NUMERICAL ANALYSIS OF OVERFLOW TRAFFIC AND A DECOMPOSITION TECHNIQUE FOR THE PERFORMANCE EVALUATION OF OVERFLOW SYSTEMS

Peter BUCHHOLZ and Bruno MÜLLER-CLOSTERMANN

Department of Informatics
University of Dortmund
Dortmund, Federal Republic of Germany

In order to achieve good and efficient approximations in the performance analysis of overflow systems a decomposition technique based on the numerical evaluation of transient Markov chains is introduced. The basic technique is the construction of substitute processes which approximate the original overflow traffic. Due to decomposition the evaluation of large realistic models can be mastered very efficiently.

1. INTRODUCTION

Modelling and performance evaluation of communication systems by means of queueing network models have become an important technique. Because of phenomena like blocking due to limited buffers (leading to rejections of messages), peakedness of overflow traffic and various mechanisms for flow and congestion control, the well-known product form networks (for an introduction see e.g. [1]) are of limited value in the context of telecommunication systems. In particular the neglect of overflow traffic characteristics would introduce unacceptable errors into the analysis.

For the mathematical treatment of these problems approximate techniques have been developed and applied successfully, see [2] for 2-moment methods like ERT, [3,4] for 3-moment methods, and [5] for a sophisticated 1-moment method. Analytical formulas for rejected and served traffic have also been developed.

In this paper we present a very general decomposition technique for the approximate analysis of overflow systems which combines known analytical formulas for traffic characteristics with the numerical evaluation of traffic streams by means of absorbing Markov chains. Due to decomposition the proposed technique can be efficiently applied to very large models and it is more accurate than the ERT-method.
2. A MODEL OF AN OVERFLOW SYSTEM

The presentation of our technique refers to the overflow model depicted in fig 1a. The system has M levels (trunks), level m, m=1,...,M, has N_m lines. Calls arrive according a Poisson process with parameter \( \lambda \) and try to occupy a line at level 1. Calls which are blocked at level i are routed to level i+1, i=1,2,...,M-1. If all lines at all M levels are occupied the call will be cleared. The holding times are assumed to have a phase type distribution, e.g. an Erlang or a Cox distribution, including the exponential case. This open wait-loss model can be transformed into an equivalent model without external arrivals, see fig. 1b. Instead of external arrivals we have a finite source and the number of requests cycling in this model exceeds the total number of lines by one. A "loss" occurs if a request is rejected at all trunks and returns to the source without success. The throughput on this route equals the loss rate. For validation purposes an exact evaluation based on the numerical evaluation of the global balance equations resulting from the model shown in fig. 1b was employed. Of course an efficient application of this brute force approach is limited to relative small models; e.g. in case of three levels, 10 lines per level and exponentially distributed holding times we obtain a model with 1321 states. Obviously there is a need for more efficient techniques.

![Diagram of an overflow system](image1)

![Equivalent model](image2)

3. APPROXIMATE ANALYSIS OF OVERFLOW TRAFFIC

The method presented here is based on model decomposition and numerical evaluation of Markov models. The main principle of the new method is the analysis of overflow traffic which will be substituted by an artificial source emitting traffic according (nearly) the same stochastic pattern, cf. fig.2. We distinguish the following three steps:

1. Determine the moments (in particular the coefficient of variation) or even more detailed characteristics of the traffic distribution. In case of overflow traffic the stochastic variable "time between successive rejections" must be evaluated.

5.1B.5.2
(2) Construct a substitute representation producing traffic which is "equivalent" to the original traffic. In order to account not only for the first moment a phase-type distribution can be used. (Fig. 2 shows a 2-phase Coxian distribution.)

(3) Finally an analysis of the aggregated model yields the desired performance measures. Dependent on the model structure steps 1 and 2 can be repeated leading to further model reduction.

This technique is not limited to the analysis of overflow systems, in particular it can be applied to various kinds of feedforward systems. Note that in case of feedback traffic the proposed technique can not be applied without modification. Now we discuss steps 1-3 in more detail.

3.1. Numerical Evaluation of the Overflow Process

The goal of step 1 is the determination of the characteristics of the overflow traffic, see fig. 3a. First we transform the model into an equivalent closed model, see fig. 3b, where N+1 requests (calls) are cycling in a closed chain between the service device and an additional station which represents the source. We assume the model to start in a state where N requests are occupying the service device and the (N+1)th message has been just rejected and restarts in the source (or in one out of several sources respectively). The first time a request reaches the sink, the process is defined to be in an absorbing (=recurrent) set of states.

For the determination of distributional characteristics of "time until absorption" we employed numerical techniques based on finite Markov chains. Two analysis techniques are sketched shortly below, both founded on the transition rate matrix Q of the Markov chain associated with a model as shown in fig. 3b.
For the computation of the k-th moment of the "inter overflow distribution", which is identical to the distribution of the time until absorption, we have to solve systems of linear equations \( \mathbf{T} \mathbf{E}_k = -k \mathbf{E}_{k-1} \) (starting with \( \mathbf{E}_0 = \mathbf{1} \)); \( \mathbf{T} \) is the submatrix representing the transitions between transient states, \( \mathbf{E}_k \) denotes the vector of the conditional k-th moments, i.e. the i-th component of the vector, \( i = 1, 2, \ldots, \text{ord}(\mathbf{T}) \), is the k-th moment under the condition that the process started at time \( t=0 \) in state i. For details, see [6].

The determination of the distribution function can be done approximately by so called randomization. The recurrent ("absorbing") subset associated with the submatrix \( \mathbf{R} \) is lumped to a single absorbing state followed by a transformation of \( \mathbf{Q} \) into a stochastic matrix (\( \mathbf{Q} \) is "randomized"). During the subsequent iteration the probability mass which flows into the absorbing state per iteration step (which represents a time unit) is computed. As result a number of values of the distribution function is obtained, which serves as approximation of the complete distribution. For details, see [7].

3.2. Substitute Overflow Processes

The following choices for the substitute process have been investigated experimentally.

M1: Only the first moment of the process is taken into account, i.e. the substitute process was modelled by an exponential distribution.

M2: The first and the second moment are used to build a 2-phase Coxian server, such that the two phases are equally utilized.

M3: The first three moments are taken to build a 2-phase Coxian server. As a consequence all three parameters are uniquely determined.

D1: Additionally to the first and second moment the original distribution function was used to obtain a good fit of the substitute process. Of course it is a problem to find parameters which yield a minimal difference between original and substitute process; to minimize the difference at some selected points an iterative approach was employed.
Note, that an interrupted Poisson process (IPP) can be transformed into a 2-phase Coxian server, see [9]. For the construction of n-phase distributions (n>2), see [10].

3.3. Evaluation of the Aggregated Model

The use of one of the substitute processes, e.g. a 2-phase Coxian server, yields an enormous reduction in the size of the model. If we apply the method repeatedly we have, in case of our sample model, to deal only with very small models (23 transient states per level), instead of a model with 1321 states. Due to the fact that the state space grows only linear in the number of lines per trunk, numerical evaluation can be applied efficiently also in case of realistic models with about 100 lines per trunk.

4. RESULTS FOR THE SIMPLE OVERFLOW MODEL

We describe some experiments using the proposed method with different substitute processes. Approximate results according the schemes M1, M2, M3 and D1 have been compared to exact results obtained by the global balance approach. As model parameters have been chosen $M=3$, $N_1=N_2=N_3=10$, $\mu_1=1.0$, $\mu_2=0.8$, $\mu_3=0.5$; the arrival rate $\lambda$ was varied from 10 to 40. Taking approximation M1 we have only to handle $M/M/1$-models which obey the Erlang loss formula, see [2]. The other approximation schemes require an initial analysis of the rejected flow from an $M/M/1$-system, what can be done analytical, see [4], followed by the successive numerical analysis of two Coxian $2/M/1$-systems. Due to the very small state space (23 states) the computation effort is very low.

![Fig. 4: Error in level-3-throughput](image)

![Fig. 5: Error in loss rate](image)
The relative errors for the most sensitive quantitative measures, namely "level-3-throughput" and "loss rate" (=amount of cleared traffic) are shown in fig. 4 and 5. Note that the accuracy of the M1-approach is very unsatisfactory, whereas the other approaches behave much better. M3 yields smaller errors than M2, whereas the expense for both methods is nearly identical. The D1-approach produces the smallest error and seems to be preferable if the problem of the parameter estimation of the Cox-distribution can be managed without to much effort.

5. MORE COMPLEX OVERFLOW MODELS

To illustrate the range of applicability of our method we briefly sketch two more complex models. We assume three independent Poisson input streams, each one having its own private trunk with \( N_i \) lines, \( i=1,2,3 \). Additionally there is a shared trunk with \( N_s \) lines. The holding times are exponentially distributed with mean 1.0 for every stream. For our experiments we assume \( N_1=N_2=6, N_3=3, N_s=5 \). Of course the approximate method can be applied to much larger models, but for the ease of validation by global balance techniques we prefer a smaller state space size. The arrival rate \( \lambda_1 \) for stream 1 was varied from 4 to 10. The arrival rates for streams 2 and 3 are 4.0 and 2.0 respectively. Now we distinguish two strategies for the assignment of the streams to the trunks.

The first strategy is to take a line from the associated private trunk, and if all lines are occupied, to try the shared trunk. In case of no success, the call is cleared. An exact analytical technique has been proposed recently for a similar model with two arrival streams [11]. The second strategy is to take first a line from the shared trunk, and if all lines of the shared trunk are occupied, the associated private trunk will be chosen. The resulting models have 1176 states (in both cases), whereas our approximation in case of strategy 1 only needs to handle systems of 13 equations and in case of strategy 2 two systems of 13 and one of 5 equations. With regard to the results shown in fig. 4 the M3-method for the construction of the substitute process was employed.

![Fig. 6: Loss Probabilities for strategy 1 and 2 (dashed lines are for strategy 2)](image-url)
The relative errors for loss probabilities turned out to be smaller than 3% and 0.1% (for strategy 1 and 2 respectively). Fig. 6 shows the values for the three streams as a function of the arrival rate $\lambda_1$ of stream 1. The dashed curves show the values for strategy 2.

6. CONCLUSIONS

The experiments show that a careful parametrization of the substitute process considerably increases the accuracy of the results. Because the models for each level are relatively small the proposed technique is very efficient when compared with global balance analysis or simulation. Note that in case of 100 lines per trunk the number of model states is about $10^6$, whereas the proposed method only requires the sequential solution of $M$ relatively small problems of about $q^*100$ states, where $q$ denotes the number of Coxian phases; as a consequence this method allows the evaluation of large realistic models.

Finally, we would like to emphasize that the availability of modelling tools which support the construction and analysis of Markovian models is very helpful for an application of the proposed method. For the work presented here, the modelling tool NUMAS [12] and partly the Markovian solver of QNAP2 [13] have been employed.

REFERENCES