EFFECTIVENESS VS. CONTROLLABILITY IN TRAFFIC ENGINEERING

Józef LUBACZ, Artur TOMASZEWSKI

Warsaw University of Technology, Poland

The paper concerns modeling the performance of teletraffic systems in fixed-length sequences (windows) of consecutive service demands acceptance/rejection decisions. The presented results are focused on evaluating and controlling the distribution of losses in a window. The proposed methodology employs a combination of discrete-time Markov processes of order $K > 0$ and simulation.

1. INTRODUCTION

The classical Erlangian trunk group engineering may be viewed as a process of controlling loss $B$ with traffic offered $\rho$ and trunk group dimension $N$, whereas the effectiveness of the control may be measured with $E = (1 - B)p/N$. Clearly, in the context of this and similar classical problems the term control can hardly be justified as the system performance is analyzed in steady-state conditions: $N$ is fixed, both $\rho$ and $B$ are stationary averages, the acceptance/loss decisions are independent on the observed performance.

The problem of the interdependence (tradeoff) between effectiveness and controllability nevertheless arises naturally if at least one of these assumptions/constraints is relaxed. The need for such generalization becomes especially clear in the context of multiservice, e.g. ATM, environments - link throughput is virtual rather than fixed, performance evaluation in terms of long term averages does not provide sufficient characterization, dynamic acceptance/rejection procedures have to be introduced.

For many cases in which the standard assumptions in traffic engineering are generalized, the central problem becomes evaluating and controlling the distribution of losses. Our research is focused on such cases.

We aim at controlling the distribution of losses in a window. The window is not a time interval - it is a finite, fixed-length subsequence of the sequence of consecutive acceptance/rejection decisions imposed on the input process of service demands (calls, packets, cells, etc.). In consequence, our analysis is based on discrete-time processes; we shall use order-$K$ Markov chains ($K > 0$; $K = 1$ corresponds to an ordinary Markov chain).

The proposed approach yields, we believe, new options in modeling the performance of teletraffic systems as it not only shifts the focus from continuous-time to discrete-time processes, but also aims at hybrid analytical/simulation performance evaluation methods purely analytical methods are infeasible in most practical cases, whatever the considered time domain).

The paper is organized as follows. First, in Sec.2, we define the notion of a window, the process associated with it and discuss several general properties of the distribution of losses in a window of arbitrary length.

Next, in Sec.3, we illustrate the proposed methodology of evaluating the distribution of losses for several basic systems, and also the controllability vs. effectiveness problem. This part concerns problems analogous to the classical analysis/dimensioning based on average loss; the difference is in that the average is substituted with the distribution.

In Sec.4 we take one more step and assume that acceptance/rejection decisions may depend on the number of currently observed losses in a window, i.e. that the distribution of losses is controlled dynamically (in the considerations from Sec.3 losses may be regarded as controlled statically).

2. LOSSES IN WINDOWS

Consider a stream of arrivals (calls, bursts, packets, etc.) to a system. Let the arrivals be numbered $1, 2, \ldots$, and valued 0 or 1, depending on whether an arrival is accepted for service or rejected (lost), respectively. In the following we shall consider binary sequences $\{a_i\}$ ($i = 1, 2, \ldots$; $a_i = 0, 1$) with the above interpretation. The sequences are trajectories of a stochastic process $A = \{A_i\}, i = 1, 2, \ldots$.

We shall model process $A$ with a Markov chain of order $K > 0$:
\[ P \left( A_{i} = a_{i}' \mid A_{i-1} = a_{i-1}', A_{i-2} = a_{i-2}', \ldots, A_{1} = a_{1}' \right) = \]
\[ P \left( A_{1} = a_{1}' \mid A_{1-1} = a_{1-1}', A_{1-2} = a_{1-2}', \ldots, A_{i-K} = a_{i-K}' \right), \quad (1) \]
for every \( i > K \), and any \( a_{1}', a_{2}', \ldots, a_{i}' \in \{0,1\} \).

In the following we shall use an abbreviated notation for transition and state probabilities:
\[ P(a_{1}'|a_{1-1}', a_{2}', \ldots, a_{i-K}') \text{ and } P(a_{i-K}'|a_{i-K-1}', a_{i-1}') \quad (2) \]

Throughout our discussion we shall assume that process \( A \) is stationary so the probabilities above do not depend on \( i \).

Consider \( w \) consecutive elements of \( \{A_{i}\} \): \( A_{i-1}', A_{i-2}', \ldots, A_{1}' \). We shall say that the elements constitute a window of length \( w \), and denote it with \( W_{w}(j) \).

Let \( P_{w}(n) \) denote the probability that \( W_{w}(j) \) contains \( n \) elements of value 1, i.e. the probability that the number of losses equals \( n \). As process \( A \) is stationary, this probability does not depend on \( j \) (on the position of the window in the sequence \( \{A_{i}\} \)).

Let \( P_{w}(n_{1}|n_{2}) \) denote the probability that the number of losses in \( W_{w}(j+1) \) equals \( n_{1} \), on the condition that the number of losses in \( W_{w}(j) \) equals \( n_{2} \). As process \( A \) is stationary and thus homogenous, this transition probability does not depend on \( j \).

Our main goal in this section is to express the distributions \( P_{w}(n) \) and \( P_{w}(n_{1}|n_{2}) \) through the quantities that determine process \( A \), i.e. its state and transition probabilities.

In intermediate computations, with \( P_{w}(a,n,W_{K}(j)) \) we will denote the probability that \( W_{w}(j) \) contains \( n \) losses, \( A_{i+1} = a \) and \( W_{K}(j) = \{a_{i+1}, a_{i+2}, \ldots, a_{i-K}\} = \{b_{1}, b_{2}, \ldots, b_{K}\} \), where \( a, b_{1}, b_{2}, \ldots, b_{K} \in \{0,1\} \). Note that \( a \) and \( W_{K}(j) \) denote, respectively, the first element and the sequence of \( K \) last elements (i.e. a window of length \( K \)) from \( W_{w}(j) \). From stationarity of \( A \) these probabilities do not depend on \( j \), so we shall omit the index.

It can be shown that:
\[ P_{w+1}(a,n,W_{K}(j)) = \]
\[ P_{w}(a,n-b_{K},1|b_{1}, \ldots, b_{K-1}) \cdot P(b_{K}|b_{1}, \ldots, b_{K-1}) + \]
\[ P_{w}(a,n-b_{K},b_{1}, \ldots, b_{K-1}) \cdot P(b_{K}|b_{1}, \ldots, b_{K-1}) \quad (3) \]

The recursion with respect to \( w \) yields an effective method for computing \( P_{w}(a,n,W_{K}(j)) \) for any \( w, n, a, W_{K} \). The starting point for the computations follows from:
\[ P_{K}(a,n,W_{K}) = \begin{cases} P(W_{K}), & \text{iff } b_{1} + \ldots + b_{K} = n \text{ and } b_{1} = a \\ 0, & \text{otherwise} \end{cases} \quad (4) \]

\( P(W_{K}) \) are determined by state probabilities of \( A \).

Finally:
\[ P_{w}(n) = \sum_{a, W_{K}} P_{w}(a,n,W_{K}) \quad (5) \]
and
\[ P_{w}(n_{1}|n_{2}) = \]
\[ \begin{cases} \sum_{a, W_{K}} P_{w+1}(0,n_{1},1|b_{2}, \ldots, b_{K}) \cdot P(b_{K}|b_{1}, \ldots, b_{K-1}), & \text{if } n_{1} = n_{2} + 1 \\ \sum_{a, W_{K}} P_{w}(a,n_{2},W_{K}) \cdot P(b_{K}|b_{1}, \ldots, b_{K-1}), & \text{if } n_{1} = n_{2} \\ \sum_{a, W_{K}} P_{w+1}(1,n_{1},1|b_{2}, \ldots, b_{K}), & \text{if } n_{1} = n_{2} - 1 \\ \sum_{a, W_{K}} P_{w}(a,n_{2},W_{K}), & \text{otherwise} \end{cases} \quad (6) \]

where the summations are taken over all possible values of \( a \) and/or \( W_{K} \).

The number of losses in a sequence of consecutive windows \( \{W_{w}(j)\}, j = w, w+1, \ldots, \) forms a Markov chain of order \( K = 1 \) with transition probabilities \( P_{w}(n_{1}|n_{2}) \) and state probabilities \( P_{w}(n) \).

Note that the state probabilities can be computed directly from transition probabilities as follows:
\[ P_{w}(n) = P_{w}(0) \cdot \prod_{k=1}^{n} P_{w}(k|k-1) \quad ; \quad k \leq P_{w}(k-1|k) \quad (7) \]
\[ P_{w}(0) \text{ is determined by the normalization condition.} \]

Note also that the average loss probability
\[ B = \frac{P(1)}{w} \quad (8) \]

independently of the window length \( w \).

In our further discussion we shall also need the probability that the first element of the sequence contained in a window of length \( w \) equals 1 (\( a = 1 \)) on the condition that the total number of losses equals \( n \):
\[ P_{w}(a = 1|n) = \frac{\sum_{W_{K}} P_{w}(1,n,W_{K})}{P_{w}(n)} \quad (9) \]

where the summation is taken over all possible values of \( W_{K} \).

We conclude this section with a useful charac-
terization of the distances between consecutive losses in \( \{a_i\} \). The distance between two sequel losses is defined as the number of consecutive elements 0 that separate two sequel elements 1. Note that a sequence \( \{a_i\} \) may be equivalently defined by a sequence of distances between losses; for example the sequence of elements ...,1,0,0,1,0,1,... is equivalently described by the sequence of distances ...,2,0,1,...

The probability that the distance \( D \) between losses equals \( d \) is defined as:

\[
P(D=d) = P(0^d|1)
\]

where

\[
P(0^d|1) = \frac{P(10^d)}{P(1)}
\]

and \( 0^d \) denotes a sequence of \( d \) consecutive 0s.

Let

\[
p = \frac{P(0^{K+1})}{P(0^K)}
\]

It can be shown that

\[
P(D=d|D\geq K) = (1-p)p^{d-K}
\]

This property is useful in simulation based evaluation of the order \( K \) of the Markov chain associated with process \( A \) (the appropriate order might be quite difficult to find based on theoretical analysis).

3. STATIC CONTROL OF LOSSES

In this section we illustrate the problem of controlling the distribution of losses in a window statically, i.e. with the dimension of system’s resources (or equivalently - with the offered traffic). The problem is thus in its essence a generalization of the classical dimensioning based on long-term average loss.

Consider first the classical Erlangian trunk group. It can be shown that in this case process \( A \) is a Markov chain of order \( K=N \), where \( N \) denotes the trunk group size. As for practical cases \( N \) is large it might be difficult to apply the analytical methods introduced in the previous section. It is thus reasonable to search for an approximation of \( A \) with a Markov chain of order \( K< N \). This is our first objective.

It can be shown that:

\[
p(1|1) = \frac{Q}{Q+N} \quad p(1|0) = \frac{N}{(1-B)(Q+N)}
\]

and

\[
P(\{1\}|10) = P(\{1\}|11) = P(\{1\} | 1)
\]

\[
P(001) = \frac{B-N}{(Q+N)(Q+N-1)}(N-1) + \frac{QN}{Q+N}
\]

where \( p \) is the offered traffic, \( B = P\{1\} \) is the loss probability given by Erlang blocking formula. Considering also that

\[
P(\{0\}|00) = P(\{0\}|01), \quad P(\{1\}|0) = \frac{P(101)}{P(10)}
\]

all the transition and state probabilities necessary for approximating \( A \) with a Markov chain of order \( K=1 \) and \( K=2 \) can be computed.

Using these, together with the methods introduced in the previous section (eq.(3-5)), we can compute the probability distribution of losses \( P_w(n) \) for any \( N,p \) and window length \( w \). In Fig.1 we present the computed \( P_w(n) \) distributions (assuming \( N=30, B=0.01, w=100 \)) for \( K=1, K=2 \) and also \( K=0 \) (which corresponds to independent, non-correlated blocking of arrivals) together with a simulated one.

![Figure 1](image)

As the results indicate, the computed distributions fit the simulated one quite well, even for small values of \( K>0 \); note that this is not true for \( K=0 \). A more synthetic indication of the quality of approximation follows from the comparison of the standard deviation of losses: 0.015, 0.017, 0.021 correspond to \( K=1,2 \) and simulation, respectively.

Fig.2 illustrates the tradeoff between controllability and effectiveness. It depicts the dependance of \( P_w(n/w > B) \) - the probability with which the actual loss in a window (expressed with \( n/w \)) exceeds a nominal loss \( B \), on the trunk group size. The presented results correspond to \( w=500 \) and fixed offered traffic.
\( \rho = 20.3 \) which yields \( B = 0.01 \) for \( N = 30 \); once again the results indicate that a small value of \( K > 0 \) yields good accuracy.

Figure 2

As can be seen, for \( N = 30 \) (nominal dimensioning) in over 40% of windows the actual loss exceeds 0.01, so the dispersion of losses, measured with \( P_w(n/w \geq 0.01) \), is considerable. To decrease the dispersion by e.g. 80% the trunk group size has to be increased by 10%, i.e. the price for gaining considerable control over the fluctuation of losses is moderate in terms of the decrease of effectiveness. Of course the choice of an appropriate measure of loss dispersion and the acceptable price for gaining control over it is an application dependent question.

We now proceed to an example of an element of a circuit switched multiservice network. Consider a link consisting of \( N \) channels offered two Poissonian traffics of intensity \( \lambda_i \) and exponential service holding time with average \( 1/\mu_i \), \( i = 1, 2 \). Service demands from traffic \( i \) require \( m_i \) channels simultaneously; complete sharing of the \( N \) channels is assumed (no access restriction).

Each of the two traffic streams may be associated with a separate process \( A \) differing in the corresponding value of the order \( K \) of the Markov chain. It is the authors’ guess, confirmed by simulation, that \( K = N/m_i \), \( i = 1, 2 \) (unfortunately an analytical proof has not been achieved).

Fig. 3 depicts the simulated distribution of distances between losses \( P(D = d) \) in logarithmic scale for the first traffic stream; the following data was used: \( N = 30, w = 300, \lambda_1 = 11, \lambda_2 = 1, \mu_1 = 1, \mu_2 = 6/11, m_1 = 1, m_2 = 6 \).

The simulation results from Fig.3 may be approximated by a straight line starting from \( N = 30 \). This value, according to property (13), estimates the order of the Markov chain corresponding to stream 1. Similarly, simulation shows that \( K = 5 \) for the second traffic stream.

Consider also a system analogous to the above, modified by adding the usual trunk reservation principle to protect the traffic handicapped in terms of loss probability. Average losses for the two traffic streams are equalized with trunk reservation level \( tr = 5 \) applied to traffic stream 1.

Sample trajectories of the evolution of loss probabilities in a sequence of moving windows \( \{W_w(j)\}, j = w, w + 1, \ldots \), obtained with simulation (for data as above), are shown in Fig.4a and Fig.4b. Fig.4a shows the trajectories (displayed every 25 call arrivals) of traffic stream 1 for complete sharing (thin line) and trunk reservation (thick line); Fig.4b corresponds to traffic stream 2.

The most important observation from Fig.4a and Fig.4b is that trunk reservation has a considerable impact on the character of losses fluctuation for traffic stream 1 (the impact on traffic stream 2 is minor). The loss probability frequently and for longer periods exceeds the nominal long-term average (it is a frequent case that the excess is above 50% and lasts for over 300 consecutive windows). Thus system dimensioning based on the classical long-term average might be unsaved and should be reconsidered; the approach proposed in this paper may be useful in this.

Unfortunately, in contrast to the Erlangian trunk group case considered before, no simple formulae can be derived for state and transition probabilities of processes \( A_n \), neither for the complete sharing nor the trunk reservation case, even if small values of \( K \) are assumed. The alternative is to simulate these values and next utilize them as data for the analytical methods introduced in Sec.2.

Using such a mixed analytical/simulation approach we have performed similar analysis to that for the Erlangian trunk group. Again it occurred that small values of \( K \) may be used to achieve satisfactory
accuracy in evaluating the distribution of losses $P_w(n)$ and its functions. For example, for the most interesting case (greatest dispersion of losses) - traffic stream 1 in the system with trunk reservation, simulation yields standard deviation of losses 0.063, while the mixed analytical/simulation gives values 0.043, 0.049 and 0.055 for $K_1 = 1, 2$ and 3, respectively.

The main qualitative conclusions from this section were also confirmed by other sample analyses in which both traffic arrival patterns and service time distribution were assumed non-Markovian.

Consider again the Erlangian trunk group, with the modification that the decision whether to accept the $(j+1)$-th service demand depends on the number of losses in the window $W_{w(j)}$. Assume a demand is rejected with probability $r_w(s,n)$ if upon its arrival $s$ trunks are busy and the number of losses in $W_{w(j)}$ equals $n$ (of course $r_w(s,n) = 1$ if $n = N$, i.e. if all trunks are busy).

Consider the state of the above system at the service demand arrivals and departures (after service completion), i.e. a discrete time process. Let $P_w(s,n)$ be the probability that the system is in state $(s,n)$ in such an instance. The following relationship between the state probabilities holds:

$$P_w(s,n) = \frac{\lambda}{\lambda + n \cdot \mu} \cdot r_w(s,n) \cdot P_w(a=1 | s) \cdot P_w(s,n) +$$
$$\frac{\lambda}{\lambda + n \cdot \mu} \cdot r_w(s-1,n) \cdot (1 - P_w(a=1 | s-1)) \cdot P_w(s-1,n) +$$
$$\frac{\lambda}{\lambda + (n-1) \cdot \mu} \cdot (1 - r_w(s,n-1)) \cdot (1 - P_w(a=1 | s)) \cdot P_w(s,n-1) +$$
$$\frac{\lambda}{\lambda + (n+1) \cdot \mu} \cdot (1 - r_w(s+1,n-1)) \cdot P_w(a=1 | s+1) \cdot P_w(s+1,n-1) +$$
$$\frac{(n+1) \cdot \mu}{\lambda + (n+1) \cdot \mu} \cdot P_w(s,n+1),$$

(17)

where $P_w(a=1 | n)$ is given by (9). Note that the state probabilities depend on the distribution of losses in the window, which is not known a priori as it depends on $P_w(a=1 | n)$; thus an iterative procedure has to be applied to solve the above equations. This, together with the complexity of the equations (two-dimensional state space) makes the problem quite complex numerically. Note also that due to the feedback introduced by the control window, process $A$ is of order $K = w$.

4. DYNAMIC CONTROL OF LOSSES

In the considerations from the previous section the distribution of losses in a window was controlled statically - the strategy of accepting service demands was independent of the current number of losses in previous windows. We shall now illustrate the problem of controlling losses dispersion assuming that the information about the current number of losses in a window is utilized in the strategy of accepting future service demands, i.e. that losses are controlled dynamically.

To illustrate the impact of the window-based control of service demand acceptance on the dispersion of losses we will compare the following versions of the system with two traffic streams considered in the previous section:

V.I: No window-based losses control. Service demand acceptance is based on trunk reservation.

V.II: Window-based losses control. For each traffic type a separate, moving control window is intro-
duced; current loss probabilities in the windows are evaluated. Service demand acceptance is based on the following rule: if at the instance of an arrival of stream 1 service demand the current loss probability for stream 1 traffic is smaller than the current loss probability for stream 2 traffic, then reject the demand. No restrictions are put on the acceptance of stream 2 service demands. Note that the service demand acceptance strategy from V.I uses only the information about the current number of busy channels, whereas the strategy from V.II uses only the information about losses in current control windows.

Fig. 5 illustrates sample trajectories of the evolution of traffic stream 1 loss probabilities in the sequence of moving windows \( \{ W_n(j) \} \), obtained with simulation for traffic data as in Sec.3; the thin line corresponds to V.I, the thick line - to V.II. The two versions were simulated concurrently, fed with the same input streams.

![Figure 5](image)

As can be seen from the figure, the strategy of service demands acceptance/rejection from V.II, though extremely simple, has significantly suppressed the fluctuation of losses with respect to V.I (trunk reservation). Even better results are obtained if the strategy is enriched.

Obviously, the power of window controlled strategies results from the fact that they introduce a feedback into systems performance. Such strategies aim at providing required performance in terms of quality of service, rather than at maximizing effectiveness in terms of resources utilization. The latter nevertheless becomes somewhat less crucial in the context of wideband multiservice systems for which we face serious problems with reliable traffic forecasts, system modelling, policing and admission control, etc.; the problem of controllability seems to dominate the problem of effectiveness.

5. CONCLUDING REMARKS

The proposed approach to controlling system performance has been illustrated with the analysis of simple elements of circuit switched networks. Nevertheless there are no methodological obstacles in using it for analysis of e.g. elements of an ATM system (finite capacity queues offered heterogenous traffic streams of complex cell arrival patterns); in fact problems encountered in the analysis of multiservice systems inspired our research. We have chosen to illustrate the discussion with classical systems rather than ATM-like ones, hoping this will provide a clearer insight into the essence of the proposed approach. Moreover, we believe that it might be useful to reconsider the good-old methodology of e.g. dimensioning telephony networks - are we really satisfied with dimensioning based on long-term average loss? The results presented in the paper have revealed that engineering based on long-term averages may be quite ambiguous and risky, especially for multiservice circuit switched networks (cf. traffic stream 1 in the system with trunk reservation), and also that considerable control over the dispersion of losses can be achieved for the price of moderate decrease of resources utilization effectiveness.

The main practical limitation in the range of application of the proposed methodology is dependent on the magnitude of the order \( K \) of the Markov chain that has to be adopted in the modelling. On one hand, as even our simple examples indicate, \( K \) has to be quite large if all details of correlation of events is to be mirrored. On the other hand though, if only some aggregated measures of the events are of interest, such as e.g. the distribution of losses or its function, then it is enough to adopt a small \( K \) to achieve satisfactory accuracy. This is likely the most important practical observation of our research.

Another difficulty in applying the methodology may arise in the cases for which the modelled events occur with very low probability (like e.g. buffer overflows in ATM system components). In such cases obtaining the required data from simulation may be a burden. But this is a general problem which could be bypassed if purely analytical approaches were available and well established, which unfortunately so far is not the case.