TRAFFIC MODELS FOR RESERVATION SYSTEMS

Peder J. Emstad and Boning Feng
The Norwegian Institute of Technology, Trondheim, Norway

Two discrete time models for placing reservations within a finite window of size $W$ are defined and studied. The system has $M$ channels. Call duration is exactly one time slot. The distribution of the number of requested reservations per time slot is general and so is the distribution of the notice interval. If all the channels in the slot requested by a customer are already reserved, he will either leave the system (LOSS model) or accept the next available slot (TNASQ model).

The LOSS model is solved exactly while the TNASQ model is solved based on some independence assumptions which are supported by simulations. The main measures studied are carried traffic, loss probability, delay probability and expected waiting time. The results from LOSS model and two of the approximate methods in TNASQ model show good accuracy.

The computational complexity of our algorithms are below $O(MW)$, which means that they can be used to solve systems with large windows and many channels.

1 Introduction
In future telecommunication systems we shall have video conferencing as a new service. This service requires high transmission capacity. With several people involved in a conference the only feasible solution is to allow advance reservation for system transmission capacity. Similar arguments also apply for point-to-point high capacity transmission channels. The increasing need for using reservation demands more attention to reservation system modeling. Some pioneering work was done by J. Roberts [1] published at ITC 10, followed up with a paper coauthored with K. Liao [2] at ITC 11 and several papers presented at ITC 12 [3, 4, 5].

The new dimension in a reservation system model is the notice interval, the time from when a reservation is placed until service commences. In our work the general problem formulation is presented. We apply discrete time models which comprise both loss and delay systems, and present some analytical solutions supported by simulations.

2 General Problem Formulation
As in any classical traffic model we have a customer arrival process and a service time distribution. In addition we need the distribution of the notice interval. Roberts [1] pointed out that it is dangerous to integrate services having different distributions of this notice interval since service is necessarily in order of requested interval. We therefore just study reservation models where all traffic has the same distribution of the notice interval.

The following notations are applied:
- $f$ - time that a customer arrives and makes a reservation
- $T$ - service starting time

1. $D$ - service duration
2. $\tau$ - notice interval ($\tau = T - f$)
3. $\lambda$ - arrival rate

Fig. 1 shows a time diagram for a reservation system and clarifies the notation.

![Time diagram for a reservation system](image)

**Figure 1.** Time diagram for a reservation system

Discrete time reservation models
We have emphasized the problems of describing the state of the system at a given time with respect to accepted reservations. The time on the channel is discretized to slots. We introduce the notion of a window. The window contains the future time slots within which a reservation is allowed to be placed, see Fig. 2.

![Discrete time diagram with window](image)

**Figure 2.** Discrete time diagram with window

The window size is $W$ time slots. We define the present slot as in position 0 while the newest slot (which...
represents the maximum allowed notice interval) in the window is in position $W$. Each time a new slot has been shifted into the window, the position numbers of the slots in the window will be decreased by one. The state of the system is the state of reservations within the window (possibly combined with the state of the reservation queue in delay systems, see below). The system has $M$ channels which equals the number of reservations accommodated for each time slot. The time unit is defined as the duration of one time slot. We assume that the service duration is 1 time slot for all reservations ($D = 1$).

We proceed with the following notation:

$$ a_i = P[i \text{ new customers arrive during a slot}] $$

$$ A(z) = \sum_{i=0}^{w} a_i \cdot z^i $$

$$ r_w = P[a \text{ customer requests position } w] \quad (w = 1, 2, ..., W, \sum_{w=1}^{W} r_w = 1) $$

$$ N_w = \text{number of requests to position } w \text{ made during one time slot} $$

$$ b_w^{(i)} = P[N_w = j], \quad B_w(z) = \sum_{j=0}^{w} b_w^{(j)} \cdot z^j $$

$$ I_i = \text{accumulated number of reservations to a slot when it goes to service} $$

so that

$$ I_i = \sum_{w=1}^{w} N_w $$

$$ C_1(z) = E(z^i) $$

The arrival process

Let us find the relation between the reservation arrival process and the accumulated number of reservations to a particular slot. We have

$$ E(z^{N_w} / I = i \text{ arrivals to the system}) = (1 - r_w + z \cdot r_w)^i $$

$$ E(z^{N_w}) = \sum_{i=0}^{w} (1 - r_w + z \cdot r_w)^i \cdot a_i = A(1 - r_w + z \cdot r_w) $$

Then Eqs. (7) and (8) give

$$ C_1(z) = E(z^i) = E(\sum_{w=1}^{w} N_w) $$

When the $N_w$'s are independent we get

$$ C_1(z) = \prod_{w=1}^{w} E(z^{N_w}) = \prod_{w=1}^{w} A(1 - r_w + z \cdot r_w) $$

If the reservation arrival process is Poisson then $A(z) = e^{-\lambda + \beta}$, and it follows from Eq. (10) that

$$ C_1(z) = \prod_{k=1}^{W} e^{-\lambda + \beta (1 - \gamma + \rho k)} = \prod_{k=1}^{W} e^{-\lambda + 2 \beta \gamma} = e^{-\lambda + 2 \rho} $$

The accumulated number of reservation arrivals is again Poisson distributed with the same parameter, regardless of the type of reservation process! This is the discrete version of the theorem stating that the output from the $M / G_\text{traffic}/S$-system is a Poisson process [6].

However, this is not true for any reservation arrival process. If there is for instance a probability $p$ for one arrival, and $1-p$ for no arrival in a time slot, then $A(z) = 1 - p + pz$. Obviously $C_1(z)$ will be different since the accumulated number of reservations may be larger than one.

If a user can not find any free channels in the requested slot, he will either abandon his call or negotiate an alternative reservation. The following two cases will be analyzed:

- Reservations for already occupied slots are lost (LOSS model)
- Take Next Available Slot with Queue (TNASQ Model)

3 Stationary Analysis of the Various Models

3.1 LOSS Model

In this model reservations for fully reserved slots are cleared. Let

$$ \pi_w^{(m)} = P[m \text{ channels have been reserved when a slot leaves position } w, 0 \leq m \leq M] $$

and

$$ L(w) = P(a \text{ reservation to position } w \text{ is lost}) $$

Following a time slot moving from position $W$ to position $I$ we easily find

$$ \pi_w^{(m)} = \begin{cases} \sum_{i=0}^{w} \pi_w^{(i)} \cdot b_w^{(m-i)} & m = 0, 1, ..., M - 1 \\ 1 - \sum_{i=0}^{w} \pi_w^{(i)} & m = M \end{cases} $$

Then $L(w)$ can be expressed as the fraction of the reservations which are lost during one time slot:

$$ L(w) = \frac{E(\# \text{lost reservations per time slot})}{E(\text{total reservations per time slot})} $$

So we get

$$ L(w) = \frac{\sum_{i=0}^{M} \sum_{j=0}^{M-i} (i - M + j) \cdot b_w^{(i)}}{\sum_{i=0}^{M} i \cdot b_w^{(i)}} $$

The expected total number of accommodated requests per slot can be found by

$$ \rho = \sum_{m=1}^{M} m \cdot \pi_m^{(m)} $$

The total loss is $\lambda - \rho$ and the overall loss probability is

$$ L = 1 - p/\lambda $$
3.2 TNASQ Model

If a customer requests a slot where all the channels are reserved, he is supposed to accept the first subsequent available slot within the window. If there are no such slots in the window, the customer will be queued. Up to M of the queued customers (the first ones if the queue follows FIFO discipline) will enter position W in the window when a new slot is shifted into it. The state of the system is now a combination of the number in queue and the state of reservations within the window.

For this model we have to take overflow into consideration. The next state of a slot in position w is no longer only dependent on what it experienced in the positions before, but due to overflow also on the state of positions 1, 2, ..., w-1 as well.

A technique with historic variables is used. We follow a general slot from it enters the window (possibly the queue) until it goes to service. The life of a slot at one position is now assumed independent of what it experiences in the next. This approach violates dependence relations in the window.

Suppose the slot we follow is now in the wth position in the window (Fig. 3).

\[ I_w = \text{accumulated number of direct reservations to the slot in position } w, \quad (I_w = \sum_{w=1}^{w-1} N_j) \] (16)

\[ J_w = \text{accumulated number of overflowing reservations received by position } w \text{ from position } w-1 \] (17)

\[ K_w = \text{accumulated number of overflowing reservations from position } w \text{ to position } w+1 \] (18)

and the state probabilities with their generating functions,

\[ e_w^{(0)} = \Pr \{ I_w = i \}, \quad C_w(z) = E(z^{I_w}) = \sum_{i=0}^{\infty} e_w^{(0)} \cdot z^i \] (19)

\[ p_w^{(0)} = \Pr \{ J_w = j \}, \quad P_w(z) = E(z^{J_w}) = \sum_{j=0}^{\infty} p_w^{(0)} \cdot z^j \] (20)

\[ q_w^{(0)} = \Pr \{ K_w = k \}, \quad Q_w(z) = E(z^{K_w}) = \sum_{k=0}^{\infty} q_w^{(0)} \cdot z^k \] (21)

\[ D_w(z) = P_w(z) \cdot C_w(z) = \sum_{i=0}^{\infty} d_w^{(0)} \cdot z^i \] (22)

3.2.1 Traffic model

For any slot the following relations are valid:

\[ K_w = \max(I_w + J_w - M, 0), \quad w = 1, 2, \ldots, W \] (23)

\[ J_1 = J_2 \] (24)

Applying the assumed independence relation between various slots, we also have

\[ J_w = K_{w-1}, \quad w = 2, 3, \ldots, W+1 \] (25)

A recurrence equation for \( P_w(z) \) can be now found by Equations (23), (24) and (25):

\[ Q_w(z) = E(z^{K_w}) = E(z^{\max(I_w + J_w - M, 0)}) \]

and since \( P_{w+1}(z) = Q_w(z) \),

\[ P_{w+1}(z) = z^0 \cdot \Pr \{ I_w + J_w - M < 0 \} + \sum_{i=0}^{\infty} z^i \cdot \Pr \{ I_w + J_w = i + M \} \]

Using Eq. (22), the equation above can be written as

\[ P_{w+1}(z) = \sum_{i=0}^{\infty} d_{w}^{(0)} z^i + \sum_{i=0}^{\infty} d_{w+1}^{(0)} z^i = z^M \left[ P_w(z) \cdot C_w(z) - \sum_{i=0}^{\infty} d_{w}^{(0)} z^i \right] + \sum_{i=0}^{\infty} d_{w+1}^{(0)} z^i \] (26)

We have \( P_2(z) = P_1(z) \) since \( J_2 = J_1 \). With Eq. (26) it follows that

\[ P_1(z) = \frac{\sum_{i=0}^{\infty} [d_1^{(0)} z^i - d_2^{(0)} z^i]}{z^M - C_1(z)} \] (27)

The numerator has totally \( M \) unknowns, \( d_1^{(0)}, d_2^{(0)}, \ldots, d_{M-1}^{(0)} \). When they are found \( P_1(z) \) is known, and then all \( P_w(z) \), \( w = 2, 3, \ldots, W, W+1 \) (where \( P_{w+1}(z) \) describes the traffic overflowing to the queue) are known through Eq. (26).

When \( M=1 \), Eq. (27) reduces to

\[ P_1(z) = [d_1^{(0)} z - d_2^{(0)}] / [z - C_1(z)] \]

where the only unknown \( d_1^{(0)} \) can be found from the normalizing condition \( P_1(1) = 1 \), which gives \( d_1^{(0)} = 1 - \lambda \).

When \( M>1 \), \( M-1 \) additional equations must be obtained in order to solve the \( M \) unknowns in Eq. (27). It may be shown (Rouché's theorem, see [7]) that whenever \( \lambda < M \) then the denominator \( z^M - C_1(z) \) in Eq. (27) has exactly \( M \) zeros on and within the unit disk of the complex plane. Since \( P_1(z) \) is analytic on and within the unit disk these \( M \) zeros in the denominator must be balanced by zeros in the numerator. The numerator is a polynomial of degree \( M \) and may be written \( K \cdot \prod_{i=1}^{M} (z - z_i) \), or
The roots $z_i, i = 0, 1, \ldots, M - 1$, are as explained equal to the roots of the denominator. One of the zeros is $z_0 = 1$ for which we know $P_1(1) = 1$. This root helps us to find the limit of Eq. (28) when $z \to 1$:

$$K = \frac{M - \lambda}{\prod_{i=1}^{M-1} (1 - z_i)}$$

(28)

The probability $\pi_w^i$ that $i$ reservations have been made to a slot when it leaves position $w$, is given by

$$\pi_w^i = \begin{cases} 
  d_w^{(i)}, & 0 \leq i \leq M - 1 \\
  1 - \sum_{m=0}^{M-1} d_w^{(m)}, & i = M 
\end{cases}$$

(32)

### 3.2.2 Delay Probability

If the accumulated number of requests for a slot exceeds $M$, then some of the reservations must be delayed. From the viewpoint of a customer who makes a request for position $w$, the delay probability is dependent on the priority between the reservations initially requested for position $w$ ("right-in reservations"), and the reservations delayed from earlier positions ("overflowing reservations"). We discuss three cases where the priority is different:

**Case I: priority to overflowing reservations**

Suppose $m$ channels of a slot are reserved when it moves to position $w$. For an arbitrary right-in reservation to the position not to get delayed, the following conditions must be satisfied:

- The number $i$ of overflowing reservations to the slot must be less then $M-m$.
- The right-in reservation must get one of the remaining $M-m-i$ channels.

Let $H_w$ be the number of overflowing reservations to slot $w$, $s_w^{(0)} = Pr[H_w = h]$ and $S_w(z) = \sum_{i=0}^{\infty} s_w^{(i)} \cdot z^i$. We also define

$$R_w(j) = Pr[a \text{ right-in reservation is not delayed/it requests pos. } w \text{ and there are } j \text{ channels available}]$$

Then

$$\Omega_{h,w} = \sum_{m=0}^{M-1} \pi_w^m \cdot \sum_{i=0}^{M-m-1} s_w^{(i)} \cdot R_w(M-m-i)$$

(35)

$s_w^{(0)}$ and $R_w(j)$ are independent of each other and can be calculated separately.

First we establish the relationship between $s_w^{(0)}$ and the results found in the previous section.

$$J_w = H_w + J_{w+1}$$

(36)

We thus have

$$P_w(z) = S_w(z) \cdot P_{w+1}(z)$$

(37)

So if $P_w(z)$, $w = 1, \ldots, W + 1$, are known, the numerical values of $s_w^{(0)}$ are determined.

If $m$ and $i$ are given, then the conditional probability that a right-in reservation gets one of the remaining $M-m-i$ channels equals $R_w(M-m-i)$, which is defined above (34). Now $R_w(j)$ can be expressed as the fraction of right-in reservations which does not overflow. We thus have

$$R_w(M-m-i) = \frac{\sum_{j=1}^{M-m-i} j \cdot b_w^{(j)} + (M-m-i) \cdot \sum_{j=M-m-i}^{\infty} b_w^{(j)}}{\lambda \cdot r_w}$$

(38)

For a general arrival process, applying Eqs. (37) and (38) in Eq. (35) finally gives $\Omega_{h,w}$. 

**Case II: random choice**

**Case III: priority to right-in reservations**

We define

$$\Omega_{h,w} = Pr[a \text{ reservation is not delayed/it requests position } w; \text{ Case } ij, i = I, II, III]$$

(33)
Case II: random choice

Suppose \( m \) channels of a slot are reserved when it moves to position \( w \). For an arbitrary right-in reservation to slot \( w \) not to get delayed, it must get one of the \( M-m \) free channels (in competition with overflowing and the other right-in reservations).

Assuming that the right-in reservations experience the same probability for delay as the overflowing reservations, we define

\[
R^*(j) = \text{Pr(\text{an arbitrary reservation to slot } w \text{ does not get delayed})}
\]

Let \( V_w \) be the total number of reservations,

\[
g_w^0 = \text{Pr}\{V_w = i\} \quad \text{and} \quad G_w(z) = \sum_{i=0}^{\infty} g_w^0 \cdot z^i.
\]

With similar arguments as in Case I, \( G_w(z) \) can be found by the relation

\[
G_w(z) = \sum_{i=1}^M z^i D_{w+1}(z).
\]

Numerical values can also be calculated by similar methods. Thus

\[
R^*(j) = \frac{\sum_{i=1}^j i \cdot g_w^0 + \sum_{i=0}^{\infty} g_w^0}{\sum_{i=1}^j i \cdot g_w^0}
\]

\( \Omega_{\text{II},w} \) can be found by using Eq. (40):

\[
\Omega_{\text{II},w} = \sum_{m=0}^{M-1} n_w^{(m)} \cdot R^*(M-m)
\]

Case III: priority to right-in reservations

In this case the computation is easier than in Case I because we can neglect the influence of the overflowing reservations. The probability for not being delayed is

\[
\Omega_{\text{III},w} = \sum_{m=0}^{M-1} n_w^{(m)} \cdot R(M-m)
\]

where \( R(M-m) \) can be calculated by using Eq. (38).

3.2.3 Expected Delay

Let

\[
L_O = \text{number of reservations in the queue;}
\]

\[
L_W = \text{number of reservations in window;}
\]

\(
L = \text{total number of reservations waiting for service.}
\)

We thus have

\[
L = L_O + L_W
\]

The overall delay probability is:

\[
\text{Pr}\{\text{Delay} > 0\} = \sum_{w=1}^W r_w (1 - \Omega_{\text{II},w}), \quad i = I, II \text{ or III}
\]

We define as imbedded points the set of instants just after a new slot is shifted into the window. Using Little's formula at imbedded points, the expected delay becomes

\[
E(\text{Delay}) = \frac{E(L)}{\lambda} - E(\tau)
\]

where \( E(\tau) \) is the expected notice interval. Because of the choice of imbedded points, none of the reservations are registered before a new time slot has been passed (we call this instant "the first imbedded point after arrival"). So a user who requests position \( w \) actually asks the service starts \( w - t \) time units from the first imbedded point after arrival. Therefore we have

\[
E(\tau) = \sum_{w=1}^W w \cdot r_w - 1
\]

\( E(L_W) \) can be found by:

\[
E(L_W) = \sum_{w=2}^M w \cdot \Omega_w
\]

Using Little's formula \( E(L_Q) \) can also be easily found:

\[
E(L_Q) = E(K_w) = Q_w'(1)
\]

We have, by applying Eqs. (44), (45) and (46), that

\[
E(\text{Delay} | \text{Delay} > 0) = \frac{E(L)}{\lambda} \left( \frac{\sum_{w=1}^W w \cdot r_w - 1}{\sum_{w=1}^W r_w \cdot (1 - \Omega_w)} \right)
\]

where \( E(L) \) is determined through Equations. (43), (47) and (48).

4 Numerical Evaluations and Discussions

The analytical methods have been applied to systems with the following parameters:

- The window size \( W = 40 \)
- Poisson customer arrivals with intensity \( \lambda \)
- The notice interval distribution is either
  - uniform, i.e. \( r_w = 1/W \) for \( w = 1, 2, \ldots, W \); or
  - geometric, where
    \[ r_w = \frac{p^{w-1} \cdot (1-p)}{1-p}, \quad w = 1, \ldots, W \]

and in our examples we use \( p = 0.1 \).

The analytical solutions are then compared with simulation results (the simulation time is 1,000,000 time slots for each case).

LOSS Model

Systems with different numbers of channels have been tested. All of the results are confirmed by simulations.
TNASQ Model

The equations described in Section 3.2.1 are applied to find $D_n(z)$ and $P_n(z)$, and the delay probabilities and the expected delay for all the three cases (as described in Sections 3.2.2 and 3.2.3). The theoretical results are then compared with the simulation results.

Two examples are shown below: Fig. 4 shows the mean delay probabilities for $M = 1$ where the notice interval distribution is geometric, while Fig. 5 shows the expected delay (provided there is a delay) with $M = 10$ and uniformly distributed notice interval.

![Figure 4. The expected delay probability, $M = 1$, geometric notice interval distribution](image)

![Figure 5. The expected delay (given there is a delay), $M = 10$, uniform notice interval distribution](image)

We find there is fairly good agreement between the analytical and simulated results in two of the three cases (Case I and Case III). For Case II (random choice), however, we observe that the theoretical results generally show higher delay probability and shorter delay time (provided there is a delay) than for Case I. Since the right-in reservations are given the lowest priority in Case I, these results seem to disagree. Compared to simulation results, it seems that the solution for Case II generally over-estimates the delay probability, and consequently under-estimates the expected delay (provided there is a delay). This tendency is more obvious for heavy traffic.

An explanation to this phenomenon is that the reservation demands to a slot are more and more dominated by overflowing reservations as the traffic increases. Due to the dependencies between slots, an already delayed reservation will experience a higher delay probability at the next position than a right-in reservation. This is also observed from simulations. Therefore it is wrong to use the overall delay probability seen by both right-in- and overflowing reservations to a position to approximate the delay probability experienced by right-in reservations only.

So the solutions developed for Case I and Case III can be applied to find the overall delay probability and the expected delay time. As we have seen in Section 3.2.2, these methods can also be applied to predict the conditional probability that a given reservation will get delayed, if the notice interval distribution is known.

Bibliography


