MATHEMATICAL MODEL OF TELEPHONE SUBSCRIBER

V. GECHTMAN, D. BIRAN
Office of Chief Scientist Ministry of Communications, Tel-Aviv, Israel

ABSTRACT

In this paper an attempt has been made to present the telephone subscriber as a man - machine interactive system and to investigate it in all its interrelated aspects: Socio-Economic, Psychological, Traffic, Technical and Commercial.

The aim of the study is statistical analysis of the factors characterising the subscriber and the setting up of mathematical models which interconnect them.

Development of such models enables us to forecast the traffic characteristics of the potential telephone subscriber based on his Socio-Economic and Psychological factors.

The use of such models can significantly expand our knowledge of the telephone subscriber as a human factor in the communication system and so significantly improve the methodology of forecasting and designing of the system.

INTRODUCTION

The design of future telephone systems is usually based on a forecast of the flow of call requests.

The flow of call requests is made up of elementary flows, whose sources are the potential telephone subscribers. Each elementary flow is a random process, but its statistical characteristics are dependent on the forecasted informative behaviour of the certain potential subscriber.

The informative behaviour of the telephone subscriber expresses itself in his traffic characteristics.

The following main traffic characteristics are worth noting:

- Average load of each subscriber - general, incoming, outgoing, in Er;
- Average number of calls to and from subscriber per day;
- Distribution of the subscriber's hourly activity - per day, per week, per month, per year;
- Distribution of the hours of maximum load per day;
- Level of the subscriber's maximum load;
- Time distribution of the call stages - the time before dialling, time taken in dialling, duration of conversation, etc.
- Distribution of outgoing calls among local, inter-city and international destinations;
- Main directions for calls (individual tendencies);
- Characteristics of the samples of repeat calls - probability of repeating calls, probability of abandoning calls, etc.

The informative behaviour of the telephone subscriber and his traffic characteristics are dependent on many varied factors as: Economic, Social, Psychological, Urban, and Technical.

It can be assumed that each subscriber's average load is dependent first and foremost on the type of subscriber - commercial or private. In the case of a commercial subscriber, it is dependent on the type of business and its size.

In the case of a private subscriber - on the number of people living in the apartment and using the telephone, on their ages, on the status of family head, his profession, his level of education, etc.

The average load and average number of calls is also influenced by the family's standard of living and its place of residence - capital city, large city, large or small town, etc.

The distribution of the hourly activity and maximum load level of the private subscriber are generally dependent on his individual 'needs for communication'; Work done at home, public or political activity, social life, arrangements for reasons of health, financial, legal or entertainment arrangements, etc.

Characteristics of the samples of repeat calls are actually influenced by the age of the telephone users, by the status of each one of them (working or non-working) and by their educational level.

It is reasonable to assume that all the traffic characteristics depend also to a certain extent on the type of equipment used by the subscriber. For instance, the load that an ordinary telephone carries must be lower than that of a telephone with extension or a multiline telephone.

All Socio-Economic and technical parameters, which were mentioned previously, may well be known about a potential telephone subscriber, consequently, they can be used to forecast the traffic characteristics of these subscribers.

This forecast becomes possible by using the special statistical model, that connects the telephone subscriber's traffic characteristics with economic, social, psychological, urban and technical parameters.

The aim of this study is the development of such a model.

This model will enable us to expand our knowledge of subscriber's behaviour as the human factor in the telephone system.

It will make possible:

- Essential improvement in the forecast of the traffic load in the telephone systems;
- Improvement in the planning of telephone exchanges and telephone network according to the forecasted traffic load;
- Efficient utilization of equipment by the balancing out of the load amongst the different groups of subscribers in the telephone exchanges;
- Perceptible rise in the quality of service.

The model also enables various commercial forecasts to be drawn up. These forecasts may well be the basis for economic evaluations and certain policy decisions - such as traffic, stocks, capital investment, contracts for the purchase of equipment, etc.

OBJECT OF STUDY

The model of the telephone subscriber is developed with the aid of statistical computerized methods. This approach required collecting significant amounts of data from existing subscribers for the purpose of studying their Socio-Economic and Psychological Identity and their
traffic behaviour within the telephone system.

The telephone subscriber is represented schematically as a Black box, which includes the "human part" of the telephone system (the group of people using the telephone) and the "technical part" of the system (the telephone itself).

All the elements, which characterise the object can in general be classified into two main groups - "input" factors and "output" factors.

"Input" factors are:
- \{X\} - The Socio-Economic and Psychological parameter vector, which characterizes the human part of the system;
- \{W\} - The technical parameter vector, which characterizes the technical part of the system;
- \{Z\} - The random noise vector, which characterizes the influence of the unknown and uncertain environment.

"Output" factors are:
- \{Y\} - The traffic parameter vector, which characterizes the informative behaviour of the subscriber;
- \{E\} - The business or commercial parameter vector, which characterizes the periodic payments of the subscriber.

It is possible to present the traffic and the business parameter vectors as a function of the Socio-Economic, Psychological, and Technical parameter vectors, thus:

\[
\begin{align*}
\{Y\} &= F_1(\{X\};\{W\};\{Z\}) \\
\{E\} &= F_2(\{Y\};\{Z\})
\end{align*}
\]

This system can be considered to be a statistical model of a telephone subscriber. It is important to emphasise, that the above model is effected through statistical forecasting only, since the noise vector cannot be measured and is the uncertain and unexplained part of a subscribers traffic behaviour.

THE DEVELOPMENT OF A MODEL

The development of a statistical model is composed of two major parts: The collection of the necessary statistical data and the processing of the data in order to construct the model.

Collection data necessary for establishing this specific model is difficult due to the large number of involved elements - Socio-Economic, Psychological, Technical, Traffic and Business. Every class of parameters demands a different approach for collecting the data, analysing it and even processing it (see fig. no. 2).

The Socio-Economic and Psychological parameters are collected with the aid of a Socio-Economic survey based on a wide questionnaire submitted to subscribers.

Traffic parameters are measured in the framework of the traffic survey with the aid of the line scanner, which was developed and constructed especially for this survey.

The technical and commercial parameters were learned from existing statistics in routine computer reports.

Although the surveys differ they are based on a common method and have the same format.

The main stages of the surveys are as follows:
1. Collection of a prior data and the preparation of preliminary lists of tested parameters;

Sample of Telephone Subscribers

Computer's Reports

Commercial Parameters

Traffic Parameters

Socio-economic Parameters

Socio-economic Questionnaires

Line Scanner

Common Processing Development of a Model

Fig. No. 2 - Development of a Model
1. Determination of data collection methods (questionnaires, measurements, preliminary calculations, etc.);
2. Evaluation of the size of samples;
3. Planning, Organization, and carrying out of data collection;
4. Preparation of the algorithms and computer programmes for the data processing;
5. Logical analysis, classification and coding of data;
6. Preliminary and separate processing of data;
7. Common processing of all the data and development of a statistical model.

It should be mentioned that each survey is in fact carried out three times:

- The first time as a pretest, which is mainly designed for learning and analysis of a prior data and also for testing preliminary assumptions and hypothesis.
- The second time as a pilot experiment, aimed at the collection of background material for the planning of the comprehensive survey.
- The third time as a comprehensive survey on a wider sample of subscribers, which will include varied geographic and Urban areas and different groups of subscribers.

Initially all the surveys were intended to run simultaneously and to pass through the stages of pretest, pilot and comprehensive survey at the same time. In fact, however, due to circumstances, these surveys were moved, thus one of them was fed by the results of the other, with regard to methodology, organization and content.

PILOT EXPERIMENT

The Socio-Economic survey of telephone subscribers was the first one to be carried out in 1976 in the form of a pretest on a sample of 1,000 subscribers connected to the same exchange. The results of the pretest were discussed at the “Convention Information”, which was held in Paris in 1977.

In 1978 the decision was taken to expand the pretest and to send out the Socio-Economic questionnaires to all the subscribers of the tested exchange. This decision enables us to achieve the objectives of the pilot experiment within a reasonable amount of time and without excessive investment.

In the framework of the pilot experiment, 9,086 questionnaires were sent to private telephone subscribers and 727 questionnaires were sent to commercial subscribers. In a special letter attached to the questionnaires, the objectives of the survey were detailed and complete secrecy of data recorded in the questionnaires was assured.

In general the subscribers responded to our request, understood the objectives of the survey and cooperated.

The results received from the Pilot Test confirmed the prior results of the pretest. The percentage of subscribers which answered the questionnaire was 37%.

Thus, after a logical analysis, classification and improvement, 2,570 questionnaires were found to be complete and eligible for processing, of these 2,403 were private subscribers and 167 commercial subscribers.

It was decided, as a first stage to concentrate on private subscribers.

For the processing of the survey’s results, the data obtained from the questionnaires was represented by means of 15 defined socio-economic and psychological parameters:

X1 - "Tendency to make a call" coefficient
X2 - The international call tendency coefficient.
X3 - Status of family head.
X4 - Profession of family head.
X5 - Age of family head.
X6 - Education.
X7 - The continent of origin.
X8 - Immigration coefficient.
X9 - Standard of living.
X10 - Size of the family.
X11 - Family employment coefficient.
X12 - Family education coefficient.
X13 - Family youth coefficient.
X14 - Family old age coefficient.
X15 - Employer participation coefficient.

The number of defined factors is much smaller than the number of questions in the questionnaire, despite the fact that all the questions were utilized in the list of the factors.

It can also be seen that most of the factors were introduced as relative coefficients, ranging from 0 to 1.

For example:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immigration Coefficient</td>
<td>The no. of years the family's head is living in Israel</td>
</tr>
<tr>
<td>Employment Coefficient</td>
<td>The no. of working people in the family</td>
</tr>
<tr>
<td>Family Youth Coefficient</td>
<td>The no. of members of the family under the age of 14</td>
</tr>
</tbody>
</table>

Other coefficients have been devised in a similar manner:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immigration Coefficient</td>
<td>The no. of years the family's head is living in Israel</td>
</tr>
<tr>
<td>Employment Coefficient</td>
<td>The no. of working people in the family</td>
</tr>
<tr>
<td>Family Youth Coefficient</td>
<td>The no. of members of the family under the age of 14</td>
</tr>
</tbody>
</table>

Two quantitative factors - X5 - the age of the family head and X10 - the size of the family - were introduced directly as absolute numbers.

The rest of the qualitative factors were scaled - a special scale was chosen for each factor, in accordance with its content.

In order to construct the “Standard of living” factor, three questions from the questionnaire were used - the kind of home (owned house, owned apartment, rented apartment), the number of rooms in the house (in the apartment) and car ownership.

In order to construct the “Tendency to make a call” coefficient the subscriber’s "needs for communication" were used.

In the pilot experiment, as in the pretest, only one traffic parameter was introduced - the number of call unit charged. This parameter was obtained from telephone bills.

Clearly, then, this component gives only a rough approximation of the traffic characteristics vector.

In future within the comprehensive survey, the measurement of these characteristics will be carried out by using a line scanner. Preliminary measurements have already been taken in an experimental exchange.

For the purpose of these measurements, a sample of 250 subscribers was chosen at random from the whole sample of 2,403 private subscribers, who participated in the Socio-Economic survey, and whose questionnaires were found to be complete and eligible for processing.

Every subscriber included in the measurement sample will be connected to a line scanner and will be checked over a period of two months - both on weekdays and weekends. The period of two months is the period of the telephone bill and was chosen to match the traffic parameters with commercial parameters of the same period.

For processing the data a programme file known as SPSS was employed ("Statistical Package for the Social Sciences", Multiple Regression Analysis).
THE MODEL

The results of the processing were received in the form of regression, which connects the traffic parameter - the average number of call units charged - with the 12 Socio-Economic parameters:

\[
Y = -6.8 + 2.6 X_3 + 8.8 X_4 + 0.3 X_5 + 1.8 X_6 + 9.4 X_7 + 6.2 X_8 + 9.9 X_9 + 0.5 X_{10} + 3.2 X_{11} - 7.4 X_{13} - X_{14} + 0.6 X_{15},
\]

Where

- \(Y\) - the average number of call units charged;
- \(X_3\) - the Socio-Economic parameters;
- \(X_{15}\) - the Socio-Economic parameters.

The multi-correlation coefficient \(R = 0.78\). The explanation level \(R^2 = 0.61\).

All the Socio-Economic factors which affect the number of calls made are considered relevant - the correlation, coefficients are significant by the F-test.

The most influential factors are:
- Standard of living \((Z_6 = 0.65)\), profession of family head \((Z_5 = 0.40)\), education of family head \((Z_5 = 0.40)\).

Although the multi-correlation coefficient by the F-test is significant, the explanation level is not particularly high.

Consequently, for the purpose of evaluating the possibilities of using the model, we had to test its forecasting ability. The test was carried out by using additional subscribers who did not participate in the process of the construction of the model. (See fig. no. 3).

To test the forecasting ability the forecasted parameters are compared with the real parameters, compiled from additional subscribers.

In our case an additional sample of 1000 subscribers was chosen and divided into 10 subscriber's groups.

In addition, the sample was separated into four traffic groups, according to average number of call units charged:

<table>
<thead>
<tr>
<th>Traffic Group</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0)</td>
<td>Inactive subscribers - about 100 call units;</td>
</tr>
<tr>
<td>(1)</td>
<td>Subscribers of low activity - 101 - 400;</td>
</tr>
<tr>
<td>(2)</td>
<td>Subscribers of moderate activity - 401 - 1000;</td>
</tr>
<tr>
<td>(3)</td>
<td>Subscribers of high activity - 1000+</td>
</tr>
</tbody>
</table>

Forecasted traffic groups were calculated for each subscribers of the sample.

The results of the forecast - the number of the forecasted group - compared with the number of the real group and the deviation of the forecast was calculated:

\[
\Delta i = |\hat{A}_i - t_i|
\]

Where

- \(\Delta i\) - deviation of the forecast for the subscriber number \(i\);
- \(\hat{A}_i\) - number of the forecasted group;
- \(t_i\) - number of the real group.

(See Tab. no. 1. on page 5).

The results of the forecast by model were compared to the results of the forecast by usual telephone administration method.
Tab. no. 1. RESULTS OF THE FORECAST

<table>
<thead>
<tr>
<th>No. group subs.</th>
<th>% of the instances with the deviation equal to:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
</tr>
<tr>
<td>3</td>
<td>100</td>
</tr>
<tr>
<td>4</td>
<td>100</td>
</tr>
<tr>
<td>5</td>
<td>100</td>
</tr>
<tr>
<td>6</td>
<td>100</td>
</tr>
<tr>
<td>7</td>
<td>100</td>
</tr>
<tr>
<td>8</td>
<td>100</td>
</tr>
<tr>
<td>9</td>
<td>100</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
</tr>
</tbody>
</table>

Average to 1000 subscribers 62.3 36.0 1.7 27.2 61.9 10.9

INITIAL APPLICATIONS

One of the possible applications of the model is a forecast of traffic groups among potential subscribers. The process of the application is carried out in accordance with the Fig. no. 4.

According to the above process, we are at present conducting a preliminary test for the forecast of traffic groups among people requesting a telephone, in the new exchange G» For this experiment, a random sample of 1000 people requesting a telephone was selected.

A Socio-Economic questionnaire, especially prepared for this case, was sent to each member of sample group. Nearly 90% of these questionnaires were returned.

The information contained in the questionnaire was used as a basis for the forecast of traffic characteristics among people requesting a telephone. The results of the forecast were passed on to the telephone administration in order to achieve optimal allocation of telephone numbers.

A test of the forecast's accuracy will be carried out at a later stage, when the new exchange is brought into service. Measurements will be taken by using the line scanner and will be compared with the earlier forecast.

It should be stated that the model, which was discussed here, is only preliminary and has certain limitations.

(a) The model was built on the basis of the results of the pilot experiment, which was carried out on a narrow sample of subscribers, connected to only one exchange with a homogeneous population

(b) Instead of the number of calls, the number of call units charged was taken as a dependent variable - a parameter, which is known to be significantly affected by the level of service.

These limitations will be eliminated in the future, by expanding the sample to include large numbers of exchanges from different geographic areas and also by use of the line scanner to measure traffic characteristics.

At the same time must be noted that the regression obtained forms an informative behavior model of human beings, for which reason, part of it will always remain unexplained.

But notwithstanding the human factor, the results obtained indicate significant links between Socio-Economic and traffic parameters which can doubtless be used for forecasting.

We trust that the present study will contribute towards broadening the knowledge concerning the telephone subscriber and his behaviour and make it possible in the future to locate and evaluate the role played by the human factor in telephone systems.

REFERENCES