ABSTRACT

When a signalling system such as R2 allows a delay between the request for signalling equipment, and that equipment actually becoming available to handle a telephone call, it would appear reasonable to adopt a delay criterion for deciding the quantity of equipment to be provided. That this is not generally done is due to a number of reasons such as the absence of an agreed network performance specification in terms of delay, the difficulty of performing the calculations, and uncertainty whether it would be beneficial to design a network in this way. This paper proposes a number of performance criteria and examines the consequences of applying them in a certain hypothetical network, and against the background of conventional electromechanical switching equipment.

1. INTRODUCTION

The problem considered in this paper is that of finding the necessary rate of provision of signalling equipment in a conventional telephone network using a signalling system where it is permissible for there to be some delay between a seize signal arriving at an exchange, and the actual association of receiving equipment with the incoming line. Similarly, on the outgoing side of the exchange it is permissible for there to be some delay between the decision to seize a line in a particular outgoing route and the association of transmitting equipment. It has in the past often been the custom to design such systems with a provision of signalling equipment such as to make the probability of delay in getting signalling equipment very small - this has been the result of lack of information on the effect of delay on the working of the system. For it will be appreciated that the occurrence of delay in associating (say) receiving equipment with a junction which has been seized, not only contributes to possible post-dialling delay experienced by the subscriber, but also adds to the loading of the junction itself and any associated transmitting equipment.

The work here reported was undertaken to investigate the question whether the provision of such equipment on a delay basis in a typical network is advantageous or whether the delay that may be experienced by the subscriber and the additional traffic loads on control equipment reflected from one part of the network to another, would make such a method of design inadvisable. In order to make any progress with this study it is necessary to postulate a typical network, and a typical distribution of traffic in that network. We also need to set some sort of standard for the quality of service given to the subscriber in terms of post-dialling delay. In the next section the assumed structure of the network, and a number of quality of service criteria, are described.

2. NETWORK ASSUMPTIONS AND QUALITY OF SERVICE CRITERIA

The network considered is made up of just two types of telephone exchange:

1) pure transit exchanges. These are such that the holding times, which are constant (if calculated on the assumption that they are not affected by queuing delays elsewhere in the network), and calling rates of the traffic on the MF devices are the same for all transit exchanges.

i) pure non-transit exchanges, in each of which traffic originates and terminates. The following are assumed to have the same values for all non-transit exchanges in the network:

(I) the traffic which is local
(II) the originating traffic which is not local and passes through n transit exchanges (for n = 0, 1, 2, 3).
(III) the terminating traffic which is not local and has passed through n transit exchanges (for n = 0, 1, 2, 3).

For convenience the notation is used that: a direct junction path between two exchanges is one hop, that local calls are 0-hop calls, and that a call which is not local and passes through n transit exchanges (for n = 0, 1, 2, 3) is a (n + 1) hop call. For each non-transit exchange it is assumed that the proportion of originating calls which are 1-hop calls is $P(1)$ for $I = 0, 1, 2, 3, 4$, and that no calls in the network go through more than 3 transit exchanges. The values of the $P(I)$ are chosen so as to represent the situation occurring in practice.

The following three alternative performance criteria, to be met by the network, can be used:

A) for all non-local calls, the post-dialling delay should not exceed $t$ seconds in more than $P \times 100\%$ of cases.

or B) out of 1-hop to 4-hop calls, no one category should be such that the post-dialling delay exceeds $t$ seconds in more than $P \times 100\%$ of calls in that category.

or C) for 4-hop calls, the post-dialling delay should not exceed $t$ seconds in more than $P \times 100\%$ of cases.

Typically, criterion A) is used with $t = 4.0$ and $P = 0.01$.

The objective of the method is to find the lowest cost combination of MF equipment such that the network satisfies the chosen criterion.

3. STARTING TIMES AND DEVICE TRAFFIC

A calling subscriber is assumed to dial some routing digits (the number of these depending on the hop-length of the call), followed by a fixed number of numerical digits which identify the called subscriber. The time to dial one digit is taken as constant, and digits are assumed to be dialled in immediate succession. Suppose that the numerical dialled digits are transmitted by the MF equipment in the originating exchange and received by the MF equipment in the terminating exchange as soon as the MF signalling path has been established between these two exchanges. This MF signalling path can only start setting up after sufficient routing digits have been dialled to identify the exchange at the end of the first hop. However, in choosing the optimum time, after sufficient routing digits have been dialled, to commence setting up the MF connection between the originating and
Finding the optimum commencement time for each hop-length of non-local calls is a matter of applying the computer method described below to trial values.

Once the commencement times have been chosen, then for each MF device type in the network, the holding time can be calculated from the holding times at the MF devices. To investigate by how much the actual queueing times differed from the no-queueing holding times, a technique was used which was described by Bjorklund and Eldin (1). This method involves an iterative loop in which the means and variances of the device holding times are calculated from the parameters of the delay distributions of the queues, while the latter parameters are obtained from the means and variances of the device holding times. The iteration is stopped when the quantities in question are sufficiently close to equilibrium. This investigation showed that, considering other assumptions made in the calculation of traffic on the MF devices, and hence in the calculation of the parameters of the queuing processes at the MF devices, it is actually used to transmit the numerical digits. Transmission of the numerical digits has to wait until the dialling of these is sufficiently far advanced.

The holding times calculated on assumptions of no queueing delays have been shown to be sufficiently close to the actual pertaining queueing times to be used in the calculation of traffic on the MF devices, and hence in the calculation of the parameters of the queueing processes at the MF devices. To investigate by how much the actual queueing times differed from the no-queueing holding times, a technique was used which was described by Bjorklund and Eldin (1). This method involves an iterative loop in which the means and variances of the device holding times are calculated from the parameters of the delay distributions of the queues, while the latter parameters are obtained from the means and variances of the device holding times. The iteration is stopped when the quantities in question are sufficiently close to equilibrium. This investigation showed that, considering other assumptions made in the calculation of traffic on the MF devices, and hence in the calculation of the parameters of the queueing processes at the MF devices, it is actually used to transmit the numerical digits. Transmission of the numerical digits has to wait until the dialling of these is sufficiently far advanced.

The holding times calculated on assumptions of no queueing delays, have been shown to be sufficiently close to the actual holding times to be used in the calculation of traffic on the MF devices, and hence in the calculation of the parameters of the queuing processes at the MF devices. To investigate by how much the actual queueing times differed from the no-queueing holding times, a technique was used which was described by Bjorklund and Eldin (1). This method involves an iterative loop in which the means and variances of the device holding times are calculated from the parameters of the delay distributions of the queues, while the latter parameters are obtained from the means and variances of the device holding times. The iteration is stopped when the quantities in question are sufficiently close to equilibrium. This investigation showed that, considering other assumptions made in the calculation of traffic on the MF devices, and hence in the calculation of the parameters of the queueing processes at the MF devices, it is actually used to transmit the numerical digits. Transmission of the numerical digits has to wait until the dialling of these is sufficiently far advanced.

Because there are only two types of exchange in the network there is only a small number of types of MF devices which are to be optimally provisioned so that the network satisfies the chosen performance criterion (chosen from A), B) and C) above.

4. DETERMINING EQUIPMENT QUANTITIES

For any trial combination of MF equipment quantities, the computer program which has been developed can decide whether the chosen criterion is satisfied. This is done as follows. The no-queueing mean holding times of the MF devices and the known calling rates yield the device traffics. The means and variances of the holding times are then used with the method of Bjorklund and Eldin (1) (loc. cit. 1) page 28, to obtain weighted means of the queuing parameters for constant and for negative exponential holding times, the latter being found using Thierler's methods (2) (3). The computer method then uses standard theory for the convolution of negative exponential distributions to decide whether the given MF equipment quantities allows the exchange network to satisfy the chosen performance criterion. No attempt is made to allow for the possible dependence between successive queues.

In the particular study carried out at Plessey, it was necessary to find the optimum combination of 5 types of MF equipment, 2 of which were to be provided in small groups. These two types of equipment were the incoming registers at the transit exchanges and at the terminal exchanges, and they are in small groups because of the access arrangements provided on the particular switching system under consideration. It was found that it was only necessary to examine equipment combinations involving the minimum possible quantities of the devices in small groups. For these minimum combinations, exhaustive search was performed over the remaining three types of equipment, for combinations causing the network to satisfy the chosen criterion.

5. SIMPLE APPROXIMATION

The method just outlined is really only suitable for computer evaluation because of the lengthy calculations. However a less accurate, but a computationally simpler method has been developed, which is suitable for hand calculation.

This simpler method is based on the observation that, to a first order of approximation, the probability that a call is delayed at all, on serially passing through sets of MF devices, equals the sum of the separate probabilities of delay on passing once through each set of devices. For purposes of illustration attention is here confined to criterion A) with $t = 4$ and $P = 0.01$ (see above). The technique, however, extends easily to criteria B) and C).

In the notation used above, the proportion of calls originating in an exchange which are 1-hop calls, for $I=0,1,2,3,4$, is $P(I)$.

$$Q = h \times P(I), \quad I=1$$

then

$$P(I) = P(I)/Q, \text{ for } I=1,2,3,4,$$

is the proportion of non-local calls which is 1-hop. Using this notation

$$N = \frac{1}{I} \times P(I)(1-1)$$

is the number of transit calls passing through by the average non-local call. On the assumption of additivity of the separate probabilities of delay, therefore, the probability of delay of the average non-local call is the sum of the following three quantities:

a) the probability of delay at the MF equipment in the originating exchange

b) $M^e$ (the probability of delay of the queueing process at the MF equipment in the transit of exchanges)

c) the probability of delay at the MF equipment in the terminating exchange.

Using the method described above, for any particular system of hardware it would be possible to produce tables giving the traffic counts for all the devices under study if these traffic are such that, for the device quantities considered, criterion A) with $t = 4$ and $P = 0.01$, is just satisfied by the network. The value of $M^e$ and the device quantities is therefore as follows. Select from the table that configuration (known to satisfy criterion A) ) which corresponds most closely to the problem in hand, and compute its overall probability of delay. Adjust the equipment quantities for the problem in hand, so that the overall probability of delay, for the device quantities and traffic of the problem in hand, is less than or equal to the overall probability of delay for the tabulated configuration. This can be
done using standard tables for probability of delay. It should be noted that the criterion used, as well as the arrangement of routing digits and numerical digits, must be the same for the trial situation and for the situation used for calculating the tables. The value of $M$, (average number of transit exchanges passed through) can be different.

This value of $M$ for the trial situation is used in paragraph b) above, when calculating the probability of any delay at all in the trial situation. Provided the total probability of any delay at all when this value of $M$ is used is less than or equal to the corresponding probability for the tabulated configuration, the trial configuration should satisfy the criterion for which the tables were prepared.

Thus, for a chosen criterion, and for a given arrangement of routing digits and numerical digits, provided that one possible combination of device quantities and traffics, close to the trial combination, can be extracted from the tables, this method provides an approximate test of whether another combination of device quantities and traffics, possibly with a different value of $M$, satisfies the chosen criterion. This method has been found to work quite well for one particular system of hardware.

6. CONCLUSIONS

Where signalling systems work on a delay basis, as for example R2, there are several reasons why an administration may not wish to provide equipment in exchanges according to a delay criterion. Some of these reasons are:

1. Absence of a suitable specification of quality of service to the user in terms of delay.

2. Uncertainty concerning the effect of delays in accessing equipment on the holding time and hence traffic loading of other parts of the equipment.

3. Difficulty of performing the necessary calculations to dimension a network.

The work described here has shown that, under the stated conditions, which were chosen to be typical, a reasonable standard of service criterion can be met and delay provisioning can be used without there being serious extra loads reflected into the system. Also, provided the implicit assumption of independence between the delays at successive stages in the progress of a call turns out to be justified, the computing difficulty is not excessive. Further, an approximate method suitable for hand calculation has been found, which, for one particular system of hardware, facilitates design. The general method applies to a wide variety of hardware systems. Significant savings compared with provision on a loss basis may, depending on the particular hardware configuration in use, be obtained.

7. REFERENCES

