MODELLING AND CONTROL OF TIME-VARYING TELEPHONE TRAFFIC

Janusz FILIPIAK* and Prosper CHEMOUL • and Edward CHLEBUS

* Teletraffic Research Centre, University of Adelaide, Australia
• CNET, Centre Paris A, ATR/SST, Issy-les-Moulineaux, France
○ Telecommunications Department, University of Cracow, Poland

In the paper models for analysis and control of time-varying circuit-switched traffic are developed. The first description refers to online traffic measurements. The second one models mean quantities. The relevance of both modeling techniques to real-time management and control of networks with non-hierarchical routing is discussed.

1. INTRODUCTION

Several meanings can be assigned to the notion of time-dependent traffic. To clarify them, consider an infinite size trunk group with offered traffic \( \rho(t) \), arrival intensity \( \lambda(t) \), and mean holding time \( 1/\mu : \rho(t) = \lambda(t)/\mu \). Assume first that the offered traffic is stationary. When calls arrive or are terminated we observe instantaneous changes of the number of trunks busy. The observation time scale in that case must be of order of \( 1/\lambda \). In the stationary system, the mean and variance of number of trunks busy remain constant.

Assume next, that the offered traffic is nonstationary, i.e., that the arrival intensity \( \lambda \) depends on time : \( \lambda = \lambda(t) \). As previously we can observe the instantaneous state of the trunk group. However, in this case the mean number in the system also evolves in time. Two different situations are possible : (a) change in offered traffic is abrupt ; (b) offered traffic changes slowly.

The first situation is illustrated in Fig. 1, where it is assumed that the traffic starts to be forwarded to the empty system with fixed rate \( \lambda \). We shall point out later on that in this case the mean number in the system evolves on the time scale of order of \( 1/\mu \), as contrasted to the time scale \( 1/\lambda \) of instantaneous traffic fluctuations.

![Fig. 1 - Traffic Fluctuations -](image_url)
In Fig. 2 more realistic patterns of changes in offered traffic are shown. We shall distinguish between two cases: (a) dynamic traffic flow, and (b) quasi-stationary traffic flow.

Roughly speaking, the traffic flow is dynamic if the offered traffic changes so quickly that there is a significant phase shift between the peaks of mean offered traffic and the mean number of trunks busy, see Fig. 2a. The traffic flow is quasi-stationary if that phase shift is not observed, Fig. 2b.

We relate now the above situations to traffic measurements. In modern SPC switching exchanges the following on-line measurements can be taken on each trunk group:

- $x(t_n)$: number of trunks busy at time $t_n$,
- $u_{in}(t_n)$: number of calls connected between $t_{n-1}$ and $t_n$,
- $u_{out}(t_n)$: number of calls terminated between $t_{n-1}$ and $t_n$.

The measurement cycle length will be denoted by $\Delta t = t_n - t_{n-1}$. The obvious conservation equation which holds both for stationary and non-stationary traffic is:

$$x(t_{n+1}) = x(t_n) + u_{in}(t) - u_{out}(t) \quad (1)$$

In Reference [7] the following model corresponding to (1) is derived for the $M/M/\infty$ system:

$$x^*(t_{n+1}) = ax(t_n) + bu_{in}^*(t_{n+1}) \quad (2)$$

Eq. (2) can be used to predict at time $t_n$ the value $x^*(t_{n+1})$, given the last measurement $x(t_n)$ of trunk group state and an estimate $u_{in}^*(t_{n+1})$ of offered traffic.

Assume next, that the traffic measurements are recorded during several days, $k = 1, \ldots, K$. Then, calculating averages $\bar{x}(t_n)$, $\bar{u}_{in}(t_n)$, and $\bar{u}_{out}(t_n)$, and taking into account (1) we get:

$$\bar{x}(t_{n+1}) - \bar{x}(t_n) = \bar{u}_{in}(t_{n+1}) - \bar{u}_{out}(t_{n+1}) \quad (3)$$

or, dividing by $\Delta t$:

$$\frac{d\bar{x}(t)}{dt} = \frac{d\bar{u}_{in}(t)}{dt} - \frac{d\bar{u}_{out}(t)}{dt} \quad (4)$$

2.2A.3.2
To write this equation we have assumed that average quantities evolve on the time scale which is coarse as compared to the period of measurements $\Delta t$. Then $\Delta t \to 0$ as required. Observe that in (4) $d\bar{u}_{in}(t)/dt = \lambda(t)$. Moreover, in the $M/M/\infty$ system: $d\bar{u}_{out}(t)/dt = \mu\bar{x}(t)$ [8]. Thus Eq. (4) converts to:

$$\frac{d\bar{x}(t)}{dt} = -\mu\bar{x}(t) + \lambda(t)$$  \hspace{1cm} (5)

Eq. (5) provides a model of dynamic flows. It is seen now that the average number of trunks busy $\bar{x}(t)$ changes on the time scale determined by $1/\mu$, as mentioned previously.

In the paper we investigate models (2) and (5) in more detail. Their validity is extended to finite size trunk groups. An example illustrating the modelling and control of dynamic flows in non-hierarchical networks is given.

2. MODEL OF ON-LINE TRAFFIC MEASUREMENTS ON A TRUNK GROUP

The stochastic difference equation corresponding to (2) has the form, [7]:

$$x_{n+1} = ax_n + bu_{n+1} + v_{n+1}$$  \hspace{1cm} (6)

where $v_{n+1}$ is a white noise process having a zero mean, and model coefficients are defined as follows:

$$a = \exp(-T), \quad b = \frac{1 - \exp(-T)}{T}$$  \hspace{1cm} (7)

$T$ is the update cycle length normalized by the mean holding time: $T = \Delta t/\tau$, where $\tau = 1/\mu$.

To derive a similar model for finite capacity trunk groups, we use the equation:

$$\frac{d\bar{x}(t)}{dt} = -\mu\bar{x}(t) + \lambda[1 - p(m,t)]$$  \hspace{1cm} (8)

which describes the evolution of mean state in the $M/M/m/m$ system. Eq. (8) can be easily derived from the birth-and-death equations. $p_m(t)$ in Eq. (8) is the call blocking probability.

We can express the conditional mean $E[x(t_{n+1})|x(t_n) = x_n]$ as follows:

$$E[x(t_{n+1})|x(t_n) = x_n] = E[x^d(t_{n+1})|x(t_n) = x_n] + E[x^b(t_{n+1})|x(t_n) = x_n]$$

In this equation $x^d(t_{n+1})$ is the number of trunks busy at time $t_{n+1}$, seized by calls connected before time $t_n$, and $x^b(t_{n+1})$ denotes the number of calls in progress, connected after $t_n$. Since $x(t)$ is a pure death process we calculate $E[x^d(t_{n+1})|x(t_n) = x_n]$ from (8) taking $\lambda = 0$. We get:

$$\bar{x}^d(t) = x_n \exp[-\mu(t-t_n)], \quad t \in [t_n, \infty]$$

Thus:

$$\bar{x}^d(t_{n+1}) = E[x^d(t_{n+1})|x(t_n) = x_n] = ax_n,$$

where $a = \exp(-T)$

To determine $\bar{x}^b(t_{n+1}) = E[x^b(t_{n+1})|x(t_n) = x_n]$ we make the assumption that the process in stationary and that $p_m(t) = p_m(t_{n+1})$ is fixed during time interval $[t_n, t_{n+1}]$. Then, it follows from (8) that:

2.2A.3.3
\[ x^b(t) = [1 - \exp(-\mu(t-t_n))] \lambda (1 - p_m)^2 \]
giving: \[ x^b(t_{n+1}) = \mathbb{E}[x^b(t_{n+1}) | x(t_n) = x_n] = b \bar{u}_{n+1}, \]
where \( b = [1 - \exp(-T)] \) (10)

Variable \( u_{n+1} \) corresponds to the number of calls connected in time interval \([t_n, t_{n+1}]\):

\[ \bar{u}_{n+1} = \lambda_{n+1}(1 - p_{n+1}) \Delta t \]

We introduce \( \bar{u}_{n+1} \) since this quantity is often recorded during on-line measurement process.

Taking into account (6) and (10) we obtain that:

\[ \mathbb{E}[x(t_{n+1}) | x(t_n) = x_n] = ax_n + b \bar{u}_{n+1}, \]
which gives the following regression:

\[ x_{n+1} = ax_n + b \bar{u}_{n+1} + s_{n+1} \] (12)

Comparing (9) and (10) with (2) it is easily seen that models (6) and (12) have the same structure and coefficients.

In the derivation of model (12), we have assumed an exponentially distributed service time. It can be shown that the validity of this model may be extended to more general service time distributions.

3. ANALYSIS OF MODELLING AND PREDICTION ERRORS

In order to evaluate the statistical properties of the proposed model, extensive simulations have been done. We have investigated the modelling error \( \sigma^2 \):

\[ \sigma^2 = \frac{1}{N-1} \sum_{n=1}^{N} (x^*_n - x_{n+1})^2 \] (13)

which measures the difference between traffic updates and forecasts.

Using (10), we may write (11) as follows:

\[ x^*_n = ax_n + \frac{1 - a}{T} u^*_n \] (14)

It is to be noted that linear prediction of link state based on measurements of \( x_n \) and \( u_n \) must have the form (14). The problem consists of finding the coefficient \( a \). In the previous section we have found that \( a = \exp(-T) \). We assume now that the parameter \( a \) is not known and seek for it by means of simulation. More precisely, we look for the parameter \( a_{\text{min}} \) which minimizes \( \sigma^2 \).

The dependence of \( \sigma^2 \) on parameter \( a \), obtained from simulation for \( T = 0.5 \), is shown in Fig. 3, where \( \rho \) is the intensity of offered traffic in erlangs. We have \( a_{\text{min}} = 0.6 \), which is consistent with the theoretical value \( a_{\text{th}} = 0.6065 \).

In the described experiment we have used a specific value of \( T \). Characteristics \( \sigma^2(a) \) for two other values of \( T \) are shown in Fig. 4. We present these curves in order to illustrate, that for large values of \( T \) the knowledge of \( x_n \) is irrelevant to the prediction of \( x_{n+1} \), since the link state changes significantly during a long update interval. On the other hand, for short update cycle length the simple prediction formula \( x^*_n = x_n \) could be possibly used.
4. DYNAMIC FLOW MODEL OF ROUTING IN NON-HIERARCHICAL NETWORKS

The evolution of the mean number of trunks busy in the M/M/∞ system is described by Eq. (5). For the finite capacity trunk group with exponential distributions of service and interarrival times, a similar model is given by Eq. (8). However, the form (8) of dynamic flow description is not always useful because $p_m(t)$ is not known explicitly. Therefore, we approximate the traffic rejected from the system by a function $\mu[G(\bar{x}) - \bar{x}]$ of the mean system state:

$$\lambda p_m(t) = \mu[G(\bar{x}(t)) - \bar{x}(t)]$$  \hspace{1cm} (15)

Using (15) in (8) yields:

$$\frac{d\bar{x}}{dt} = -\mu G(\bar{x}) + \lambda$$ \hspace{1cm} (16a)

where intensities of carried and rejected traffic are:

$$\lambda^C = \mu \bar{x},$$  \hspace{1cm} (16b)

and

$$\lambda^R = \mu[G(\bar{x}) - \bar{x}]$$ \hspace{1cm} (16c)

respectively. Properties of this model are investigated in [8]. For quasistationary systems we have $d\bar{x}/dt = 0$, and then:

$$G(\bar{x}) = \rho$$ \hspace{1cm} (17)

Equation (17) defines a relation between the mean number in the system and the offered traffic. For the steady state of the M/M/m/m system the same dependence is given by:

$$\bar{x} = \rho[1 - E(m, \rho)]$$ \hspace{1cm} (18)

Thus, we see, that (17) is reverse to (18), which can be used to determine the functions $G(\bar{x})$. A model similar to (16) was introduced in [11].
The model (16) is convenient for analysis of routing in non-hierarchical networks. Recently, several time-dependent routing techniques for such networks have been proposed [1-4]. Basically, they work as follows. Traffic $r_i(t)$ offered at node $i$ is forwarded to destination $k$ through the direct link $(i, k)$. If all trunks are busy then an alternative connection through the tandem exchange $j$ is attempted. There are several possibilities as regards to the choice of the alternative path. Usually, a routing algorithm tends to choose slightly loaded trunk groups. In each case a route is composed of two links at most.

We model the routing in non-hierarchical networks by means of control variables $a_{ijk}(t)$. Routing variable $a_{ijk}(t)$ define the portion of overflow from $(i, k)$ which is routed to the alternative path $(i, j, k)$. We look for routing variables $a_{ijk}(t)$ such that the mean number of calls rejected in the time period $[0, T]$, $T \to \infty$, is minimized. Using the introduced notation the total number of calls discarded from the network is:

$$Q = \int_0^T \sum_{(u,v) \in \mathcal{L}} [G_{uv}(x_{uv}) - x_{uv}] dt$$

(19)

The problem of minimizing (19) by using one-moment models was investigated by several authors see [8-12] where further references are given. We adopt here the solution given in [8]. Define the state dependent length $\Omega(x_{ij}, x_{jk})$ of route $(i, j, k)$:

$$\Omega(x_{ij}, x_{jk}) = 1 - \omega_{ij}(x_{ij}) \omega_{jk}(x_{jk})$$

(20)

where $\omega_{uv}(x_{uv}) = [-dG_{uv}/dx_{uv}]^{-1}$

(21)

It is shown in [8] that the number of blocked calls is minimized if all alternative routes used by the traffic have equal length in the sense of definition (20) and are shorter than unused routes. This property can be used to calculate the optimal states $x^*$ and optimal controls $a^*$.

The foregoing results suggests several routing algorithms.

Algorithm 1: The optimal controls $a^*(t)$ and optimal flows $f^*(t)$ are calculated. The flow realization techniques (e.g., the sequential routing) are used to implement the optimal flow pattern. In principle, this is the DNHR approach [1], [6].

Algorithm 2: The optimal trunk group states $x^*_{uv}(t)$ are calculated. Calls are routed so as to keep the actual network state $x_{uv}(t)$ in close proximity to the optimal state $x^*_{uv}(t)$, [9].

Algorithm 3: The routes are ordered according to (20). Each call tries $n$ routes in succession starting with the shortest one.

In this paper we investigate in more detail the Algorithm 3. It allows us to omit the arduous calculations of optimal state and control variables. The Algorithm 3 is similar to the algorithm used in the Toronto field trial [3] except that another measure of trunk group congestion was used. At first, the residual capacity $m-x$ of each trunk group was calculated. The route quality was then determined according to the formula:

$$\Omega_1(x_{ij}, x_{jk}) = \min(m_{ij} - x_{ij}, m_{jk} - x_{jk})$$

(22)

The general idea behind both approaches is to forward the traffic along the routes having the largest number of idle trunks.
To apply our approach the function \( G(x) \) must be found. In most cases we have tried the following simple approximation [8]:

\[
G(x) = \begin{cases} 
    x, & \text{for } x < m/2 \\
    m/2 - a \ln \frac{2(m-x)}{m}, & \text{for } m/2 < x < m
\end{cases}
\]

(23)

which for congested links, \( x_{uv} > m_{uv}/2 \), gives:

\[
\Omega_2(x_{ij}, x_{jk}) = 1 - \frac{m_{ij} - x_{ij}}{a_{ij}} \frac{m_{jk} - x_{ik}}{a_{jk}}
\]

(24)

Thus, what we propose may be viewed as using (24) instead of (22) in the algorithm proposed in [3]. The deeper insight shows, however, that the absolute values of route lengths are not so important as the resulting ordering of possible routes.

The performance of the proposed algorithm has been evaluated by means of simulation. In order to get comparable results we have simulated the Toronto experimental network, which parameters are given in [3]. The sample results are shown in Tables 1a and 1b. It is seen that for slightly loaded network the proposed algorithm gives worse results than the residual capacity algorithm, while for bigger loads it performs better. The poor performance of our approach for small loads is partly due to the simplicity of approximation (23). It is to be noted that for big update periods the network efficiency may be improved by admitting multiple overflows.

<table>
<thead>
<tr>
<th>Number of overflows</th>
<th>Update cycle length (s)</th>
<th>Number of overflows</th>
<th>Update cycle length (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>15</td>
<td></td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td></td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td></td>
<td>60</td>
</tr>
<tr>
<td>a</td>
<td>0.049</td>
<td>0.066</td>
<td>0.181</td>
</tr>
<tr>
<td>b</td>
<td>0.082</td>
<td>0.148</td>
<td>0.264</td>
</tr>
<tr>
<td>a</td>
<td>0.033</td>
<td>0.049</td>
<td>0.099</td>
</tr>
<tr>
<td>b</td>
<td>0.049</td>
<td>0.066</td>
<td>0.181</td>
</tr>
<tr>
<td>1</td>
<td>a</td>
<td>b</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>a</td>
<td>b</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>b</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>a</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>b</td>
<td></td>
</tr>
</tbody>
</table>

Table 1 - Traffic Loss (%) in the Toronto Field Trial Network - (a : Residual Capacity; b : Proposed algorithm)

5. CONCLUDING REMARKS

Due to space limitations the paper is significantly abridged. In particular, the relevance of prediction methods developed in Sections 2 to 3 to dynamic routing derived in Section 4 should be elucidated. We only note here that variables \( x_{ij} \) and \( x_{jk} \) which appear in (24) are mean quantities which can be evaluated by using directly the prediction formulae given in Sections 2 and 3. Alternatively, the Kalman filter can be applied to improve the prediction quality, cf. [5]. Such approaches are a subject of further study.

2.2A.3.7
REFERENCES


