Method for Estimating Parameters of 3G Data Traffic

A. Krendzel, Y. Koucheryavy, J. Harju
Institute of Communication Engineering
Tampere University of Technology (TUT)
Tampere, Finland
{krendzel, yk, harju}@cs.tut.fi

S. Lopatin
St.-Petersburg Research and Development Institute of Telecommunications (LONIIS)
St.-Petersburg, Russia
derj@loniis.org

***Abstract*—A large diversity of services based on packet switching in the Third Generation (3G) wireless system leads to dramatic changes in the characteristics and parameters of the data traffic. It is expected that the traffic in the 3G systems will undergo both quantitative and qualitative changes. In particular, there will be a considerable increase of a rate of transactions and an amount of information transferred during the transactions. Besides, it is necessary to take into account the notion of self-similarity that will take place in the 3G systems due to the high variability of burstiness of the multiservice traffic. For these reasons, the prediction problem of data traffic characteristics becomes more complex when planning the 3G wireless systems. In this paper the estimation method for the main parameters of the data traffic in the 3G mobile networks taking into account the above-mentioned 3G features is considered. It is based on a probabilistic model of events initiated by calls of 3G services. It is proposed to make both a decomposition of 3G services into some subsets in accordance with an average amount of transferred data during transactions and a distribution of 3G prospective users into some subgroups in accordance with the Pareto law.**

Keywords-the Third Generation (3G) wireless system, 3G data traffic estimating, the Pareto law; the probability model of 3G calls

I. INTRODUCTION

One of the main features of the 3G wireless systems is to support for a high-speed transfer of a large amount of multimedia information between users. The examples of 3G services are voice telephony, real-time interactive games, videotelephony, instant messaging, emergency calls, multimedia conferencing [1]. As a result, the traffic in the 3G systems will expand considerable [2]. It may lead to increasing the rate of transactions and the traffic intensity on a level of the Packet Switched Core Network (PS CN) domain. Besides, traffic patterns generated by 3G services are quite different from traditional Poisson models used for circuit switched voice traffic and may have self-similar nature due to high variability of burstiness [3,4,5]. According to [6] self-similar nature of the traffic may be taken into account assigning values 0.6….0.7 to the Hurst parameter.

Taking into consideration these facts the estimation problem of the main parameters of the data traffic in the 3G systems becomes more complicated. In order to overcome the difficulties, the estimation method for the traffic generated by 3G services on a level of the PS CN domain in the 3G systems is considered in this paper. The method is based on the probabilistic model of events initiated by 3G calls that is formed in accordance with both a decomposition of a 3G services set into subsets and a distribution of 3G potential users into subgroups.

The structure of the paper is organised as follows. We start with the decomposition of a 3G services set into subsets. As criterions for the decomposition an amount of transactions in a busy hour and an amount of data transferred during a transaction are used. Then, the distribution of 3G users into subgroups is fulfilled taking into consideration a non-uniform demand for 3G services in accordance with the Pareto law [7,8]. Thus, the probabilistic model of 3G calls comprises two exhaustive classes of statistically independent events concerning 3G calls. It allows segregating the common traffic into some components. After that, we calculate the rate of transactions and determine the traffic intensity on nodes of the PS domain. Finally, we present the case study of the calculation of the main parameters of the data traffic using obtained expressions.

II. THE DECOMPOSITION OF A 3G SERVICES SET INTO SOME SUBSETS

The main feature of the data traffic in the 3G systems is a large diversity depending on the profile of services provided to 3G users [9]. It is reasonable to make the decomposition of 3G services set into subsets taking into account that services belonging to one subset will generate quite equal traffic intensity. In this paper we present the approach for the decomposition of a 3G services set into three subsets (I = 3). Emphasis, that there are not special restrictions for a choice of a number of subsets.

The first subset of services (i = 1) deals with a transfer of a small amount of information (E-mails, Web-pages and so on) characterizing by integrity of information blocks. The subset including such services is named "Messaging". Below the subset is denoted by the abbreviation "MS". A demand for the "MS" services may be increased, for example, during national holidays; however, an average amount of transferred information per a transaction for this type of services is quite low.

The second subset includes a transfer of texts, sounds, graphical files, for instance, a transfer of melodies, color cards and so on. The subset is named "Melody" and denoted by the abbreviation "ML" below. A demand for the "ML" services and an average amount of transferred information per a transaction may be characterized as the mean level.

The third subset comprises multimedia services, which deal with a huge amount of information. The subset is called "Multimedia" and has the abbreviation "MM" below. An amount of transferred information for the "MM" services may achieve several tens of MB. The initial information concerning a specific (per each subset of services) distribution of the total amount of transactions in a busy hour (γ_i, i = 1,2,3) may be used as the numerical criterion for the decomposition of a 3G services set into subsets. These initial parameters form the exhaustive set and may be denoted as

$$\gamma_i, \quad i = 1,2,3, \quad \gamma_1 + \gamma_2 + \gamma_3 = 1, \tag{1}$$

Thus, the result of the submitted decomposition is the division of 3G services into three subsets. Such approach has the following advantage. 3G services within each subset will generate the current traffic intensity that has a variance of values of parameters substantially smaller than one relating to the traffic intensity of the total 3G services set.

III. THE DISTRIBUTION OF 3G USERS INTO SOME SUBGROUPS

A demand for 3G services depends on both a solvency of users and tariffs on 3G services. It is clearly that tariffs on 3G services from the different subsets of 3G services will be unequal. As a result, a non-uniformity of a demand for 3G services from different subsets will take place between 3G users. In order to take into account the fact it is necessary to distribute all 3G users into subgroups in accordance with their demand for 3G services from the different subsets defined earlier.

Parameters of the non-uniformity of the distribution of 3G services for each subset may be considered as the input data. Usually the non-uniform distribution of incomes between population is characterized by the Gini coefficients [10,11]. Values of the Gini coefficients may be defined on a basis of statistical information and marketing research regarding a demand for 3G services.

Let the non-uniformity of a demand for each subset of 3G services be defined as

$$K_{G,i}, \qquad i = 1,2,3. \tag{2}$$

Here $K_{G,i}$ is the the Gini coefficient for i - th subset of 3G services.

Let's assume that the non-uniformity of the distribution of 3G services for each subset corresponds to the Pareto law. Then parameters of the Pareto distribution may be calculated as follows [12]

$$\alpha_i = \frac{0.5(K_{G,i}+1)}{K_{G,i}}, \qquad i = 1,2,3. \tag{3}$$

It is obviously that the least non-uniformity of the distribution will take place for the "MS" subset and the largest one will take place for the "MM" subset, i.e.

$$\alpha_1 > \alpha_2 > \alpha_3 > 1. \tag{4}$$

It is quite easy to distribute users in accordance with a demand for 3G services when given (3,4). As a rule for the distribution of users into three subgroups (J = 3) the following one described below may be applied. Emphasis, that there are not special restrictions for a choice $J > 3$.

The wealthiest users ("rich") producing 90% of a demand for 3G services from the "MM" subset are included into the third (j = 3) subgroup. Note that Lorenz curves corresponding to the Pareto distribution with the parameter α may be written as follows [12,13,14]

$$Q(\alpha, x) = 1 - (1 - F(x))^{\frac{\alpha-1}{\alpha}}. \tag{5}$$

Using the Lorenz curves it is possible to determine the relative number of users in the "rich" subgroup (F_3) as

$$F_3 = 0.9^{\frac{\alpha_3}{(\alpha_3-1)}}. \tag{6}$$

Let users from the third subgroup and the second subgroup create 90% of a demand for 3G services from the "ML" subset. Then the relative number of users in the second (j = 2) subgroup named "middle" is determined as

$$F_2 = 0.9^{\frac{\alpha_2}{(\alpha_2-1)}} - F_3. \tag{7}$$

At last, the relative number of users in the first subgroup (j=1) named "poor" may be found as

$$F_1 = 1 - F_2 - F_3. \tag{8}$$

The expressions (6,7,8) give the rule of a definition of the third subgroups of 3G users in accordance with their demand for 3G services from the subsets of 3G services defined in the section 3.

IV. THE CALCULATION OF THE RATE OF TRANSACTIONS

A. The problem formulation

The decomposition of 3G services and the distribution of 3G users give a possibility to form the probabilistic model of the initiation of transactions based on an intersection of events from two statistically independent exhaustive classes.

The events included in the first class are denoted by index i= 1,2,3. They correspond with demands for services from the "MS", "ML", "MM" subsets respectively. The events included in the second class are denoted by index j = 1,2,3. They correspond with demands initiated by users from the "poor", "middle" and "rich" subgroups respectively. The first class and the second class of events are exhaustive by a definition since the expressions (1,8) are true. It is supposed that these classes of events are independent, in other words, the probability of an intersection of the events is equal to the product of probabilities each of the events.

It should be explained in detail. Both the above-mentioned decomposition of 3G services into three subsets and the above-mentioned division of 3G users into three subgroups allow defining nine events. The events form the exhaustive set and are denoted below by the dual index i,j: $i = 1,2,3$; $j = 1,2,3$. The first event ($i = 1$, $j = 1$) is that an user from the first subgroup (j=1) initiates a demand for a service from the first subsets (i=1). At the second event ($i = 2$, $j = 1$) an user from the first subgroup initiates a demand for a service from the second subset. At the third event ($i = 3$, $j = 1$) an user from the first subgroup makes a demand for a service from the third subset. The fourth event ($i = 1$, $j = 2$) is that an user from the second subgroup makes a request for a service from the first subset and so on.

Thus the probabilistic model of the events allows segregating nine segments from the common flow of transactions initiated in a busy hour by procedures of such services supporting. A variance of values of parameters each of random flows will be less than one relating to the common flow. This is because of the above-mentioned rules of the decomposition of 3G services and the distribution of 3G users.

Then the problem is to determine average values of the rate of transactions between nodes of the PS CN domain in the 3G system. It corresponds with an intersection of two events from two formed classes of events.

We consider the approach for the calculation of the specific rate of transactions (λ_{ij}) per an user in a busy hour for nine (i,j=1,2,3) intersections of events from two above-mentioned classes. It is based on solving for a system of three equations that should be formed for each of subsets of 3G services.

Note that each system of the equations may be assigned on a basis of values of the parameter β_{ij} where β_{ij} is a share of transactions in a busy hour relating to users of the j-th subgroup when services from the i-th subset are initiated, i,j=1,2,3. Values β_{ij} may be defined by two different ways that give a possibility to form the required equations below.

B. *The definition of β_{ij} on a basis of the non-uniformity of a demand for 3G services*

Values β_{ij} should be calculated for each of nine events at given α_i, i=1,2,3; F_j, j=1,2,3. Taking into account the expressions (3,5,6,7,8) and that $\Sigma_j \beta_{ij}=1$, i,j =1,2,3 the following systems of the equations may be defined as

$$\begin{gathered} \beta_{i1} = 1-(1-F_1)^{\frac{(\alpha_i-1)}{\alpha_i}}, \\ \beta_{i2} = 1-(1-F_1-F_2)^{\frac{(\alpha_i-1)}{\alpha_i}} - \beta_{i1}, \\ \beta_{i3} = 1-\beta_{i1}-\beta_{i2}. \end{gathered} \tag{9}$$

Three (i=1,2,3) systems of the equations (9) allow calculating the distribution of transactions between three subgroups of users taking into consideration the non-uniformity of a demand in each subset of 3G services.

C. *The definition of β_{ij} over the rate of transactions λ_{ij}*

Using the parameter λ_{ij} (i,j = 1,2,3) defined earlier for the specific rate of transactions in a busy hour per an user it is quite easy to compose the expression for the rate of transactions in a busy hour that generated by all users of the j-th subgroup when they initiate services from the i-th subset as follows $F_jN\lambda_{ij}$ (i,j=1,2,3), where N is a number of 3G users.

Note that the number of 3G users may be estimated approximately as a number of wealthy 2G users, i.e., $N=dN_n$, where N_n is a number of 2G users, d is a share of wealthy 2G users.

Averaging over "j" the expression $F_jN\lambda_{ij}$ the following one for the rate of transactions generated by calls of 3G services from the i-th subset may be obtained: $N\Sigma_j\lambda_{ij}F_j$, i,j = 1,2,3. The required expression for the calculation β_{ij} over λ_{ij} is derived by a division of the first expression into the second one that gives

$$\beta_{ij} = \frac{F_j\lambda_{ij}}{\sum_j \lambda_{ij}F_j}, \qquad i,j = 1,2,3. \tag{10}$$

The expressions (9,10) give the system of nine equations with nine unknown λ_{ij}, i,j=1,2,3.

D. *Solving for the system of equations*

Let values β_{ij}, i,j=1,2,3 be known in accordance with (9). The system of nine equations (10) may be transformed into the system that consists of tree matrix-vector equations.

$$\begin{pmatrix} (\beta_{i1}-1)F_1 & \beta_{i1}F_2 & \beta_{i1}F_3 \\ \beta_{i2}F_1 & (\beta_{i2}-1)F_2 & \beta_{i2}F_3 \\ \beta_{i3}F_1 & \beta_{i3}F_2 & (\beta_{i3}-1)F_3 \end{pmatrix} \begin{pmatrix} \lambda_{i1} \\ \lambda_{i2} \\ \lambda_{i3} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \tag{11}$$

or, more compactly

$$A_i\lambda_i = \mathbf{0}, \qquad i = 1,2,3. \tag{12}$$

The elements (3x3) of matrixes $\boldsymbol{A_i}$ and vectors $\boldsymbol{\lambda_i}$,$\mathbf{0}$ are given in (11) respectively.

The equation (12) is uniform. It has the nonzero solution only if matrixes $\boldsymbol{A_i}$ are particular, i.e. their determinants are equal to zero. It is easy to show that this condition is true because rows each of the matrixes are linear-dependent since a sum of all rows each of matrixes $\boldsymbol{A_i}$ gives the zero row taking into account (8, 9). The linear combination of any two rows of a determinant each of matrixes $\boldsymbol{A_i}$ does not give the zero row. Thus a rank of matrixes $\boldsymbol{A_i}$ is equal to 2 and a defect of matrixes $\boldsymbol{A_i}$ is equal to 1.

Taking into account such features of matrixes $\boldsymbol{A_i}$ the system (12) has three one-parameter families of solutions

$$\lambda_{ij} = k_iC_{inj}, \qquad i,n,j = 1,2,3; \tag{13}$$

where C_{inj} are algebraic cofactors any row of matrixes $\boldsymbol{A_i}$; k_i are parameters of three families of solutions.

In order to concretize the solutions (13) it is necessary to add to the input data the parameter value λ_{11} that may be quite easy estimated on a basis of statistical information concerning a demand for 2G services.

Then using (13) we obtain $k_1 = \lambda_{11}/C_{1n1}$ when $i = 1$. After this we select, for example, the row number three. For this row algebraic cofactors of the matrix $\boldsymbol{A_1}$ are determined, i.e. $C_{131}=F_2F_3\beta_{11}$. Then we have a possibility to get the rest two solutions for the first family as follows

$$\lambda_{12} = \frac{\lambda_{11}\beta_{12}F_1}{\beta_{11}F_2}; \qquad \lambda_{13} = \frac{\lambda_{11}\beta_{13}F_1}{\beta_{11}F_3}. \tag{14}$$

In order to find the family of solutions when $i = 2$ it is necessary to assign a value of the parameter k_2. For that it is worthwhile to use the input data defined in (1), i.e. values of the parameter γ_i, i=1,2,3, which connect values $N\sum_j \lambda_{ij}F_j$, i,j=1,2,3 with each other as

$$\begin{aligned}\gamma_2\sum_j \lambda_{1j}F_j &= \gamma_1\sum_j \lambda_{2j}F_j,\\ \gamma_3\sum_j \lambda_{1j}F_j &= \gamma_1\sum_j \lambda_{3j}F_j.\end{aligned} \tag{15}$$

Taking into account the obtained sum $\Sigma_j \lambda_{1j}F_j$ and expressions (10,15) we have the rest two families of solutions when i = 2,3 as

$$\lambda_{ij} = \frac{\gamma_i\beta_{ij}\sum_j \lambda_{1j}F_j}{\gamma_1F_j}; i = 2,3;\ j = 1,2,3. \tag{16}$$

It is obviously that the solutions assigned by expressions (13,16) are equivalent, however, the expression (16) is simpler taking into account the selected input data, i.e., it is not necessary to assign values of the parameters k_2 and k_3.

Thus we have obtained the estimations (14,16) of the rate of transactions in a busy hour for 8 segments of the data traffic generated by procedures of 3G services supporting. The value of the first segment (i,j = 1) is included in the input data.

V. The Calculation of the Traffic Intensity Generated by 3G Services on Nodes of the PS Domain

In order to calculate the traffic intensity in Erlang, in addition to the rate of transactions it is necessary to determine values of the parameter T_i (i=1,2,3) corresponding to procedures of 3G services supporting from the above-mentioned subsets, where parameter T_i is the average duration of a transaction. It is worthwhile to add to the input data both the value of the average amount of information transferred during a transaction and the value of the average rate of transferred data in a radio-channel on the interface Mobile Station / Base Station (Node B). These parameters are denoted respectively as

$$\begin{aligned}&w_i, kbit/trans; i = 1,2,3;\\ &c, kbit/\text{sec}.\end{aligned} \tag{17}$$

Using these symbols the average duration of a transaction corresponding to procedures of 3G services supporting from the i-th subset may be found as

$$T_i = \frac{w_i}{c}, \text{sec}/trans. \tag{18}$$

The required values of the specific (per a user) traffic intensity generated by calls from the i-th subset of 3G services that are initiated by users from the j-th subgroup are obtained as follows

$$S_{i,j} = \frac{\lambda_{ij}T_i}{3600}, Erl. \tag{19}$$

The traffic intensity created by calls from the i-th subset of 3G services that are initiated by all 3G users may be found as

$$S_i = \frac{T_i\sum_j F_j\lambda_{i,j}}{3600}, Erl. \tag{20}$$

and, at last, the final expression for the estimation of the total traffic intensity is

$$S = \frac{\sum_i T_i\sum_j F_j\lambda_{i,j}}{3600}, Erl\ . \tag{21}$$

The presented expressions have been obtained when both the number of subsets of 3G services (I) and the number of subgroups of 3G users (J) are equal to three. However, as it has been emphasized above all derived expressions may be easy generalized in cases when values I and J are more then three.

VI. The Case Study

The example of the estimation of the parameters of the data traffic generated by 3G users is considered in this section.

The initial values are presented in Table 1.

TABLE I.

Para-meters	**"MS" subset**	**"ML" subset**	**"MM" subset**	**Meaning**
α_i	3.5	1.05	1.02	Pareto coefficient
γ_i	0.9	0.085	0.015	Share of transactions
w_i, kbit/trans	1	100	10000	Av. amount of data / trans
N	100000			Number of 2G users
d	0.45			Share of wealthy 2G users
c, kbit/sec	50			Av. data rate
λ_{11}, trans/user	0.5			Rate of transaction
k	0.1			Concentration factor

The calculations have been made using the approach and the expressions presented in previous sections. The most essential results of the data traffic estimating are the follows

- the total traffic intensity is 30 Erl in a busy hour;

- the rate of transactions per an user in a busy hour is 0.67 trans/user for the “MS” subset, 0.06 trans/user for the “ML” subset, 0.01 trans/user for the “MM” subset;
- the amount of data arriving at the PS CN domain from users of the “MS” subset of 3G services during a month is 1130 MB, from users of the “ML” subset is 10690 MB, from users of the “MM” subset is 188570 MB;
- the average duration of a transaction is 0.02 sec for the “MS” subset of 3G services, 2 sec for the “ML” subset and 200 sec for the “MM” subset.

It is seen from the results that services from the “MS” subset generate a small part of the traffic intensity on network elements, however, initiate the most number of transactions. At the same time, services from the “MM” (Multimedia) subsets give the main part of the traffic intensity, but have the least number of transactions. Hence it is worthwhile to impose tariffs on 3G services depending on as a number of initiated transactions as an amount of transferred data.

VII. CONCLUSION

In this paper the estimation method for parameters of the data traffic generated by 3G users has been presented. The method is based on the probabilistic model of 3G calls that has been composed in conformity with the rules of the decomposition of 3G services into subsets and the distribution of 3G users into subgroups. It allows segregating all traffic into nine segments. Thus the submitted method gives a possibility to combine separated statistical data concerning the deployment of 3G services (based on packet switching) into the unified probabilistic model and estimate parameters of the data traffic that has self-similar nature.

With the help of the method the following parameters of the data traffic may be estimated:

- the specific (per an user) rate of transactions in a busy hour in each of the defined segments of the data traffic;
- the average duration of 3G calls in each subset of 3G services;
- the specific (per an user) traffic intensity created by 3G calls in a busy hour in each subset of 3G services and in each subgroup of 3G users;
- the total traffic intensity on nodes of the PS domain in the 3G systems.

The use of the method allows making easier both the preparation of business-plans for development of wide range of 3G services and the preparation of system projects for deployment of the PS CN domain.

REFERENCES

[1] M. C. Bale, “Voice and Internet multimedia in UMTS networks,” BT Technology Journal, vol. 19, №.1, 2001.

[2] ”Long term potential remains high for 3G mobile data services,” Report # 18 from the UMTS Forum, February 2002.

[3] M. Jiang, M. Nicolic, S. Hardy, L. Trajkovic, “Impact of self-similarity on wireless data network performance,” *ICC 2001*, *USA*, June 2001.

[4] W. Willinger, M.S. Taqqu, W.E. Leland, D.V. Wilson, “Self-similarity in high-speed packet traffic: analysis and modeling of Ethernet traffic measurements”, Statistical Science, vol. 10, no. 1, pp. 67-85, 1995.

[5] W. Willinger, M.S. Taqqu, R. Sherman, D.V. Wilson, “Self-similarity through high-variability: statistical analysis of Ethernet LAN traffic at the source level”, IEEE/ACM Transactions on Networking, vol. 5, no. 1, pp. 71-86, 1997.

[6] Koucheryavy Y., Krendzel A., Lopatin S., Harju J. "Performance Estimation of UMTS Release 5 IM-Subsystem Elements", 4-th International Workshop on Mobile and Wireless Communications Network, Stockholm, Sweden, September 9-11, 2002, ISBN 0-7803-7606-4.

[7] J. A. Schumpeter, “Vilfredo Pareto: in ten great economist from Marks to Keynes”, OUP, 1951.

[8] M. Allais, “Pareto, Vilfredo: contributions to economics”, International Encyclopedia of the Social Sciences, Macmillan and Free Press, vol.11, 1968

[9] H. Kaaranen, A. Ahtiainen, L. Laitinen, S. Naghian, V. Niemi, “UMTS networks. Architecture, mobility and services,” John Wiley & Sons, 2001.

[10] M. Kendall M, A. Stuart, “The advanced theory of statistics. Distribution Theory”, London, C. Griffin & Co., 1962.

[11] P. M. Dixon, T. Mitchell-Olds, R. Woodley, “Bootstapping the Gini coefficient of inequality”, Ecology 68, pp. 1548-1551, 1987.

[12] L.E. Varakin, “The Pareto law and the rule 20/80: the distribution of incomes and telecommunication services”, MAC proceedings, pp.3-10, no.1, 1997.

[13] P. Derusso, Roy R., Close C., A. Desrochers, “State variables for engineers”, John Wiley & Sons, 1998.

[14] C.V. Brown, P.M. Jackson, “Public sector economics”, Oxford, Blackwell, 1990.