AN ANALYSIS OF TRAFFIC VARIATIONS IN THE FRENCH TELEPHONE NETWORK

Annie PASSERON and Simone RIVAT

CENTRE NATIONAL D'ETUDES DES TELECOMMUNICATIONS, PAA/ATR
Issy Les Moulineaux, France.

ABSTRACT

This paper deals with a statistical study on traffic measurement data. The study is especially concerned with traffic variation analysis with the goal of comparing different possible ways of choosing representative values for use in final group engineering.

Analysed data concern outgoing traffic in transit centres and were recorded by the French traffic supervisor in 1982 and 1984. They mainly take the form of moving daily peak hour trunk group offered loads.

Statistical tests are first used to check normal and gamma distribution hypotheses for traffic load variations. We then compare the relative merits of different possible methods for trunk group sizing.

1. INTRODUCTION

The general goal of this study is to compare different possible ways of choosing representative values from traffic measurements, for both telephone traffic administration and final group sizing.

More precisely, the study consists in the statistical analysis of a set of traffic data which is large enough to be representative of a national long distance telephone network. This analysis is carried out with two aims. The first is to judge the existing French procedure for choosing representative values in relation to sizing criteria. The second is to simulate other possible methods in order to be able to propose a different procedure which would lead both to improved daily grades of service and decreased network costs.

The existing process for trunk provisioning was defined several years ago when very few traffic measurements could be made and when customers had rather homogeneous needs. Traffic characteristics were modelled simply, resulting in the following system:

- each month at the time consistent busy hour [1], five hourly load measurements are made on all trunk groups; the second highest value is kept as the monthly representative value; the yearly representative value is the second highest monthly value in the year. The latter is thus a sort of busy season representative value. In the sizing process, the Erlang formula and Wilkinson's ERT are applied to the yearly representative values of offered traffic for a given blocking probability on final groups (currently 1%).

In fact, trunk dimensioning is planned once a year using forecasts of yearly representative values on groups at the end of the study period and is monitored each month using the latest traffic measurements in conjunction with the above sizing models. The performance of the existing network dimensioning method may be directly monitored from measurements on final groups which constitute last choice routes for calls trying to reach their destination. These final groups therefore constitute an important class of trunk groups.

This process may be considered inadequate in the present network for essentially two main reasons.

Firstly, the magnitude of season to season, day to day or hour to hour variations seems to be increasing while the position of the post selected busy hour varies within a wide interval. For instance, for a set of 88 trunk groups, Table I gives the width of the period of the day containing the post selected busy hour during each of 11 working days in May 1984. We exclude busy hours in off-peak call charging periods. The busy hour appears to fluctuate widely. This effect is accentuated for small trunk groups or trunk groups going to transit centres.

Table I: Width of the busy period

<table>
<thead>
<tr>
<th>Number of hours</th>
<th>&lt;2</th>
<th>2 to 3</th>
<th>3 to 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of groups (%)</td>
<td>15</td>
<td>18</td>
<td>67</td>
</tr>
</tbody>
</table>

Secondly, the use of electronic exchanges should permit more sophisticated traffic measurements. A statistical study on detailed traffic data has therefore become necessary to improve network administration and trunk provisioning procedures.

This paper deals only with hourly traffic loads and is not concerned with traffic variations inside the hour.
Analysed data are more detailed than those used in the current trunk provisioning method as described above. They are extracted from traffic records made by the new French traffic supervisor which is now being brought into service. These records only concern trunk groups outgoing from transit centres towards either subscriber centres or other transit centres. In the French network organization, the former are in fact final choice or other transit centres. In the French network records only concern trunk groups outgoing from groups while some of the latter may overflow on to a special transit centre but only for security reasons. In the planning process, they are all sized for a grade of service of 1% blocking probability.

Before comparing sizing methods on these data, a preliminary statistical analysis of traffic load variations is made leading, in particular, to theoretical distribution functions for traffic offered to trunk groups.

2. AVAILABLE DATA

2.1. The French Long Distance Network Supervision Centre

After some preliminary trials [8], since 1982, there exists in the French telephone network a supervision centre, whose purpose is to monitor the grade of service of traffic outgoing from regional transit centres.

In each supervised centre a terminal station collects traffic measurements describing the state of outgoing trunk groups and some register groups, and sends these data to the supervision centre every fifteen seconds. Measurements are supposed to be made every working day, all day long from about 9 a.m. to about 11 p.m. For each trunk group, raw data include:

- carried load, number of seizures, number of successful seizures in every four minute period,
- test of all trunks busy state every fifteen seconds,
- number of trunks in service every eight minutes.

As data are recorded on a magnetic tape (1 tape per day), off-line data processing may be performed. Each day, for the normal charging period of the day and for the remainder of the day, the traffic load Moving Daily Peak Hour (MDPH) is calculated, to within an accuracy of 4 minutes, for each individual trunk group and for each centre. Load curves for different trunk groups or centres or groups of them may also be established.

2.2 Data Volume

Data were collected during the whole of 1982 and a large part of 1984. Data for 1983 were very incomplete. In fact, the system is still being brought into service and is not yet completely operational. Due to data transmission errors and equipment failures, measurements are available for less than 17 working days each month.

During 1982 an 1984, some centres were newly connected while others were closed down (reflecting the fast modernization and digitalization of the French network). Overall, we dispose of observations for 58% of working days in 1982 and 49% in 1984, but not for all centres.

After data validation (for zero data, clock coherence and observations breaks), 57% of possible daily measurements from all centres could be kept in 1982 (53% in 1984), among which 83% (76% in 1984) were complete, i.e. all day long, for all trunk groups.

For this study we selected a set of consistent data large enough to be representative of the French network.

2.3. Data Used In The Study

We analysed only hourly loads on final groups outgoing from three regional transit centres, located in Limoges, Nice and Orléans in 1982, and from one, located in Limoges in 1984.

We mainly took into account MDPH loads which are defined during the normal charging period of the day for all available daily measurements.

Table II shows how much data we obtained throughout 1982 for the above three centres. Trunk groups which were not present during the whole year do not appear.
Data are quite evenly distributed from January to November (5 to 14 days per month) and among different days of the working week.

Table II : Studied data for 1982.

<table>
<thead>
<tr>
<th>Centre</th>
<th>Number of groups</th>
<th>Maximum number of days</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>total to transit centres</td>
<td></td>
</tr>
<tr>
<td>Limoges</td>
<td>93</td>
<td>111</td>
</tr>
<tr>
<td>Orleans</td>
<td>79</td>
<td>99</td>
</tr>
<tr>
<td>Nice</td>
<td>76</td>
<td>97</td>
</tr>
</tbody>
</table>

During 1984 Orleans and Nice transit centres were replaced by time division systems and data from Limoges were kept for only 62 days.

Raw data from May 1984 have also been analysed with the aim of making some rough comparisons between MDPH loads and hour by hour peak hour (HDPH) loads (i.e. loads measured between 8 am and 9 am, 9 am and 10 am, etc.).

By using an hourly carried load, \( T_e \), and the corresponding number of all trunks busy states, \( N_O T \), observed in 240 fifteen second cycles it is easy to calculate the offered load to be studied:

\[
T_O = \frac{T_e}{(1-N_O T/240)}
\]

\( N_O T \) was never very high and was often zero, trunk groups tending to be overdimensioned. For some groups (60% of studied trunk groups) it was not possible to get this information and we made the approximation:

\[
T_O = T_e
\]

3. TRAFFIC LOAD DISTRIBUTION

It is well known that measured busy hour loads vary from day to day and studies aiming to analyse and model this variability are far from being a recent phenomenon. In [10], for instance, day to day variations were studied during a busy season of 6 weeks on a set of data on Time Consistent Busy Hour (TCBH) loads concerning some 20 trunk groups. A gamma distribution was fitted to the traffic load data and tested graphically. In [6], combined normal distributions were tested graphically for a set of measurements made during two consecutive years and concerning hour by hour loads.

In this section we try to update these results using recent data. The analysis is made for each available trunk group and concerns MDPH loads. In order to throw light on the best choice of yearly representative values, variations are studied inside the whole year and inside a practical busy season.

3.1. Histograms

Figures 2 and 3 show two typical shapes of the histograms we plotted for MDPH offered load for each of the above trunk groups.

3.2. Choice Of Distribution Functions

3.2.1. Normal Distribution [4]

This distribution is symmetrical and unimodal. The probability density function is:

\[
f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[ -\frac{1}{2} \left( \frac{x-m}{\sigma} \right)^2 \right]
\]
It is defined for real values of $x$. Given a sample of $n$ values $X_i$,

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

is a maximum likelihood unbiased estimator of $\mu$ and

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

is an unbiased estimator of $\sigma^2$.

The normal distribution can be used as an approximation for other distributions. A known distribution may be replaced by a normal distribution with the same expected value and standard deviation (method of moments). This distribution is easy to use in calculations.

3.2.2. Gamma Distribution Or Type III Of Pearson's System [4]

The probability density function is defined for positive $x$ and (if only two parameters, $\alpha$ and $\beta$, are used) is of the form:

$$g(x) = \frac{x^{\alpha-1} \exp(-x/\beta)}{\beta^\alpha \Gamma(\alpha)} \quad (\alpha > 0 \ ; \ \beta > 0)$$

where $\Gamma(\alpha)$ is the gamma function.

To estimate $\alpha$ and $\beta$, the method of moments may be used leading to the following simple formulae:

$$\frac{\sum X_i}{n} = \bar{X}$$
$$\frac{\sum X_i^2 - \left(\sum X_i\right)^2}{n-1} = \bar{X}^2$$

where $\bar{X}$ and $\bar{X}^2$ are the arithmetic mean and second central moment of a sample of $n$ values $X_i$.

However, these estimators may be less accurate than maximum likelihood estimators $\hat{\alpha}$ and $\hat{\beta}$, for which the following approximation may be used if $\frac{\hat{\alpha}}{\hat{\beta}}$ is large enough:

$$\hat{\alpha} \hat{\beta} = \sum X_i$$

$$n-1 \sum \ln X_i - \ln \hat{\alpha} \hat{\beta} = \ln (\hat{\alpha} - 0.5) - \ln 2$$

For comparison, we used both estimation methods in this study.

The gamma distribution has a single mode if $\alpha > 1$, and, as $\alpha$ increases, it tends to the normal distribution.

3.2.3. Kolmogorov-Smirnov Distribution Test

This test is used to test gamma and normal distribution hypotheses for traffic loads. Given a sample of $n$ ($n > 35$) independent observations the Kolmogorov-Smirnov test tests whether or not the data may be considered as a random sample of observations from a specified distribution.

For the test, the largest absolute deviation between the sample cumulative distribution and the theoretical distribution is computed. This deviation is compared to a value which is exceeded with a probability of 0.10, under the specified hypothesis. This value is computed using the Kolmogorov-Smirnov limiting distribution.

3.3. Test Results

Table III shows results for data on trunk groups observed in the whole year 1982.

<table>
<thead>
<tr>
<th>origin centre</th>
<th>normal</th>
<th>gamma</th>
</tr>
</thead>
<tbody>
<tr>
<td>Limoges</td>
<td>71</td>
<td>75</td>
</tr>
<tr>
<td>Orleans</td>
<td>67</td>
<td>67</td>
</tr>
<tr>
<td>Nice</td>
<td>86</td>
<td>86</td>
</tr>
</tbody>
</table>

In fitting the gamma distribution, in particular for trunk groups going to transit centres, we sometimes find very high values of $\alpha$. In this case we use the normal approximation. In all cases $\alpha$ is never less than 10. Both estimation methods give almost identical parameters ($\alpha, \beta$) on our specific data (except for a few small trunk groups), which leads us to select the simple method of moments.

For some trunk groups, only one of the two tested distributions is accepted. For instance, the gamma distribution may be accepted for very small trunk groups when the normal distribution hypothesis is rejected. The reason may be that data are near zero but do not include negative values for which the theoretical normal distribution is defined.

Results are not significantly different if we distinguish trunk groups going to transit centres from those going to subscriber centres. In general, if both normal and gamma distributions are not accepted, the coefficient of variation is quite high. The rejection decision does not appear to depend on the level of the traffic load.

For Limoges in 1982 and in 1984, we have tried to define a practical busy season by looking for a period of four consecutive months for which the mean load is the highest in the year for most trunk groups. In fact, such a busy season is only obvious for trunk groups going to subscriber centres. Table IV shows that statistical tests give better results with data taken only in the busy season than with data taken in the whole year. The above comments on gamma fitting are still valid for these data.

<table>
<thead>
<tr>
<th>year</th>
<th>normal</th>
<th>gamma</th>
</tr>
</thead>
<tbody>
<tr>
<td>1982</td>
<td>89</td>
<td>88</td>
</tr>
<tr>
<td>1984</td>
<td>93</td>
<td>94</td>
</tr>
</tbody>
</table>

In conclusion, the normal approximation for busy season data appears to be acceptable for most trunk groups. Because of its remarkable pro-
4. FINAL GROUP SIZING

4.1. Different Possible Methods

The objective is to compare different methods for final group sizing. All these methods need traffic data which should be chosen so as to represent the highest loads to be carried during the reference period (one year, for instance). After sizing, a final group is considered to offer a good service if it always meets a given grade of service, estimated, for example, by the proportion of blocked calls, day after day, with the possible exception of a few busy hours. For each sizing method, the above set of traffic data allows us to simulate, for a large number of trunk groups the number of days or hours in a whole reference period for which different blocking levels are exceeded. We suppose that the blocking function which gives the blocking probability on a group from its number of trunks and its offered load is the Erlang function.

Following [2] and [9], we distinguish three approaches as follows.

a) A sufficiently high reference load is chosen and the group is sized to a given blocking probability (using the blocking function).

b) The group is sized for an average blocking objective in a busy season. The distribution function of the load and the blocking function are used.

c) The hourly load which will not be exceeded more than once in a given period is estimated by using an extreme value distribution. The group is sized for this reference load.

The first approach includes both the existing French process, as detailed in the introduction, and the CCITT recommendations [1], i.e. the reference load is the mean of the 30 highest TCBH loads in the year. The second is used in North America [5]. The third belongs to Extreme Value Engineering (EVE).

Taking into account results of the preceding section, we analyse data only on trunk groups for which the normal distribution was accepted.

4.2. Comparison Between CCITT And Existing French Procedures

The comparison is made using MDPH data instead of the TCBH data. From MDPH data we define an equivalent French Monthly Representative Value (YRV) based on order statistics. For instance, if 5 MDPH values are available in a month, we keep the second highest, but if 11 MDPH values are available, we keep the fourth highest. A preliminary comparison was made between these supervisor MRVs and the current TCBH MRVs on Limoges data during the whole of 1982.

Depending on the size of each trunk group, the supervisor MRVs are between 4% and 25% higher than those of the TCBH. As the CCITT definition is also applied to MDPH loads by taking the mean of the 30. \( n/250 \) highest of \( n \) available MDPH data (\( n = 111 \) for Limoges, cf Table II), the comparison is thought to be meaningful.

Table V gives global results of the comparison for data in 1982. With the CCITT method the Yearly Representative Value (YRV) is exceeded, on average, on 11 days per year, i.e. 4% of working days. With the French method the YRV is exceeded, on average, on 37 days per year, i.e. 15% of working days.

Table V : Undervalue of French YRV (in %) compared to CCITT YRV.

<table>
<thead>
<tr>
<th>Limoges</th>
<th>Orleans</th>
<th>Nice</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.7</td>
<td>9.8</td>
<td>9.6</td>
<td>9.3</td>
</tr>
</tbody>
</table>

The average traffic load undervalue of 9.3% leads to an average undervalue in the number of circuits of 8.5% for a blocking probability equal to 1%.

The mean of the 3 highest data in the year for each trunk group is, on average, 23% higher than the French YRV.

The normal distribution modelling allows us to approximate each of the above YRVs with a function the expected value, \( m \), and the standard deviation \( \sigma \):

\[ YRV = m + K \cdot \sigma, \]

of which an unbiased estimator is

\[ \bar{X} + K \cdot s \] (\( \bar{X} \) and \( s \) are defined above).

\( K \) is computed from the normal distribution function applied to our data.

On average, we find:

\[ K = 1.73 \]

for the CCITT YRV and

\[ K = 1.04 \]

for the French YRV.

Such a model could be used to compare both methods on a new small sample of data or to predict the performance of these methods.

4.3. Comparison Between CCITT Recommendations And Average Blocking Method

Only the Limoges data from 1982 are used.

Average blocking, \( B \), on a trunk group of \( N \) circuits is calculated as follows [5]:

\[ \overline{B} = \int_{0}^{\infty} E[a,N]f(a)da, \]

where \( f(a) \) is the normal probability density function of the distribution of loads offered to the trunk group and \( E[a,N] \) is the Erlang blocking function.

For each trunk group, \( f(a) \) is estimated for the month where the average load is highest. Sizing is then made successively with the CCITT YRV and for an average blocking level of 1%.

Roughly speaking, if the coefficient of variation is lower than 0.15, the average blocking criterion leads to fewer trunks than the CCITT criterion; the opposite is true if this coefficient is greater than 0.15.

The average blocking method seems better suited to traffic variability and, when used over
the highest month of a year, sizing is, on average, only 4.2% higher than that of CCITT recommendations.

4.4. First Comparisons With EVE

Limoges data from May 1984 are used for this comparison.

In telecommunications two extreme value distributions have already been proposed. In [7], the classical extreme value distribution function (also called the Gumbel distribution) is used:

\[ Pr[X < x] = \exp[-e^{-\beta(x-a)}] \]

This is a limiting distribution of the greatest value in a random sample of infinite size. In [3], the normal distribution power 6 is proposed. In order to test these distributions we first calculate the HDPH of 7 typical trunk groups using the traffic profiles of several working days. Figure 4 shows, for instance, the traffic profile observed on May 10, 1984 on a medium size trunk group.

These data are used to estimate the parameters of the above distributions. In EVE, the load used for sizing is that which will be exceeded once on average during a return period (usually 20 working days), as calculated from an extreme value distribution. After estimation of parameters by the method of moments, the first distribution leads to the return period load:

\[ r_{20} = \bar{X} + 1.866 \text{ s} \]

and the second gives:

\[ r'_{20} = \bar{X} + 1.744 \text{ s} \]

By taking a return period equal to the number of available data we could check the number of times the return period load is exceeded for each of the two distributions. However, due to the small values of the coefficient of variation it is not possible to draw any positive conclusions. Further studies are necessary but as, on average for the 7 trunk groups, the two definitions of \( r_{20} \) differ by only 1.4%, we make the arbitrary choice of the normal 6 modelling. We come to the same conclusions if we calculate a return period load using MDPH data for May 1984.

In comparing the return period load, for hour by hour data (a), the return period load for MDPH data (b) and the MDPH French YRV (c) for May 1984 we find that (c) is always the lowest, that (a) is, on average, 10.9% higher than (c) and that (b) is on average, 17.8% higher than (c).

Lastly, we compare the return period load \( r_{20} \) calculated from MDPH data over 4 months to the equivalent CCITT load calculated from the same data. We find practically the same loads, which is not surprising since the CCITT load was modelled by:

\[ \bar{X} + 1.73 \text{ s} \quad \text{and} \quad r_{20} \text{ is calculated with} \quad \bar{X} + 1.744 \text{ s} \]

Another reason is that both loads have similar definitions on MDPH data: In 1982 on our data the CCITT load was exceeded on 4% of working days on average and \( r_{20} \) is theoretically the load which will be exceeded, on average, once in 20 working days.

5. CONCLUSIONS

Data recorded by the French long distance supervision centre have supplied rich information on traffic load variations. Daily peak hour data may be defined on trunk groups for each working day with a precision of 4 minutes. When, ultimately all regional transit centres are connected, the supervisor could be used to provide data for trunk provisioning and for validating traffic variation models in addition to its primary function of traffic supervision.

From the present study it appears that the normal approximation for traffic variations within the year is acceptable and is as good as an approximation by the gamma distribution. Moreover, the normal approximation is even better for data in a practical busy season of 4 consecutive months.

Concerning YRVs used for trunk group sizing, we first showed that, with our data, MDPH loads are between 4% and 25% higher than TCBH loads. If considering only MDPH data, the French method leads to a YRV which is exceeded, on average, on 15% of working days while the CCITT YRV would be exceeded on just 4% of working days. An average blocking method applied to a busy season (in the study we took the highest month in a year) leads to nearly the same sizing as the CCITT method while closely fitting traffic variability and being more easy to apply. Only busy season measurements are needed and sizing may be realized by use of the normal distribution. EVE applied to a period of 4 months gives similar results but in this field further studies are necessary.
Concerning traffic monitoring and monthly representative values we have found tentative relations between engineering with hour by hour data, MDPH data and French MRV data. This analysis requires further study using additional data.

REFERENCES


