MEASUREMENT AND ANALYSIS OF SUBSCRIBER TRAFFIC RATES

E. W. GREGORY, D. J. SONGHURST

ABSTRACT
The principal method used by the British Post Office to predict future levels of telephone traffic for individual exchanges requires knowledge of the average quantities of traffic per line originated by both residential and business customers during the exchange busy hour. Consequently considerable efforts have been, and are being, devoted to the problems of estimating these 'class traffic rates'. In particular, consideration is being given to estimating them directly by statistical methods using random samples of connections.

Pilot studies to investigate this approach are being undertaken at a number of exchanges and this paper discusses the method of analysis proposed and presents some preliminary results and provisional conclusions. A statistical model is proposed for the observations and used to produce an expression for the variance of the traffic rate estimate obtained from the random sample measurements. Methods of estimating this variance from sample data are proposed and used to indicate the effects of
- increased sample sizes.
- extended measurement periods.
- reduced equipment scanning rates.

on the accuracy of the traffic rate estimate.

INTRODUCTION
The British Post Office (BPO) uses an integrated approach to the forecasting of future levels of busy hour telephone traffic for individual exchanges. Exchange busy hour originating traffic forecasts are not produced in isolation but are directly related to expected future changes in system size (numbers of lines) and composition (the relative proportions of residential and business connections). Specifically, future exchange busy hour traffic predictions are derived from the product of independent forecasts for the numbers of residential and business lines and the corresponding average levels of traffic per line originated by each category of subscriber.

Business and residential are the two main rental categories employed by the BPO. A small number of exchange lines will usually belong to neither category, being used for provision of public call office etc. The total traffic generated by these 'unclassified' lines is usually measured and accounted for separately. Alternatively, they may be included in the business category. In either event we may consider the traffic quantity of interest to be the sum of the traffic generated by residential and business connections.

Formally, the required forecast, $A_t$, of busy hour traffic at time $t$ ahead is produced from the relationship

$$A_t = B_t + B_t$$

where $B_t$ and $B_t$ are the forecast numbers of residential and business connections at time $t$ respectively and $R_t$ and $R_t$ are the corresponding values of busy hour traffic per line originated by each class of subscriber.

Traditionally the quantities $R_t$ and $B_t$ are referred to within the BPO as 'class calling rates'. This choice of terminology is unfortunate in that the quantities are in fact measures of traffic not calls. In this paper therefore we adopt the more accurate description 'class traffic rates' (CTR).

It is important to take separate account of residential and business traffic growth in this way, not only because the corresponding traffic rates differ appreciably but also because past experience demonstrates that future system growth and trends in calling behaviour will usually be different in each case. Generally speaking, residential telephone demand in the UK is strongly influenced by changes in tariffs but is relatively insensitive to variations in economic conditions, at least in the short term. Business growth on the other hand is, within broad limits, relatively insensitive to tariff increases but sensitive to economic fluctuations. Thus, ideally, each market sector needs to be treated separately and the BPO method represents one way of achieving this aim.

However the approach is only viable provided that reasonably accurate forecasts can be made of business and residential connection and traffic rate quantities and an essential prerequisite for this is the availability of regular and reliable information on present traffic rate values. The advantages of separate class traffic rate forecasting will be largely negated if the CTR values used are inaccurate. Consequently in recent years the BPO has undertaken a number of studies of methods which have been proposed for deriving CTR values in order to establish the accuracy of the estimates produced by, and the costs associated with, each method. In particular the possibility of using estimates derived from random samples of connections has been, and is being, extensively examined.

From the purely statistical viewpoint, the measurement of random samples of connections belonging to a particular class represents potentially the most desirable method of evaluation. Not only is the selected sample guaranteed, in a specific sense, to be representative of all exchange connections of that class but the statistical basis of the selection process also enables the accuracy of the estimate obtained to be inferred from the data. This is an important advantage compared to alternative approaches previously examined by the BPO which, in general, depend on assumptions of unknown veracity and consequently give rise to CTR estimates of indeterminate precision. In general too, only one class of connection needs to be sampled, an assessment of the CTR for the remaining class being readily obtained provided only that an estimate of total originating traffic is available (by definition, all exchange lines belong to one class or the other - see above).

The selection, identification and subsequent measurement of random samples of exchange connections in a particular class is, however, a relatively expensive and time consuming matter. Before considering application of the method generally within the BPO therefore, a trial program of exchange measurements on samples of both residential and business connections in being undertaken using a measurement device designed by the BPO. This
device estimates CTR values using the switch-count method with a 5 minute scanning interval. The purposes of this trial are:

- to examine the operational practicality of the procedures involved and derive a provisional estimate of costs.
- to study the effects of different sample sizes and measurement periods, equipment scanning rates etc on estimation accuracy, in order to determine the most appropriate values for practical use.

This paper concentrates on the second of these objectives, describes the method of analysis proposed and provides details of the information collected so far.

**ANALYSIS METHOD**

The data available for each sample consists of the average traffic originated by each sampled connection over 5 days (Monday-Friday) during the exchange busy hour, a separate value being obtained for each week during the period measured. Thus, if a sample of size n connections is measured for w weeks there will be nw values xij, where xij refers to the value associated with sample connection i in week j.

The method of analysis proposed is described in detail in Appendix 1. The main points to note are that the required CTR value is estimated by the average level of traffic observed on the sampled connections during the measurement period and that the accuracy of this estimate can be assessed from the measurement data itself. In general, this variance can be expressed as the sum of 3 components, as follows:

- A component reflecting uncertainty in the estimate due to sampling; the contribution from this component decreases with increasing sample size.
- A contribution reflecting uncertainty due to random variations in the traffic originated by individual connections in different weeks (including the effect of switch-count errors in the measurement process); this component can be reduced by increasing either the size of the sample or the duration of the measurement period or by decreasing the scanning interval of the equipment (but see Appendix 2).
- A contribution associated with weekly fluctuations common to all connections; in practice it was expected that this contribution might well be negligible and this expectation has been borne out by the limited results available so far. Nevertheless, it was considered essential to allow for the possibility of a significant effect in the analysis.

The above can be quantified by the equation (see Appendix 1)

\[
\text{var}(E) = \left( \frac{N-n}{N} \right) \left( \frac{S^2}{n} + \frac{T^2}{w} + \frac{V^2}{wN} \right)
\]

where \( E \) = CTR estimate

\[ n = \text{sample size} \]

\[ N = \text{total lines, in class being sampled, on the exchange} \]

\[ S^2 = \text{'residual variance'} \]

\[ V^2 = \text{'between weeks variance'} \]

\[ T^2 = \text{'between samples variance'} \]

Unfortunately, measurements from only two exchanges, both of samples of residential connections, have been completed in time for inclusion in this paper. Whilst such limited data does not enable conclusions of any great generality to be reached, discussion of the results obtained may still be of interest as an example of the general approach envisaged once more extensive data becomes available. The exchanges concerned are:

- Brixton, an exchange in central South London, where a random sample of 288 of the 3072 residential connections on the exchange were measured for a 4-week period.
- Leicester, a provincial city centre exchange, where a small sample of 86 residential connections from a total of about 6000 were measured, again over a 4-week period.

Both sets of measurements were taken in late 1978. (Complete data is available from the authors on request).

In both cases calculation of the Analysis of Variance table (see Appendix 1) indicated, as expected, that the 'between weeks' component of variance was not significant and this was therefore taken to be zero in the subsequent analysis. The CTR estimates and their accuracy, and the values calculated for the quantities \( S^2 \) and \( V^2 \) are shown in Table 1. CTR values are expressed in milli-erlangs (m.e.). The components of variance are expressed as (m.e.)². The relative standard deviations, sd, of the estimates are expressed as a percentage of the calculated CTR values. The high level of uncertainty associated with the Leicester estimate is apparent.

<table>
<thead>
<tr>
<th>Exchange</th>
<th>CTR</th>
<th>( S^2 )</th>
<th>( V^2 )</th>
<th>sd (as % CTR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brixton</td>
<td>13.9</td>
<td>450</td>
<td>1045</td>
<td>11%</td>
</tr>
<tr>
<td>Leicester</td>
<td>8.1</td>
<td>255</td>
<td>247</td>
<td>24%</td>
</tr>
</tbody>
</table>

**TABLE 1: SAMPLE RESULTS**

Using the formula for var(\( E \)) given above with \( T^2 = 0 \) and substituting the calculated values for \( E \), \( S^2 \) and \( V^2 \) it is possible to predict the accuracy attainable from using different sample sizes and measurement period durations. These results can be presented graphically as in Figures 1 and 2 which show the 'relative standard deviation' (the standard deviation expressed as a percentage of the mean) as a function of sample size and...
length of measurement period. These can then be used to formulate practical recommendations concerning sample sizes and measurement period durations. For example, both Figures 1 and 2 indicate that the additional improvement in estimation accuracy resulting from the measurement of additional weeks decreases rapidly beyond about 4 weeks. Combined with the possibility of temporal changes in CTR affecting measurements made over much longer periods and the increased difficulty of avoiding unrepresentative periods, such as public holidays, with extended measurement periods this suggests that, for residential samples at least, 4 weeks may represent the optimum length for measurement. This result is influenced to some extent by the apparent absence of any 'between weeks' contribution to the variance of the estimate in the cases studied but it is thought that it may well be confirmed when more extensive data becomes available.

It is clearly inappropriate to try and reach any general conclusions on sample size requirements from such sparse information. Encouragingly however, both sets of results suggest that estimates of tolerable precision (relative standard deviation less than 10%) can be achieved from samples of less than 500 connections, which is considered to be about the largest sample size which would be operationally feasible. Any conclusions beyond this however must await the availability of more extensive sample information.

ADEQUACY OF SCANNING RATE

As described above, the measurement device used in the BPO studies measures traffic quantities by the usual switch-count method using a 3 minute scan interval. Since the total traffic measured on a small sample of connections, particularly residential connections, may amount to less than one erlang, the adequacy of this scanning rate in a situation where average call holding times are of the order of 2 minutes has been questioned. This point has been examined theoretically and details are given in Appendix 2. The conclusion reached is that a 3 minute scan rate is adequate for traffic rate measurement purposes.

FURTHER WORK

As more study data becomes available it is intended to extend the work described above in a number of ways, in particular

- It is expected that traffic variations, as measured by the components of variance \( S^2 \), \( T^2 \) and \( V^2 \), will tend to increase with increasing CTR values. Once more exchange measurements have been completed, it is hoped to establish empirical relationships between the components of variance and the corresponding CTR values. These relationships would then be used in a more rigorous examination of sample size and measurement period requirements. Also, once 'live' information on relative costs of setting up a sample of given size and obtaining weekly measurements from it is available, it will be possible to determine the optimum arrangement for specified cost between measuring a small sample for a long period or a larger sample for a shorter period.

- As noted previously, the general availability of reliable information on total exchange originating traffic means that only one class of subscribers need to be sampled and measured to obtain CTR estimates for both classes. Some preliminary work has been done on defining the circumstances in which residential measurement is preferable to business and vice versa but a comprehensive treatment requires more information than is presently available. In particular, the empirical relationships referred to above may allow solution of the problem.

- Consideration is being given to limiting sample measurements on business subscribers to 'single-line' connections only. Multi-line business (private branch exchange) subscribers are served
by separate groups of exchange equipment and the total traffic originated by them can, in certain circumstances, be identified and measured separately. Where feasible, this would have the effect of making the business connections sampled more homogeneous and so increase the accuracy attainable from a sample of given size.

CONCLUSION

Preliminary results from the BPO study of CTR assessment using random samples of exchange connections are promising and suggest that the method may be suitable for general implementation. Much work remains to be done however.

REFERENCES

1 A Huitson "The Analysis of Variance: A Basic Course" (Griffin, 1971)

2 K M Olsson "Calculation of Dispersion in Telephone Traffic Recording Values for Pure Chance Traffic" (Tele, English Ed. No 2, 1959)

3 J Riordan "Telephone Traffic Time Averages" (BSTJ Vol 30, 1951)

ACKNOWLEDGEMENTS

The authors are indebted to Mr K Parry of the BPO/SEBD for his assistance with the analysis and for its development into a computer program (in PDL - details available from the authors on request).

Acknowledgement is made also to the Senior Director, Development, of the British Post Office for permission to publish this paper.
APPENDIX 1 DETAILED ANALYSIS

We assume that a random sample of connections of a particular class, size N, has been selected at an exchange where the total number of connections in the class is N. Measurement on the selected lines is made over a period of w weeks.

The EPO measurement device records values of the observed traffic rate, from the switch count method, for each sampled line in each week of the period. Let \( x_{ij} \) (i = 1, ..., w; j = 1, ..., n) represent the observed value for sample connection i during week j.

Analysis of the data is based on the following statistical model for the observations:

\[
x_{ij} = \lambda_i + w_j + \epsilon_{ij}
\]

(\text{Note: } i = 1, ..., N \text{ here})

where \( \lambda_i \) represents the mean CTR value associated with connection i, \( w_j \) is a random variable having zero mean and variance \( \sigma^2 \) representing (stochastic) week-to-week fluctuations common to all connections in the class; \( \epsilon_{ij} \) is a random variable with zero mean and variance \( \sigma^2 \) representing random week-to-week fluctuations specific to connection i, including those arising from switch-count estimation errors.

The quantity to be estimated is \( \bar{C} = \sum_{i=1}^{N} c_i / N \)

and the proposed estimate is the sample mean

\[
E = \frac{1}{Nw} \sum_{i=1}^{N} \sum_{j=1}^{w} x_{ij} / nw
\]

It is trivial that \( E \) provides an unbiased estimate for \( \bar{C} \) and it can also be shown (details available from authors on request) that

\[
\text{var}(E) = \frac{\bar{C}^2}{Nw} + \frac{\bar{C}^2}{nw} + \frac{(N-n) \bar{C}^2}{n}
\]

where \( \bar{C}^2 = \sum_{i=1}^{N} \sum_{j=1}^{w} x_{ij}^2 / nw \) and \( \bar{C}^2 = \sum_{i=1}^{N} \sum_{j=1}^{w} (c_i - \bar{C})^2 \)

Estimates of the quantities \( \bar{C}^2, \bar{C}_w^2 \) and \( \bar{C}^2 \) are required. These are obtained from the usual 2-way Analysis of Variance table derived from the readings \( x_{ij} \) (see, for example, reference 1 for explanation of Analysis of Variance tables).

Denote the mean squares calculated from this table (reference 1) as follows-

- **Connections mean square** = \( \frac{1}{w} \sum \left( \sum_{i=1}^{N} (x_{ij} - \bar{C})^2 - \text{var}(E) \right) / (N-1) \) by \( M_c \)
- **Weeks mean square** = \( \frac{1}{N} \sum \left( \sum_{i=1}^{N} (x_{ij} - \bar{C})^2 - \text{var}(E) \right) / (w-1) \) by \( M_w \)
- **Residual mean square** = \( \frac{1}{Nw} \sum \left( x_{ij} - (\bar{C} - \text{var}(E)) \right) / (Nw-1) \) by \( M_R \)

Then, after some algebra, it can be shown (details again available from authors) that, under the above model for the observations, the expected values of the 3 mean squares are given by

\[
E(M_c) = \bar{C}^2 + \sigma^2 / w
\]

\[
E(M_w) = \bar{C}^2 + \sigma^2 / N
\]

\[
E(M_R) = \bar{C}^2
\]

Hence unbiased estimates of \( \bar{C}^2 \), \( \bar{C}_w^2 \) and \( \bar{C}^2 \) are provided by the quantities \( (M_c - M_R) / w, (M_w - M_R) / N \) and \( M_R \) respectively.

Tests for significance of the quantities \( \bar{C}^2 \) and \( \bar{C}_w^2 \) can be made in the usual way by comparing the ratios \( M_c / M_R \) and \( M_w / M_R \) respectively to tables of the F-distribution with appropriate degrees of freedom (reference 1). Where these tests indicate \( \bar{C}_w^2 < 0 \) or \( \bar{C}_w^2 > 0 \), revised estimates of the variance components can be calculated in the usual way (reference 1). In particular, where \( \bar{C}_w^2 < 0 \) is indicated, the usual case, we use revised estimates for \( \bar{C}_w^2 \) and \( \bar{C}^2 \) given by

\[
\bar{C}_w^2 = \frac{M_w}{N}
\]

\[
\bar{C}^2 = \frac{M_c}{N}
\]

The appropriate estimates are then substituted in the formula for \( \text{var}(E) \) to give an estimate of the accuracy of the sample CTR value and to predict the accuracy attainable with different sample sizes and measurement periods.

APPENDIX 2 ADEQUACY OF SCANNING RATE

Reducing the scan interval used in the switch-count measurement procedure will result in a reduction in the "residual variance" \( \bar{C}^2 \). The following theoretical analysis demonstrates that even the greatest reduction possible, associated with changing from a 3 minute scan to continuous traffic measurement, would result in only a very small improvement in accuracy. A scanning rate of 3 minutes is therefore considered adequate.

Olsson (reference 2) gives an expression for the variance of traffic measured by scanning which is appropriate for our current purposes. This is that the variance \( \bar{V} \) is given by

\[
\bar{V} = \frac{A}{T} \left[ \frac{\bar{F} + 1}{f} - 2fn (1 - \exp (-T/h)) \right]
\]

where \( A \) is traffic intensity measured in erlangs

\( f \) is the scanning interval

\( h \) is average call holding time

\( T \) is the duration of the measurement period

\( n \) is the number of weeks

\( \bar{F} \) represents the quantity \( \exp (s/h) \)

The minimum value of this expression, corresponding to continuous traffic measurement, was previously given by Niordan (reference 3) as

\[
\bar{V}_C = \frac{2A}{T} \left[ \frac{f}{f - 1} - 1 + \exp (-T/h) \right]
\]

Both expressions are reduced by a factor \( \frac{T}{f} \) where measurement is over \( N \) periods of duration \( T \).

The maximum possible reduction in the "residual variance" \( \bar{V} \) from a reduction in the scan interval is given by

\[
\bar{V}_R = \frac{2A}{T} \left[ \frac{f}{f - 1} - 1 + \exp (-T/h) \right]
\]

Both expressions are reduced by a factor \( \frac{T}{f} \) where measurement is over \( N \) periods of duration \( T \).

\( \bar{V}_R \) is very small this gives us an approximate value for the maximum possible reduction in \( \bar{V} \) of

\[
\frac{c}{5} \left[ 50 \left( \frac{\exp (s/h) - 1}{\exp (s/h) - 1 - \exp (-T/h)} \right) - \frac{50}{1} \right]
\]

on substitution \( f = 2.5c \) (me) on evaluation.

A reduction of this order in \( \bar{V} \) is negligible for most purposes. For example, it would reduce the \( \bar{V} \) value for Brixton (Table 1) by just 3%, and that for Leicester by 8%. On measurement over 1 week this would improve the standard deviation of the estimate by only 1% of its present value in the case of Brixton and 2% for Leicester.