

This work mostly focuses on the investigation of the behavior of different groups of users on the throughput capacity of the network and it offers to reduce their mutual influence by introduction of a multiple-level servicing of calls for different categories of users. The multiple-level contract on the quality of service - *SLA* (*Service Level Agreement*) concluded between the operator and the user must stipulate the dependence of quality of service from the sum of monthly payments. A certain amount of services will be provided for a certain monthly payment and the excess use will be additionally charged. In case of introduction of certain restrictions customers must be offered some additional services.

2. THE CRITERION OF ACCOUNTING OF THE MUTUAL IMPACT BETWEEN THE INFORMATION AND THE SIGNAL NETWORK

The structure of a mobile network of the GSM standard (phases 2G and 2G+) in line with ETSI 3G TS 29.002 version 3.3.0 [3] recommendation contains four levels: user level, level of the information network (IN), level of signaling (SN) and a level without data. The information of users is transferred via IN on the basis of the channel commutation method. The management information of setting connection processes, information between databases of the Home (HLR) and Visitors (VLR) registers and information of providing users with additional service is transmitted via SN, which is built on the basis of the signaling system No.7.

Special attention in the process of design and forecasting of loads is paid to the throughput capacity of IN. During the function of a network processes that are going on in the IN influence the processes that are going on in the SN and vice versa. Consequently the increase of the number of applications of connection in the IN will result in increased loads in SN and, as a consequence, to the increase of the time needed for establishing a connection. At the same time the increase of time needed to establish a connection leads to higher load on the information network because of a longer preliminarily booking time of information channels.

To ensure the quality of service to users of a mobile network there should a certain ratio established between time of enablement of channels in the IN and time needed to establish a connection with the SN and this ratio should not be disrupted in the course of exploitation of the network. Paper [2] introduces the notion of the coefficient of the ratio between the information load and the signaling load (ψ_{fn}) for the purposes of assessment the relation between the information and signaling load created by a flow of applications between any two commutation nodes (CN) in the network, for example, between CN_f and CN_n. Coefficient ψ_{fn} is equal to the relation between the average time of engagement of the information path and the average time of establishing a connection along the signaling path in the given direction. Knowing coefficients ψ_{ij} ($i, j = \overline{1, M}$, where M – is a number of commutation nodes CN in the network) for all pairs of corresponding network nodes we may get the integral value of the coefficient ψ for the whole network. This coefficient will then determine the balance between IN and SN loads.

As a contrast to the stationary (fixed) networks, in the mobile networks there is a more dynamic change of characteristics of the load created by users [2]. More specifically, the likelihood of load distortions in such networks caused by stochastic movement of users inside zones of network coverage and their connection to different MSC (Mobile Switching Center) is much higher (more often). One may also notice a dynamic increase (or decrease) of intensity of calls and average length of conversations caused by a wrong tariff policy of operators. This is accompanied by the change of coefficient ψg which mirrors the change in the functioning of the network. The coefficient ψ also depends on the size of applications flows for basic and additional services received in the network and also on the probability of engagement and failure to answer on the side of users, on the presence of satellite segments on the signaling path, on the average length of one signaling message, etc.

3. STRUCTURAL COMPOSITION OF USERS OF A MOBILE NETWORK

Similarly to users of fixed networks all users of the mobile network may also be split into categories, each of which has its own parameters of the created load and other specificities. Three categories of users of mobile networks were singled out during the observation of the load [2], which are conditionally called "orthodox", "business" and "unconscious" users (Table 1).

Table 1

Categories of users of mobile network and characteristics of the generated load

Parameters	Unconscious	Business	Orthodox
Average length of a conversation (sec)	4	30	60
Average length of wait of the user's answer (sec)	2	10	10
Average number of calls from users in maximal load time (number)	7	3	1
Coefficient ψ	**15**	**150**	**300**
Typical additional services	Short messages (SMS) (3 pieces from the user in maximal load time)	Retention of calls rerouting	–

"Unconscious users" occupy the first place in the table. This is the category, which currently is the target of competition between operators. The appearance of this category is caused by the fact that operators of cellular networks do not impose a tariff on the first few seconds (normally five) of connection, which is used by the mentioned group of users, who make a few short, free of change connections. A very short average call time is typical of this category (less than 5 seconds) and as well as a very high call intensity (7 calls from each user in the maximal load time). Users of this category also very actively use the SMS message service.

The next category – "business" users. It is characterized by an average length of call (30 seconds in the average) and a high number of calls from users in maximal load time. The given category comprises also active users of additional services (AS) i.e. they most frequently use retention of calls and rerouting.

And the last category of orthodox users may be characterized in terms of low call intensity (1 call) and a high average length of conversation (one minute). The given category is not a typical user of additional services.

4. SIMULATION

The created model mirrors transmission processes of SU (signaling units) in the SN with the due account taken of buffering mechanisms; correction of errors and the mechanism of overload control (with the help of the s-called "thresholds"). The network topology was set with matrices, determining the number of information channels between CN (commutation nodes) which at the same time also serve the function of signaling points (re-transmitters) n the SN. The presented approach presupposes the interconnected mode of work in the SN. Signaling points are characterized by then time of processing of SU, buffer size and the size of thresholds. The SC (Service Channel) is characterized by the time of information transmission. Although it was said before that the mechanisms ensuring faultless SU transmission were considered in the simulation, authors also assumed the absence of distortions or losses of information in channels. Inter-node load was set by the attraction matrix in the IN network. There was a separate attraction matrix set for SMS (short messages in mobile networks).

The authors set the following initial data: structure of IN and SN networks, parameters of the received application flows, time and other characteristics of the system as well as nullified or entered a value for default variables and massives needed for the program. After setting of the initial variables, the process of simulation was started.

The mobile network is the object of modeling in the given work. The mobile network consists of two transit nodes of commutation and a series of final commutation nodes (CN). When selecting the size of the sample, the needed relative discrepancy of results was assumed as 0.01 with credibility of 0.95. The basic output data were as follows for each pair of CN: total number of sent SMS, the number of received SMS, the number of attempts of connection, the number of successful connections, average time of establishing a connection and average time of engagement of the information network. For each segment of signaling the following was determined: the number of SU received for transmission, the number of refusals of telephone calls (both because of absence of information channels and because of overload of the signalization segment), the number of refused SMS, and the number of lost SU (because of the overfill in the buffer). The following parameters were calculated on the basis of the received simulation data: percentage of refusals of SN (with the breakdown by caused of refusals) and IN and the coefficient ψ for calls between every pair of CN. After that the integral coefficient is calculated ψ for the whole network and the share of refusals for all the calls received in the network.

5. RESULTS

With the increase of the number of "unconscious users" the load on information channels did not increase very much, IN productivity was adequate and therefore there were no refusals in the IN. The situation in the SN was quite opposite: with the increase of "unconscious users", refusals appeared more often caused by activity of thresholds triggered by the overfill of buffers. Consequently, coefficient ψ is going down reflecting the degree of disproportion of the load of each category of users in the IN and SN. Such situation is made even worse when SMS service is used. In this case even with the initial number of users the share of refusals in the SN is very high. Coefficient ψ falls down even more, mirroring the situation. Obviously, that "unconscious users" generate a lot of load in the SN creating "underload" in the IN, which is proved by the rapid decrease of the coefficient ψ.

Losses in the SN were permanently staying on a relatively low level whereas losses in IN after a certain moment of increase in the number of "business users" started to rapidly grow up. Notably, that with this, coefficient ψ did not change. After introduction of AS the situation stayed unchanged with the only difference that losses in the IN started to appear somewhat earlier alongside with the increase of the signaling load and prolongation of the average time of establishing a connection. This resulted in the increase of the time of occupancy of channels in the mobile network. It was noted that with insignificant increase of the number of "business users" their number is still insufficient to cause a visible impact on the average time of IN occupancy. Only afterwards, after the appearance of refusals in the IN, average time of network occupancy started to go down (influence of calls that were refused causing reduction of the average length of calls). With regard to SN, the productivity of the latter is enough to "sustain" the twofold increase of the number of "business" users. The given behavior of the coefficient ψ turns out to be fully explainable even without a too detailed analysis: "business users" load both IN and SN to the same extent and, therefore, even with the increased of the aggregate load, the balance expressed by coefficient ψ does not go up.

With the increase of "orthodox" users the situation becomes quite opposite to that with the "unconscious users", i.e. their increase does not cause an overload in the CN, the productivity of the network is sufficient. However, it must be noted that if we compare "business" and "orthodox" users, if their numbers increase gradually then "business users" tend to be quicker to load IN and are faster to receive refusals because of the lack of information channels. As a contrast to "business users", with the growth of the number of "orthodox users", there was a slight increase of coefficient ψ in the network. Such behavior proves time and again the change in characteristics of the incoming load to the network, as "orthodox users" were the ones that generated most of the load in the IN.

6. CONCLUSION

To bring down loads in IN and SN generated by different categories of users in line with the approach that suggests that different categories of users are offered different levels of service, it is necessary to develop special tariff plans which take account of the behavior of users. For instance, when concluding a contract a user selects certain tariff plan and after that the system of network management in the course of a certain time (for example 2 months) monitors the behavior of a customer from the point of view of the information and signal load he/she generates and the range of the used AS and calculates a personalized coefficient ψ. On the basis of this observation the subscriber is issued recommendations concerning the selection of the most optimal tariff plan, which may be voluntarily chosen by the user.

To ensure the above, it is necessary to develop recommendations of network management to ensure that, on the one hand, users get access to any services in the network and, on the other hand, the number of users is growing.

REFERENCES

1. *Denisieva O., Gobzemis A.* Structural optimization of local area and metropolitan area networks // Proceedings of the International Conference on Methods of Mathematical Analysis and Optimization of Telecommunication Networks. Mn.: BSU, 2003. P. 70–73.
2. *Bogomolova N. E.* Capacity of signaling path in Integrated Services Digital Network analysis // Elektrosvyaz. 2003. № 3. P. 41–43.
3. Mobile Application Part (MAP) Specification (3G TS 29.002, Version 3.3.0, Release 99).