Method for Estimating a Number of Potential 3G Users

A. Krendzel[1], Y. Koucheryavy[1], V. Derzhavina[2], J. Harju[1]

***Abstract*--It is expected that some network planning problems will take place when deployment of the Third Generation (3G) wireless systems due to a considerable increase of multiservice traffic amount and its large diversity. The estimation problem of a number of prospective users of 3G services is the initial one of network planning. The successful solution of the problem gives a basis to solve other topical 3G network planning problems such as 3G traffic estimating, performance evaluation of 3G network nodes and others. In this paper the method for estimating a number of potential 3G users in a region is considered. It is based on application of the Pareto law, the Lorenz curves and the Gini coefficient.**

***Keywords* -- the Third Generation systems, the Pareto Law, the Lorenz curve, the Jipp curve, the Gini coefficient**

I. INTRODUCTION

One of the main features of the data traffic in the Third Generation (3G) wireless systems is a large diversity depending on the profile of services provided to 3G users. It is supposed that there will be dramatic quantitative and qualitative changes in the parameters and nature of the traffic.

In particular, there will be a considerable increase in the rate of transactions and in the total traffic intensity in 3G systems [1]. Besides, it is necessary to take into account the notion of self-similarity that may occur in 3G systems due to the high variability of burstiness of the multiservice traffic [2,3].

As results of it, some network planning problems arise when deployment 3G wireless systems, namely [4,5]

- the estimation problem of 3G traffic parameters,
- the problem of the performance estimation of Packet Switched Code Network (PS CN) domain elements taking into consideration the self-similar nature of the multiservice traffic,
- the estimation problem of the necessary number of trunk lines to support an interaction between PS CN domain elements and others.

The solution of the above-mentioned problems enables planning 3G networks in such a way that both technical and economical advantages can be achieved when deployment and exploiting the networks.

However, in order to solve the above-mentioned problems, at first, it is necessary to determine an amount of prospective users of 3G services. It may be considered as the preliminary problem of 3G planning. The successful solution of the problem will enables forming initial data for other 3G network planning problems. In this paper the estimation method for a potential number of users of 3G services is developed. The method is based on use of the Pareto law, the Jipp curve, the Lorenz curves, and the Gini coefficient.

II. THE PARETO LAW, THE JIPP CURVE AND A NUMBER OF WEALTHY 2G USERS

More than a century ago the Italian sociologist and economist Vilfredo Pareto studied the distribution of personal incomes for the purpose of characterizing a whole country's economic status [6,7,8]. His law is a mathematical formulation on a basis of empirical conclusions in which he attempts to prove that the distribution of incomes and wealth in society is not random, but exhibits a consistent pattern. This law, which Pareto derived from British data on income, shows a relationship between each income level and the number of people who receive more than that income. According to the Pareto law the more individual income level in a subgroup the less individuals are in the subgroup.

Let's reflect values of the normalized income G/G_{min} into value area of random variable $1 \leq x < \infty$, where G is one of income values, G_{min} is the minimum income value. Then, in accordance with the Pareto law, the number of individuals who have income more than x is

$$R(x) = x^{-\alpha}, \quad x = G/G_{\min}, \quad 1 < \alpha < \infty, \qquad (1)$$

where α is the distribution parameter called the Pareto parameter.

In term of integral distribution we have the follows

$$R(x) = P(\xi > x) = 1 - F(x) = \begin{cases} x^{-\alpha}, & x \geq 1 \\ 1, & x < 1 \end{cases}. \qquad (2)$$

Here, $F(x) = P(\xi < x)$ is the integral function of probability distribution (the monotone increasing function), $P(\xi < x)$, $P(\xi > x)$ are the probabilities of

[1]Authors are with Institute of Communication Engineering, Tampere University of Technology, Tampere, Finland, http://www.cs.tut.fi/tlt/, e-mail: {krendzel, yk, harju}@cs.tut.fi. [2]Author is with St.-Petersburg Research and Development Institute of Telecommunications (LONIIS), St.-Petersburg, Russia, http:/www.loniis.ru, e-mail: derj@inser.loniis.spb.su.

events when a value of a random variable ξ less/more than x.

By definition the probability density of the random variable ξ with the integral distribution $F(x)$ is follows

$$w(x) = \frac{dF(x)}{dx} = \frac{-dR(x)}{dx} = \frac{\alpha}{x^{\alpha+1}}, \quad x \geq 1 \qquad (3)$$

It should be emphasized that $\int_1^{\infty} w(x)dx = 1$ and $w(x) = 0$ when $x < 1$.

The random variable ξ given by the probability density $w(x)$ has n-th moment of distribution in the area where $\alpha > n$ only. The n-th moment may be defined as

$$M_n = \int_1^{\infty} x^n w(x) = \int_1^{\infty} \frac{\alpha x^n}{x^{\alpha+1}} dx = \frac{\alpha}{\alpha - n}, when\, \alpha > n \qquad (4)$$

When $\alpha < n$ we have integral divergence in (4). It is seen from (4) that the average value of random variable distributed by the Pareto law exists and may be obtained as

$$M_1(\alpha) = M(\alpha) = \frac{\alpha}{\alpha - 1} \qquad (5)$$

Taking into account (5) and the normalization rule of income values it is possible to get the expression $G_{min}(\alpha) = G_0 / M(\alpha)$. By substituting the expression into (1) we have

$$R(G/G_{\min}) = \left(\frac{G}{G_0} M(\alpha)\right)^{-\alpha}, \qquad (6)$$

where G_0 is the average value of personal annual income.

The relationship between infocommunication density and an individual level of Gross Domestic Product (GDP) may be described by the well-known Jipp curve [9]. Mathematically it may be approximated as [10]

$$\ln \Delta = \ln A + a \ln G_0, \qquad (7)$$

where Δ is the number of income earners who receive incomes higher than G_0, A is the normalizing dimension factor, a is the power index when G_0 in the non-logarithmic form of notation.

Taking into account that we deal with the estimation problem of prospective 3G users it is reasonable to interpret the parameter Δ as the penetration level of the Second Generation (2G) services. It may be defined as $\Delta = N_{mt}/100$, where N_{mt} is the average number of 2G mobile terminals per 100 individuals.

Investigations are shown that when analyzing the penetration level of 2G terminals it is worthwhile to substitute the value $a = 1$ in the expression (7). Transforming (7) we have the follows

$$\frac{T}{\Delta} = \frac{G}{G_0}. \qquad (8)$$

It gives a possibility to form the expression for determination of the subgroup of wealthy users as follows

$$R(T, \Delta, \alpha) = \left(\frac{T}{\Delta} M(\alpha)\right)^{-\alpha}. \qquad (9)$$

Here the ratio T/Δ determines the excess of the penetration level of 2G services in the subgroup of wealthy users above the average value of the 2G penetration level relating to all individuals. As a rule, values of the parameter Δ are known and may be found in statistical literature. Therefore, for estimation of a number of users in the wealthy subgroup (when given T) it is necessary to find the Pareto parameter α.

III. The Pareto Parameter and the Lorenz Curves

Solving the above-mentioned problem it is worthwhile to take into account that the parameter value α depends generally on inequality of the income distribution. In our case it depends on inequality of distribution of a number of phones or computers between individuals. It is often the Lorenz curves are applied to determine inequality of the income distribution in a subgroup of individuals [11,12,13]. If all individuals have the same income, the Lorenz curve is a straight diagonal line, called the line of equality. If there is any inequality in income then the Lorenz curve falls below the line of equality. The Lorenz curves show relationships between the normalized (to the parameter value M) average income in a subgroup of population Q and the normalized number of individuals in the subgroup $F = 1 - R$.

The analytical function $Q(F)$ is derived from Pareto distribution [14,15]. So, by definition, we have the following system of two equations for Q and F

$$1 - Q(x) = 1 - \frac{1}{M}\int_1^x z w(z)dz = x^{1-\alpha}, \qquad (10)$$

$$F(x) = \int_1^x w(z)dz = 1 - x^{-\alpha}. \qquad (11)$$

From (10, 11) we have the analytical assignment of a set of the Lorenz curves as follows

$$Q(\alpha, x) = 1 - (1 - F(x))^{\frac{\alpha - 1}{\alpha}}. \qquad (12)$$

The set of the Lorenz curves shows that with an increasing the parameter value α income in a subgroup is becoming more evenly distributed. It should be emphasized that the relationships $Q(\alpha, x)$ are the generalization of the well-known inequality of the income distribution between population that can be given by as

$$\frac{\rho}{1-\rho}, \qquad \rho < 0.5. \tag{13}$$

The ratio means that in many aspects of human activities the less part of population (ρ%) earns the most part of cumulative income ($1-\rho$)%. The inequality 20/80 corresponds to the Pareto law given by the Lorenz curve $Q(\alpha,x)$ with α=1,16. It should be taken into consideration that the value α=1,5 is assigned as the mean statistical estimation. It corresponds approximately to the parameter of inequality ρ=0.3 [15].

Usually the Gini coefficient K_G is considered in statistical literature as a measure of inequality in a population [16,17]. The Gini coefficient is the ratio between the area enclosed by the line of equality $Q = F$ and the Lorenz curve $Q(F)$, and the total triangular area under the line of equality.

As a rule, in each stage of 2G development it is known only a number of 2G users on a basis of statistical information. In other words, it is known a number of individuals in a subgroup that has a completely even distribution of 2G services. It is obviously that the Lorenz curve in this case transforms into the broken line (on the assumption that one user has only one mobile terminal). It may be expressed as follows

$$Q(\Delta,F)=\begin{cases}0, & 0<F<1-\Delta\\ \dfrac{F-1+\Delta}{\Delta}, & 1-\Delta\le F\le 1\end{cases}, \tag{14}$$

The graph of the distribution (14) when the penetration level Δ = 0.3 and the Lorenz curves when α = 1.05; 1.15; 1.5; 3.0 are shown in Fig. 1.

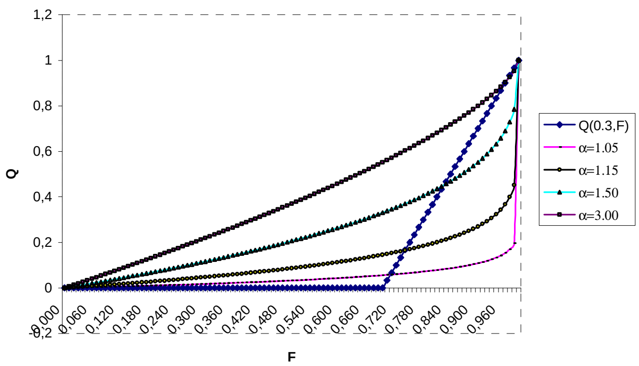

Fig. 1. The set of Lorenz curves

Note that a mobile terminal of a subscriber may be used by some other people, for example, subscriber's family members, colleagues, friends and so on. Taking into account this fact, the distribution of 2G services between individuals is "more equal" then the distribution of 2G terminals. Since in most cases this distribution corresponds to the Pareto law, in order to solve the problem of determination of a number of wealthy users it is necessary to find a transition from the distribution curve of subscribers $Q(\Delta, F)$ to the distribution curve of users $Q(\alpha,F)$. In other words, it is necessary to approximate the distribution (14) depending on the parameter Δ by the function (12) corresponding to the Pareto distribution with the parameter α.

IV. Solution of the Approximation Problem

The approximation problem may be formulated as follows. When the function $Q(\Delta,F)$ is given it is necessary to find the function $Q(\alpha,F)$ where the parameter α satisfies the following requirement

$$\alpha_{opt} = \arg\min d(Q(\Delta,F);Q(\alpha,F)). \tag{15}$$

Here $d(Q(\Delta,F);Q(\alpha,F))$ is the distance between the approximated function and the approximating one.

According to (15) it is necessary to assign the functional $d()$ to solve the formulated problem.

In this paper we present the approach for determination of optimal value of the parameter α based on the approximation with help of the Gini coefficient.

As a measure of the distance between the functions $Q(\Delta,F)$ and $Q(\alpha,F)$ the value of difference module between corresponding the Gini coefficients, i.e.

$$d(Q(\Delta,F);Q(\alpha,F)) = \mathrm{mod}(K_G(\Delta)-K_G(\alpha)), \tag{16}$$

where $K_G(\Delta)$, $K_G(\alpha)$ are the Gini coefficients relating to the functions $Q(\Delta,F)$ and $Q(\alpha,F)$ respectively.

It is shown in [15] that for the distribution (12) the Gini coefficient $K_G(\alpha)$ may be expressed as follows

$$K_G(\alpha)=\frac{1}{2\alpha-1}. \tag{17}$$

For the distribution (14) the Gini coefficient $K_G(\Delta)$ may be found as

$$K_G(\Delta)=1-\Delta. \tag{18}$$

It gives a possibility to find the solution (15) taking into account (16) as follows

$$\alpha_{opt}=\frac{0.5(2-\Delta)}{1-\Delta}. \tag{19}$$

Substituting (19) into the right part of the expression (9) we get the follows

$$K=\left(\frac{\Delta^2}{T(2-\Delta)}\right)^{\frac{2-\Delta}{2(1-\Delta)}}. \tag{20}$$

Emphasis that the value $R(T,\Delta)$ have not to exceed 1. Therefore, $R(T,\Delta)$ is equal K only at such values of the parameter Δ when K is less or equal to 1. At the value area of the parameter Δ where $K > 1$, the value $R(T, \Delta)$ is equal to 1. Analytically the condition may be formulated using the constrain function.

$$R(T,\Delta) = 1 + (K-1)rect(K,1), \qquad (21)$$

where

$$rect(x,y) = \begin{cases} 1, & |x| \le y \\ 0, & |x| > y \end{cases}$$

is the constraint function.

Note that the value of the parameter T should be on the interval $0.5 < T < 0.7$ because only at such values of the penetration level of 2G services in the subgroup of wealthy users, the members of the subgroup are the potential users of 3G services. In the Fig. 2 the relationship $R(T, \Delta)$ when $T = 0.6$ is shown in accordance with (21).

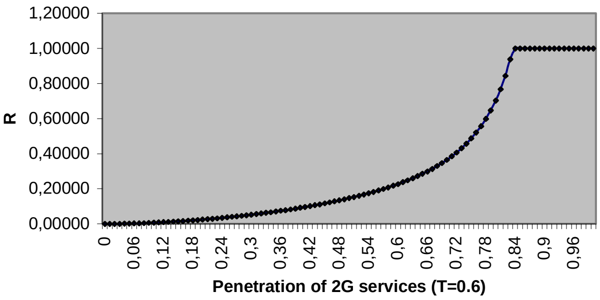

Fig. 2. The relationship between the relative number of 3G users and the penetration level of 2G services

The absolute value of the number of 3G potential users is

$$N_{3G} = R(T,\Delta)N, \qquad (22)$$

where N is the number of population in a region.

The expressions (20,21,22) give a possibility to evaluate the number of 3G potential users before deployment of 3G wireless systems in a region if the penetration level of 2G services in the region is known.

V. CASE STUDIES

In this section we consider two simple examples illustrating the submitted method. The first example is follows. According to statistical information there is 4.7 million of population ($N = 4.7 \cdot 10^6$) in one of a large European city. The penetration level of 2G services is $\Delta = 0.34$. Using the expressions (21,22) it is quite easy to estimate both the relative number of potential UMTS users in the city as $R(0.6; 0.3) = 0.067$ and the absolute number of potential UMTS users as $N_{3G} = 3 \cdot 10^5$.
In the other large city the number of population is $N = 8.5$ million and the 2G penetration level is $\Delta = 0.55$. For this city the relative number of prospective UMTS users is $R(0.6; 0.55) = 0.18$ and the absolute number of prospective UMTS users is about $N_{3G} = 1.5 \cdot 10^6$. As it is seen from these two examples in the cities there are enough potential users in order to develop 3G services.

VI. CONCLUSION

In this paper the estimation problem of prospective 3G users in a region has been considered. In particular, the solution method of the problem based on the Pareto law, the Lorenz curves, the Gini coefficient has been proposed. If the number of population in a region and the penetration level of 2G services in the region are known then the method allows evaluating the number of prospective 3G users. The use of the submitted method enables making easier the preparation of system projects for deployment of 3G equipments.

REFERENCES

[1] "Long term potential remains high for 3G mobile data services," Report # 18 from the UMTS Forum, February 2002.
[2] M. Jiang, M. Nicolic, S. Hardy, L. Trajkovic, "Impact of self-similarity on wireless data network performance," ICC 2001, USA, June 2001.
[3] K. Park, W. Willenger, "Self-Similar network traffic and performance evaluation," John Wiley & Sons, 2000.
[4] A. Krendzel, Y. Koucheryavy, S. Lopatin, J. Harju "Estimation method for data traffic generated by 3G users," the IEEE International Conference on Communications (ICC-2004), Paris, France, June 20 – 24, 2004.
[5] Y. Koucheryavy, A. Krendzel, S. Lopatin, J. Harju "Performance estimation of UMTS release 5 IM-subsystem elements," the 4-th IEEE Conference on Mobile and Wireless Communications Networks (MWCN 2002) proceedings, Stockholm, Sweden, September 9-11, 2002, ISBN 0-7803-7606-4.
[6] V. Pareto, "Le cours d'Economie Politique," Macmillann, London, 1897.
[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] A. Jipp, "Wealth of nations and telephone density," Telecommunication Journal, No.6, 1963.
[10] I.E. Varakin, "Economics, telecommunications and development of the society: macroeconomic mechanisms of telecommunications development," Electrosvyaz Journal, no. 1, 1994.
[11] M. O. Lorenz, "Methods for measuring the concentration of wealth," Amer. Stat. Assoc., no. 9, pp. 209-219, 1905.
[12] S. Kotz, N.L. Johnson, C.B. Read, "Encyclopedia of Statistical Science," New York, Wiley, 1983.
[13] C. Dagum. "The generation and distribution of income, the Lorenz curve and the Gini ratio," Econ. Appl. No. 33, 327-367, 1980.
[14] M. Kendall, A. Stuart, "The advanced theory of statistics. Distribution Theory," London, C. Griffin & Co., 1962.
[15] 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.
[16] C. Gini, "Indici di concentrazione e di dipendenza," Biblioteca delli'eccomista, no.20, 1922.
[17] P. M. Dixon, T. Mitchell-Olds, R. Woodley, "Bootstapping the Gini coefficient of inequality," Ecology 68, pp. 1548-1551, 1987.