A QUEUING PROBLEM IN DATA TRANSMISSION

Eckart Wollner
German Post Office
Darmstadt, Federal Republic of Germany

ABSTRACT

This paper investigates a queuing system with non-preemptive priorities and several classes of finite numbers of subscribers with different birth and service rates who need different numbers of time slots per clock for data transmission in a time division multiplex system. An analytical approach is made to determine the state probabilities. The traffic characteristics are evaluated numerically for a particular example and compared with the corresponding values obtained in the simulation of the problem.

1. INTRODUCTION

An electronic data switching network allows data transmission by time division multiplex from a concentrator via exchanges to another concentrator. Discrete lines are only required between the subscriber and the concentrator. This paper deals with the path between the concentrator and the exchange.

The data transmission is effected via a synchronous time division multiplex system in the following way: The time address, the subscriber's address, and the direction of the polarity reversals are transmitted in a time division multiplex channel which is not being held at the moment. Throughout the duration of the transmission the data supplied by one subscriber may occupy completely different channels of the time division multiplex system [1]. We do not intend to discuss in this paper the amount of transmission capacity saved because only the polarity reversals are transmitted. The time division multiplex channels are regarded in this context as service stations for data transmission and are considered as being held throughout the duration of a call.

Several classes of finite numbers of subscribers are connected to the concentrator who want to transmit data and have, for this purpose, common access to a TDM-system between the concentrator and the exchange. Subscribers of a certain class transmit their data at the same rate, different classes may use different rates. Whenever a call of a subscriber belonging to a certain class seize more channels per clock than another call offered by a subscriber of another class, clearly a larger amount of information is transmitted. Thus, a certain transmission rate is generated by simultaneously seizing a certain number of channels per clock. A subscriber is in one of the following 3 states: he is either idle, or waiting for the possibility to transmit data, or he is transmitting data. The times during which a subscriber of a certain class is idle are random variables with an identical distribution which may be different for different classes. They are independent of the idle times of other subscribers and of the holding times. Also the holding times of the calls of each class are independent identically distributed random variables. If the TDM-system does not provide, at a certain moment, a sufficient number of channels required for a subscriber's call attempt, the call attempt may wait in an unlimited waiting room. The waiting discipline for subscribers belonging to one class is FIFO, between subscribers of different classes there is a non-preemptive priority.

This model is also used in other fields than data transmission: for instance, in the maintenance of a stock of engines. In this case, only one service-man is required for one simple repair whereas several service-men have to work simultaneously in order to make one difficult repair.

344/1

ITC 7
2. DESCRIPTION OF THE MODEL

The symbol \( r \) indicates the number of subscriber classes connected to the concentrator. Each class \( i \) consists of \( q_i \) subscribers, the total number of subscribers connected to the concentrator being

\[
\sum_{i=1}^{r} q_i = N < \infty
\]

The waiting room is sufficiently dimensioned with \( N \) waiting places. A subscriber of class \( i \) transmits his data at the transmission rate \( v_i \), \( v_1 \leq v_2 \leq \cdots \leq v_r \).

The transmission of data via the TDMA-system requires channels to be held per clock \( T \). The duration selected for

\[
T = \frac{1}{v_1}
\]

ensures that a subscriber of class 1 needs exactly one channel per clock to transmit his data. In general, \( v_i \cdot T \) channels are held by a set of data transmitted by a subscriber of class \( i \).

Each clock \( T \) is assumed to be divided into \( n \) channels. If, at a certain moment, less than \( v_i \cdot T \) channels are available for a new call-offered, for example, by a subscriber of class \( i \) - the new call attempt should be allowed to wait for service in an unlimited queue.

Voluntary departures are not possible. The service discipline for each class \( i \) is "first come, first served". There are, however, priorities among the individual classes, but they do not preempt existing connections. A higher priority is given to subscribers belonging to a class with a higher index (higher transmission rate). This ensures that a new call attempt of a subscriber belonging to a class with a higher index, i.e., with a greater channel demand, prevents other subscribers from initiating their calls until after the completion of a certain number of calls enough channels are available for the higher index subscriber to transmit his information.

The following problem has to be solved:

1. \( r \) sources of finite size with different birth and service rates
2. \( n \) parallel service stations
3. a queuing system with non-preemptive priorities
4. simultaneous use of several service stations

Our following considerations shall be based on the assumption that \( r = 2 \).

3. DESCRIPTION OF THE STATES AND TRANSITIONS

Class 1 is assumed to consist of \( q_1 \) subscribers requiring 1 channel each; class 2 consists of \( q_2 \) subscribers requiring 4 channels, i.e., the transmission rate is four times as high. Class 2 has non-preemptive priority over class 1, the waiting discipline within the same classes is FIFO.

In order to describe the state of the system, allowance has to be made for the number of waiting and seizing subscribers of each class.

Denotations: The system is in the state

\[
\mathbf{z} = \{j_1, j_1; k_2, k_2\}
\]

where \( j_1 \) and \( k_2 \) are the numbers of class 1 and class 2, respectively, waiting or seizing subscribers of the system.

According to this description, the system is in one of the states \( \mathbf{z} \) at any time \( t \); i.e., the number of all possible states of the state space \( S = \{\mathbf{z}\} \) is finite.

The calling rate per class, which depends on the state of the system, is assumed to be as follows:

\[
\lambda_j^{(1)} = \begin{cases}
\frac{q_1-j}{a_1} & \text{for } j = 0, \ldots, q_1 \\
0 & \text{otherwise}
\end{cases}
\]

\[
\lambda_j^{(2)} = \begin{cases}
\frac{q_2-k}{a_2} & \text{for } k = 0, \ldots, q_2 \\
0 & \text{otherwise}
\end{cases}
\]

The superior index \( (i) \) denotes the class, \( \frac{1}{a_i} \) is the calling rate per subscriber per unit of idle time \( i = 1, 2 \). Accordingly, also the service rate depends on the state of the system. We assume:

\[
\mu_j^{(1)} = \begin{cases}
j_1/b_1 & \text{for } j_1 = 1, \ldots, n \\
0 & \text{otherwise}
\end{cases}
\]

\[
\mu_k^{(2)} = \begin{cases}
k_2/b_2 & \text{for } k_2 = 1, \ldots, \left[\frac{N}{4}\right] \\
0 & \text{otherwise}
\end{cases}
\]

where \( b_i \) is the mean holding time per subscriber of class \( i \), \( i = 1, 2 \).
The state \( Z \) of the system is assumed to be \( \{1, j1, k2, k2\} \) at the time \( t \). For convenience we shall use the term \( 1 \)-call for call attempts of the class \( i, i = 1, 2 \). The transition probabilities from the state \( Z \) to a new state during a small interval \( (t, t+h) \) shall be defined as follows: Formula (1):

\[
\begin{align*}
\alpha) & \ Pr \left\{ \text{1-call arrives during } (t, t+h) \right\} = \\
\beta) & \ Pr \left\{ \text{2-call arrives during } (t, t+h) \right\} = \\
\gamma) & \ Pr \left\{ \text{1-call ends during } (t, t+h) \right\} = \\
\delta) & \ Pr \left\{ \text{2-call ends during } (t, t+h) \right\} = \\
\epsilon) & \ Pr \left\{ \text{several events during } (t, t+h) \right\} = \\
\zeta) & \ Pr \left\{ \text{no event during } (t, t+h) \right\} = \\
\end{align*}
\]

where \( h \) is an arbitrarily small given time. By \( \alpha \) - \( \zeta \) the original state \( Z \) is changed in the following way:

\[
\begin{align*}
\alpha) & \quad j_j \rightarrow j+1, \quad k_2, k_2 \text{ remain unchanged} \\
\beta) & \quad k_j \rightarrow k+1, \quad j_1, j_1 \text{ remain unchanged} \\
\gamma) & \quad j_1 \rightarrow j-1, \quad k \text{ remains unchanged} \\
\delta) & \quad k \rightarrow k-1, \quad j \text{ remains unchanged} \\
\end{align*}
\]

It should be noted that in the case of transitions \( \gamma \) and \( \delta \) changes may occur in more than 2 components of the original state vector, e.g.: 

\[
\begin{align*}
\psi > 0, \{j_j, j_1, k_2, k_2\} \rightarrow \psi > 0, \{j_j, j_1, k_2, k_2\} \\
\end{align*}
\]

In other words, we assume that the idle times of one subscriber of one class are negative exponentially distributed and independent of the state of the system. The same assumption is made with respect to the holding time of one subscriber.

4. KOLMOGOROV FORWARD EQUATIONS AND THEIR ERGODICITY

Let \( F(Z, t) \) be the probability that we have the state \( Z \) at time \( t \) and \( F(Z, t) = F(Z, t)^{\infty} \), the conditional probability of state \( Z \) at the time \( t \leq t < t^0 = 0 \) if we have the state \( Z \) at the time \( t \). This assumption yields the following equation:

\[
F(Z, t) = \sum_{Y \in S} F(Y, t) F(Z, t)^{\infty} F(Y, t) = \sum_{Y \in S} F(Y, t)^{\infty} F(Z, t)^{\infty} F(Y, t) F(Z, t)^{\infty},
\]

where \( 0 \leq t < t^0 \), because in conformity with our assumptions the system discussed in this paper is described by a Markov chain \( X \) with a continuous time parameter \( t \) and a finite state space \( S \). X has stationary transition probabilities, i.e. \( F(Z, s+t|Z, s) \) is independent of \( S \) for all \( Z, X \in S \) and \( t \in (0, \infty) \). The infinitesimal generator of the Markov chain has been defined in the last section of (1). If the system is adequately dimensioned, the Markov chain may be assumed to be irreducible with a finite number of positive recurrent states. Then it follows, according to [2], that there exists uniquely a limiting distribution of the transition probabilities and that it coincides with the limiting distribution of the state probabilities

\[
\lim_{t \to \infty} F(Z, t) = p(Z)
\]

This enables us to find \( p(Z) \) with the aid of the equation system

\[
\sum_{Y \in S} p(Y) \cdot \Lambda(Y) = 0,
\]

where \( \Lambda(Y) \) is a sum of the infinitesimal generators of the Markov chain \( X \). In this way the following equation system is obtained:

\[
\text{(see next page)}
\]

5. NUMERICAL EVALUATION

\( q_1 = 9 \) subscribers of class 1, \( q_2 = 2 \) subscribers of class 2, and \( n = 6 \) channels yield as many as 91 possible states \( Z \) of the system. With the aid of the above equation system (2) the probabilities of state \( p(Z) \) were calculated for particular values of \( b_1 \) and \( 1/a_1 \). Some traffic theoretical quantities may easily be determined from the probabilities of state \( p(Z) \).

Def. The mean number of waiting calls is

\[
A_{w1} := \sum_{j_1 > 0} p(j_1, j_1; k_2, k_2) \text{ for class } 1
\]

\[
A_{w2} := \sum_{k_2 > 0} p(j_1, j_1; k_2, k_2) \text{ for class } 2
\]

The traffic offered = the mean number of calls arriving during the mean holding time is

\[
A_1 := b_1 \cdot \sum_j p(j_1, j_1; k_2, k_2) \text{ for class } 1
\]

\[
A_2 := b_2 \cdot \sum_j p(j_1, j_1; k_2, k_2) \text{ for class } 2
\]

The mean calling rate is

\[
A_1 := \frac{A_{w1}}{b_1}, \quad A_2 := \frac{A_{w2}}{b_2}
\]

The probability of delay is

\[
P_1(>0) := \frac{1}{A_1} \cdot \sum_j p(j_1, j_1; k_2, k_2) \text{ for class } 1
\]

\[
P_2(>0) := \frac{1}{A_2} \cdot \sum_j p(j_1, j_1; k_2, k_2) \text{ for class } 2
\]

where we sum over the states in which a state of delay is caused by the arrival of a new \( 1 \)-call, \( i = 1, 2 \).
Formula (2):

\[(\lambda_1^e, \lambda_2^e)_{j_1, j_2} \in \mathbb{R}^2 \]

\[p(j_{i1}, j_{i2}; k_{i1}, k_{i2} | \nu) = \lambda_{i1}^{j_{i1}} \lambda_{i2}^{j_{i2}} p(j_{i1}, j_{i2}; k_{i1}, k_{i2} | \nu) + \lambda_{i1}^{j_{i1}-1} \lambda_{i2}^{j_{i2}} p(j_{i1}, j_{i2}; k_{i1}, k_{i2} | \nu)
\]

for \( n \leq i_1 \), \( j_{i1} > 0 \), \( k_{i1} > 0 \)

\[p(j_{i1}, j_{i2}; k_{i1}, k_{i2} | \nu) = \sum_{k=0}^{\infty} p(j_{i1}, j_{i2}; k_{i1}, k_{i2} | \nu)
\]

All components of \( p(j_{i1}, j_{i2}; k_{i1}, k_{i2} | \nu) \) are not negative. If one of them turns negative, \( p(j_{i1}, j_{i2}; k_{i1}, k_{i2} | \nu) \) is to be omitted.
The mean waiting time is

\[ t_{W1} = \frac{A_{W1}}{X_1}, \quad t_{W2} = \frac{A_{W2}}{X_2}. \]

The mean delay of calls delayed is

\[ t_{D1} = \frac{t_{W1}}{P_1(>0)}, \quad t_{D2} = \frac{t_{W2}}{P_2(>0)}. \]

For the numerical evaluation a mean holding time of \( b_1 = b_2 = 1 \) is selected. The mean idle time

\[ \frac{1}{\lambda_1} = \frac{1}{\lambda_2} = \frac{1}{\lambda} \]

varies from 0.04 (0.01) 0.16. Figures 1 to 4 specify the traffic offered \( A \), the probability of delay \( P(>0) \), and the mean delay \( t_W \) and \( t_D \) as functions of the mean idle time \( 1/\lambda \) of both classes.

Fig. 1: The traffic offered \( A \) as a function of the mean idle time \( 1/\lambda \) of both classes.

Fig. 2: The probability of delay \( P(>0) \) as a function of the mean idle time \( 1/\lambda \) of both classes.

Fig. 3: The mean waiting time \( t_W \) as a function of the mean idle time \( 1/\lambda \) of both classes.

Fig. 4: The mean delay of calls delayed \( t_D \) as a function of mean idle time \( 1/\lambda \) of both classes.

The same input parameters were used for a "time-true" simulation [3], in which first the interval between the present moment and the last call attempt is selected as a negative exponentially distributed random variable. Then, a class is selected according to the traffic offered and a source of this class. If it is already included in the system, a new interval, a new class, and a new source are selected. If the source is not yet included in the system, a negative exponentially distributed holding time is determined for it. After selecting the inter-
val, the system is up-dated in each case, i.e. it is ascertained whether calls have been completed since the last up-dating, which perhaps enabled waiting subscribers to initiate their calls.

For each parameter constellation 500 000 experiments were made and the 95 % confidence intervals for the characteristic traffic theoretical quantities were determined. The results show that all numerically determined values of the analytical model described in this paper were within the 95 % confidence intervals of the simulated values. In order to demonstrate the precision, which we were able to represent graphically only in Fig. 4, we specify in the following table some other values for comparison:

<table>
<thead>
<tr>
<th>1/a</th>
<th>( A_1 )</th>
<th>95 % conf. limits</th>
<th>( P_1(\cdot 0) )</th>
<th>95 % conf. limits</th>
<th>( \mu_1 )</th>
<th>95 % conf. limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.04</td>
<td>0.3461</td>
<td>0.3434</td>
<td>0.3462</td>
<td>0.0057 0.0054 0.0060</td>
<td>0.0040 0.0056 0.0044</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.0768</td>
<td>0.0759</td>
<td>0.0772</td>
<td>0.0424 0.0405 0.0429</td>
<td>0.0399 0.0382 0.0421</td>
<td></td>
</tr>
<tr>
<td>0.08</td>
<td>0.6658</td>
<td>0.6625</td>
<td>0.6677</td>
<td>0.0275 0.0269 0.0283</td>
<td>0.0184 0.0175 0.0193</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.1472</td>
<td>0.1453</td>
<td>0.1479</td>
<td>0.0275 0.0975 0.1002</td>
<td>0.0833 0.0895 0.0858</td>
<td></td>
</tr>
<tr>
<td>0.12</td>
<td>0.9596</td>
<td>0.9557</td>
<td>0.9642</td>
<td>0.0661 0.0657 0.0683</td>
<td>0.0451 0.0438 0.0472</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.2113</td>
<td>0.2104</td>
<td>0.2140</td>
<td>0.0661 0.1673 0.1687</td>
<td>0.1325 0.1289 0.1354</td>
<td></td>
</tr>
<tr>
<td>0.16</td>
<td>1.2271</td>
<td>1.2204</td>
<td>1.2298</td>
<td>0.1176 0.1154 0.1188</td>
<td>0.0846 0.0812 0.0862</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.2689</td>
<td>0.2657</td>
<td>0.2705</td>
<td>0.2476 0.2410 0.2481</td>
<td>0.1877 0.1812 0.1897</td>
<td></td>
</tr>
</tbody>
</table>

6. CONCLUSIONS

This paper intends to present an approach to a queuing model which so far has not been investigated and may be applied in many operations research problems. Although general solutions of the problem have not yet been found, it was possible to use the approach to determining the state probabilities for demonstrating the quality of the simulation by means of a simple example. In the case of larger numbers of subscribers and channels the number of states rises so quickly that simulation procedures have to be used because the computers available at present are not able to solve such large equation systems.

7. REFERENCES

