FAST CALCULATION OF THE CAC CONVOLUTION ALGORITHM USING THE MULTINOMIAL DISTRIBUTION FUNCTION.

J.L. Marzo+, R. Fabregat+, J. Domingo*, J. Sole* [1]

ABSTRACT. This paper is focussed on one of the methods for Connection Acceptance Control (CAC) in an ATM network: the convolution approach. With the aim of reducing the cost in terms of calculation and storage requirements, we propose to use the multinomial distribution function. This permits direct computation of the associated probabilities of the instantaneous bandwidth requirements. This in turn makes possible a simple deconvolution process. Moreover, under certain conditions additional improvements may be achieved.

1. INTRODUCTION. The ATM transport network is based on packet switching using small fixed-size packets. ATM permits flexible bandwidth allocation, so that an important objective is to obtain the maximum statistical gain on a physical channel. A system of traffic control is therefore required to ensure adequate Quality of Service; the CAC is a tool designed for this purpose.

The CAC is a procedure that decides whether a new connection can be accepted by the ATM network. The decision is based on the resources occupied by the existing connections and the characteristics of the new connection.

In this paper connection types are classified 0,1,...,T-1. All connections within one class have identical traffic parameters and identical Quality of Service requirements. Given that $x_0, x_1, \ldots, x_{T-1}$ connections of classes 0,1,...,T-1 are present on a link of capacity C, the CAC polices the network to decide when a new call of class i will be accepted, given that connection requirements have to be met for all classes. There are four well known algorithms that attempt to answer this question: peak rate allocation, linear CAC procedure, the two-moment allocation scheme and the convolution approach [1].

The convolution approach is based on the probability density function of the instantaneous bandwidth requirements of each established connection. This is the most accurate method but it has a considerable computation cost and a high number of accumulated calculations. Nevertheless, in critical near-congestion situations, the convolution is the only algorithm that gives enough accuracy.

The objective of this study is to evaluate the usage requirements of the CAC convolution approach. Two different implementations have been studied, the first based on the convolution expression, and the second based on the use of the Multinomial Distribution Function (MDF). The first implementation is studied in section 2; the MDF and its associated methods of calculation are analysed in section 3. Some improvements based on the fast convolution function will be explained in sections 3.4 and 3.5.

Both studies have been carried out according to the GMDP (General Modulated Deterministic

---

[1] + Universitat de Girona. EPS. Girona (Spain). Tel.: +34 72 210262. Fax +34 72 210326
* Universitat Politècnica de Catalunya. DAC. Barcelona (Spain). Tel.: +34 3 4016982. Fax +34 3 4017055.
This work has been supported by CICYT (Spanish Education Ministry) under contract TIC92-1289-PB.

Process) model. The GMDP model describes the behavior of a traffic source at cell and burst level. In each state, STi, a distributed number of cells are sent with regular interarrival times (constant rate r_i) during the corresponding sojourn time TS_i. From mean sojourn times we readily obtain the state probabilities under a set of given assumptions.

So, for each type of source there is a corresponding vector: the size of the vector is determined by the number of source states possible. There are two fields: Rate, and Probability; the Rate is the rate of transmission in a given state and the Probability is the probability that the source is in this state.

2. ANALYSIS OF THE CAC CONVOLUTION APPROACH. This section contains the calculation based on the convolution expression:

$$P(Y+X_j) = \sum_{k=0}^{n} P(Y=n-k) \cdot P(X_j=k)$$

where Y is the bandwidth requirement of the already established connections; X_j is the bandwidth requirement of a connection of traffic type j, and n denotes the required bandwidth.

To store all the transmission rates possible for the connections established at a given moment, and to store the probability that the sources will emit at those rates a System Status Vector (SSV) is needed. This vector has the same two fields, Rate & Probability. When a new connection is made the SSV must be updated. The corresponding source vector is used to make this update in the following way:

for each old SSV element a set of new SSV elements is generated. The number of these elements is determined by the number of emission states possible from the new connection. The Rate of each new element is the sum of the existing rate and the rate corresponding to the state of the new source. The Probability of each new element is the product of the existing probability and the probability corresponding to the state of the new source.

Figure 1 shows an example for two simple states of the system. Fig 1.1 shows the status corresponding to one source of type 0, (1,0,...,0). Fig 1.2 shows the convolution of two sources of type 0, (2,0,...,0). A source of type 0 is defined thus : it is in state E_0 (Rate = 0.2 Mbit/s) with Probability 0.7; it is in state E_1 (R = 0.4 Mbit/s) with Pr. 0.2; and it is in state E_2 (R = 2 Mbit/s) with Pr. 0.1.

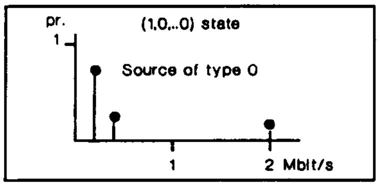

Fig 1.1

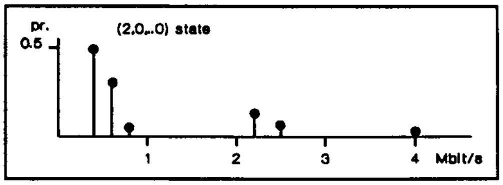

Fig 1.2

The same rate may appear more than once. The size of the SSV may be reduced by sorting and combining the repeated rates; the amount of reduction which can be achieved depends on the rates of the source type.

2.1 Implementation Cost.

Let **N** equal the number of total sources,

where $N = \sum_{k=0}^{T-1} N_k$ and where N_k is the number of sources of type k.

Let **T** be the number of types of source, and

$\mathbf{T_k}$ be the number of states of the source type k. The number of elements necessary to store the SSV is: $NumElem = \prod_{k=0}^{T-1} T_k^{N_k}$ and its calculation cost $NumElem * T_k * (pc*1 + sc*1)$

where **pc** is the product cost, and **sc** is the sum cost.

1) Note the great amount of memory storage required by the SSV. This requirement increases with the number of sources and possible source states.

2) The probability as expressed in the SSV is the result of a large number of previous calculations.

3) Furthermore, this process does not admit of easy deconvolution [2]. In order to arrive at an exact calculation of the vector which would result from a connection release it is necessary to calculate from scratch.

3. FAST CONVOLUTION FUNCTION. We have studied a new, faster method in order to overcome these drawbacks. As we have seen in the previous section, after convolution the same rate may appear more than once in the SSV. Which elements are repeated? How many times?

3.1 Multinomial Distribution Function (MDF). We shall study the Multinomial Distribution Function [3]. We assume one source type only, emitting in k states. Each source state has an associated rate R_i. Therefore, for n connections, x_1 sources are in state E_1; x_2 sources are in state E_2, and X_k sources are in state E_k; we can consider a k-dimensional random variable $(E_1,E_2,\ldots,E_k)$. Therefore a random event that has been repeated n times has the characteristic $(x_1,x_2,\ldots,x_k)$. It is necessary to calculate the probability of E_1 occurring x_1 times, E_2 occurring x_2 times, and E_k x_k times. For this purpose we now consider generalized Bernoulli trials. As in the previous situation assign to the point

$$(\underbrace{E_1,E_1,\ldots,E_1}_{x_1},\ \underbrace{E_2,E_2,\ldots,E_2}_{x_2},\ \ldots,\ \underbrace{E_k,E_k,\ldots,E_k}_{x_k})$$

the probability $p_1^{x1}p_2^{x2}\ldots p_k^{xk}$. This is the probability assigned to any sequence having x_i occurrences of E_i, i=1,2,..,k.

To find the number of such sequences, first select x_1 positions out of n for the x_1's to occur: $\binom{n}{x_1}$ permutations of x_1 identical elements out of n. Then select x_2 positions out of the remaining n-x_1 $\binom{n-x_1}{x_2}$ and so on, x_k in E_k $\binom{x_k}{x_k}$. Note that $x_1+x_2+\ldots+x_k=N$ for all combinations.

Thus the number of sequences having exactly x_1 connections in state E_1, and x_2 connections in state E_2, ... and x_k connections in state E_k is:

$$\binom{n}{x_1}\cdot\binom{n-x_1}{x_2}\ \cdots\ \binom{x_k}{x_k}\ =$$

$$=\frac{n!}{(n-x_1)!\ x_1!}\cdot\frac{(n-x_1)!}{(n-x_1-x_2)!\ x_2!}\cdots\frac{(n-x_1-\ldots-x_{k-1})!}{(n-x_1-\ldots-x_k)!}=\frac{n!}{x_1!x_2!\ \ldots\ x_k!}$$

The probability assigned is:

$$P[E_1=x_1;E_2=x_2;\ \ldots\ E_k=x_k]\ =\ \frac{n!}{x_1!x_2!\ \ldots\ x_k!}\ p_1^{x_1}\ p_2^{x_2}\ \ldots\ p_k^{x_k}$$

where $p_1+p_2+\ldots+p_k=1$.

When $x_1,x_2,\ldots,x_k$ are non-negative integers whose sum is n, this probability is called the multinomial distribution function. Note that the probability of each source being in state E_i is independent of the probability of the other source states.

3.2 Calculation algorithm. The multinomial approach is applied to groups of the same type, and the global state probabilities are evaluated by convolution of the partial results obtained from the different existing groups of sources.

A System Status Matrix (SSM) stores all possible combinations relating to system state. This matrix is defined as follows: for each type of source k there is an associated sub-matrix.

$$\begin{bmatrix} \cdots & \cdots & \cdots \\ \cdots & X_{ij}^{\ k} & \cdots \\ \cdots & \cdots & \cdots \end{bmatrix}^{(k)}$$

where x_{ij}^{k} is the generic element. This generic element stores the number of connections for column i-state, row j-combination. This sub-matrix has M_k rows; this value is the number of possible

combinations of the type k source $M_k\ =\ \sum_{i=1}^{N_k+1} i=\frac{(N_k+1)(N_k+2)}{2}$

$N_k\ =\ x_{j1}^{k},x_{j2}^{k},\ldots,x_{j\pi k}^{k}$, for all row $j=0,\ldots,M_k-1$,

The fact that all possible state combinations relating to a given source type are contained within one sub-matrix means that source disconnection is simply a question of processing this sub-matrix:

deconvolution. Deconvolution is essentially a means of evaluating the relative lightening of the system load after a disconnection.

In this way a generic element of the global SSM is generated by concatenation of all the possible combinations between the rows of sub-matrices.

A global system status row

$$< row_{j1}^{1},\ldots, row_{jk}^{k},\ldots, row_{jT}^{T} >$$

is defined for all $j_p=0,..,M_p-1$ and for all $p=0,\ldots,T_k$, and an element is defined as:

$$\begin{bmatrix} \ldots & \ldots & \ldots \\ \ldots & X_{lm} & \ldots \\ \ldots & \ldots & \ldots \end{bmatrix}$$

It stores the number of connections for each state, where l is the row obtained by combining the sub-rows obtained previously, and where the value of j varies $(j_1, j_2,\ldots,j_T)$; m is the state column corresponding to the partial state i of a type of source k. Therefore, the size of the global SSM will be of dimension

$(T_1+T_2+..+T_k+..+T_T)$ columns * $(M_1*M_2*..*M_k*..*M_T)$ rows.

This is MaxCols * MaxRows. It is not necessary to store either the rate or the probability corresponding to each global state. The rate of each row m is:

$$\sum_{m=0}^{MaxCols} x_{lm} * SourceRate_{i_k}^{k}$$

and p_j^k is the probability for sub-row_{jk}^{k}. It is calculated with the MDF, as we have seen in the previous section, and the total probability for row j is

$$\prod_{k=0}^{T} p_l^{k}$$

Given this system of calculation, results are obtained directly from each element and from there alone, unlike the standard convolution system which obtains results by accumulation.

3.3 Example of Implementation Cost.

For T = 2 and $T_k=3$ for all $k=1,\ldots,T$, the number of all calculations is:

$$M_0 * M_1 * [opc_0 + opc_1 + pc * 1 + sc * 2]$$

where $opc_k = pc*4*N_k + sc*3*(N_k-1) + ec*1 + dc*1$

pc is the product cost, sc = sum cost, dc = division cost and ec = exponential cost.

3.4 Balanced Algorithm. In the global SSM the repeated sub-matrices correspond to the same associated calculations. To reduce this calculation cost, we can store partial results in a set of vectors associated with each type of source.

In this method each partial probability, corresponding to a type of source, is evaluated by using the MDF. Therefore the partial probability for any element is the product of the pre-calculated values, probabilities and rates corresponding to the different classes of source.

The Balanced Algorithm solution combines appropriate storage requirements with low calculation cost.

3.5 Cut-off calculations. The process of evaluation using information contained in the System Status Matrix has one clear aim: to indicate whether or not a new connection will be accepted. For this decision the global Cell Loss Probability (CLP) is compared with a previously set value, in order to guarantee a specific QOS. Therefore, if during the process of calculation the accumulated CLP exceeds a previously set value, the process can be stopped, and the calculation cost thus reduced.

A further reduction in calculation cost is obtained as follows: calculation of probability is only carried out in those cases where the associated rate exceeds the physical channel rate. Note that the calculation of probability is considerably more complex than the calculation of rates, so that we select the objects of our calculation on the basis of the latter.

Furthermore, in each block, the rows generated are not examined at random, but are graded according to rate, so that when the pre-set minimum rate C is reached, the process is terminated, and a further saving achieved.

4. SUMMARY. The convolution approach is based on the probability density function of the instantaneous bandwidth requirements of each established connection in an ATM network. This is the most accurate method but it has a considerable computation cost and a high number of accumulated calculations.

In order to reduce the implementation cost of the CAC convolution algorithm, in this paper we propose to use the multinomial distribution function. This approach is applied to groups of sources of the same type, and the global state probabilities are evaluated by the convolution of partial results obtained from the different existing groups of sources. This permits the system to reduce the number of calculations by stopping the evaluation algorithm when the accumulated cell loss probability is greater than the maximum permitted. Furthermore, if the different combinations are examined in the appropriate order the number of operations may be reduced significantly.

REFERENCES

[1] COST 224 Final Report. "Performance Evaluation and Design of Multiservice Networks. Paris October 1991.

[2] William H. "Numerical recipes in Pascal the art of scientific computing". Ed. Cambridge 1.990.

[3] Ash, Robert B. "Basic Probabilty Theory". John Wiley & Sons. Illinois 1969.