ATM Call Control by Neural Networks

Joaquim E. Neves*, Luis B. de Almeida**
and Mário J. Leitão***

** DEI, Universidade do Minho, Azurém, P-4800 Guimarẽs*
*** DEEC-IST, Avenida Rovisco Pais, P-1000 Lisboa*
**** DEEC-FEUP, Universidade do Porto, Rua dos Bragas, P-4000 Porto*
INESC - Instituto de Engenharia de Sistemas e Commputadores
,* Largo Mompilher 22, P-4000 Porto*
*** Rua Alves Redol 9, P-1000 Lisboa*
eneves@inescn.pt, lba@sara.inesc.pt, mleitao@inescn.pt

Abstract

The resource allocation in the Broadband Integrated Services Digital Network (B-ISDN) must guarantee the quality of service negotiated with new and existing calls, taking into account the Asynchronous Transfer Mode (ATM) statistical characteristics. A quality of operation function, characterizing the overall network performance, is proposed, and based on this function, it is introduced a new strategy for the admission control and routing of the ATM call connections. As it is shown by simulation results, feedforword Neural Networks trained with the backpropagation algorithm, can learn the traffic patterns in previous traffic situations, and can be used to predict the quality of operation changes caused by each new call.

1 Introduction

The transport of B-ISDN services, like interactive video telephony, high quality video and audio broadcast programs or data file transfer, requires a call establishment phase. The Asynchronous Transfer Mode provides the transport and the switching of information flow, generated by B-ISDN services, in fixed size data packets called ATM cells which can be transported in the existent Plesiochronous (PDH) or Synchronous (SDH) Digital Hierarchies or in the new cell-based transmission systems [1].

At the call establishment, the user has to negotiate with B-ISDN control entity the traffic characteristics of the call and the quality of service requested. The network control entity can accept the request and allocate network resources for the support of the service, or propose a lower quality of service, and in the limit, rejected the call, if not enough network resources are available. B-ISDN has mechanisms for usage parameter control by policing the call traffic at user interface and taking appropriate actions if the usage values of the information flow parameters, namely, the average and the peak cell rate and the burstiness, are exceeded in a virtual channel.

Neural networks have been proposed to control ATM networks. For example in reference [7], three hierarchical levels of neural networks are suggested to implement the control functions. Neural networks can also be applied in service coding [9] (e.g. video and audio compression), in the B-ISDN policing function, (e.g. the control of a selective cell discard), in switching control [2, 8] and in the processing of the operation and maintenance (OAM) signals. Neural Networks can learn traffic patterns of the network operation in previous situations, to be used in the cost prediction of each new call admission.

This paper discusses the application of neural networks in ATM call control. Section 2 introduces the network quality of operation function and describes its use for call control proposes. Section 3 describes and compares simulation results with different call admission control methods. Section 4 summarizes the main issues of this paper.

2 ATM Call Control Technique

The resource allocation for ATM call connections can be made by taking the peak bit rate of ATM cells sources as reference, but no statistical gain is obtained by multiplexing many sources. If the resource allocation is made by the average bit rate of ATM cell sources, the statistical gain obtained by multiplexing many sources is maximum but the simultaneous occurrences of the peak periods of some sources can drastically increase the delay and cell loss rate. The strategy proposed for call admission control establishes a compromise between the maximum number of calls accepted and the satisfaction of the quality of services negotiated for calls established.

2.1 Quality of Operation of the B-ISDN

The Quality of Operation concept integrates the parameters of the quality of service negotiated by the network in the call establishment, the availability of network resources, and the equilibrium between the call rejection rate of different ATM service classes. The proposed quality of operation function (QO) can be written by the following expression:

$$QO = \sum_j (\alpha_j A_j + \beta_j B_j - \chi_j X_j - \sum_i \delta_{ji} \Delta_{ji}) \tag{1}$$

where α_j, β_j, χ_j, and δ_{ji} are non-negative real control parameters, and A_j, B_j, X_j, and Δ_j are functions. A_j quantifies in terms of quality of operation the bit rate allocated to each service class j; B_j quantifies the bit rate free to be allocated to each service class j; and X_j quantifies the deviation of the call rejection rate of the service class j from the average call rejection rate of all service classes; Δ_{ji} quantifies the main quality of service requirements of each service class j, namely the cell loss rate (i=0), the delay (i=1) and the delay variation (i=2).

The values of the control parameters are dependent of the B-ISDN operation scenarios, the predominant services and the desirable network load. The variables of the quality of operation functions can be average or accumulated traffic values. The acquisition times of these variables have to be compatible with the time constants of the services and network.

The bit rate allocated function A_j is taken as a nonlinear function of the allocated bit rate to the service class j, with an increasing contribution to quality of operation if the allocated bit rate to that service class does not reach a certain threshold, and with a decreasing contribution if the allocated bit rate is bigger than that value. This allows the dynamic partition of band between service classes.

If a static partition is desired, the sum of the threshold values of the bit rate allocated to each service class j is smaller than the transmission capacity; in this case the bit rate free to be allocated function B_j and the deviation of the calls rejection rate function X_j do not need to be considered $(\beta_j = \chi_j = 0, \forall j)$.

The B_j and X_j functions produce opposite effects, which allows that in heavy traffic situations, the resources allocated to narrowband services do not block the access to broadband services. In reference [6], the proposed solution for this problem is static partition of band.

The quality of service functions express the contribution of the corresponding factors (the cell loss rate Δ_{j0}, the delay Δ_{j1}, and the delay variation Δ_{j2}) to the quality of operation. The values of the Δ_{ji} functions for each service class j, are proportional to the fraction of the bandwidth used by the class.

2.2 Traffic Prediction by Neural Networks

Patterns of the traffic load in a node or link can be collected during the operation of the B-ISDN, in different traffic situations, to be used as learning patterns of neural networks. The delay and the cell loss rate that will be introduced by the call are not known at the time of the call establishment. When the resources of the new call connections are established, the vector of the allocated bit rate of each service class has to be stored and the traffic load pattern is evaluated later, when the new call is generating traffic.

Neural Network inputs are the allocated bandwidth to each service class, and the outputs can be the expected delay, cell loss rate, and the maximum and minimum buffer occupation, the latter leading directly to the delay variation. Another output is included (the number of arrived cells) to allow a better behavior of the training process.

2.3 Call Admission Control

When a connection request for resource allocation to a call arrives to a node, the bit rate requested and the maximum delay and error rate are declared, and the connection is integrated in a suitable service class. The decision if the connection request can be accepted or has to be rejected is based in the network quality of operation expected in each node of the call route.

Each node processor asks the neural network the expected traffic load pattern for the node and the outgoing link, with and without the inclusion of the new connection. The network answers with the expected patterns, then the quality of operation is evaluated in both cases, and the resources are allocated to the call if the expected quality of operation in every B-ISDN node and link of the call route is higher if the new connection was accepted.

2.4 Routing of the ATM Calls

In case where a call has available alternatives routes, every network node and link of all available call paths is inquired about the network quality of operation expected for the call. This allows to discard unacceptable routes.

The cost function of the routing algorithm can be obtained from the quality of operation expected in the nodes and in the links of each call path, with suitable values of the control parameters. For instance, to route calls through paths less loaded, the control parameters of the allocated bit rate have to be set with a small or even null value.

If the number of alternative routes, and the number of nodes of each route, is small, the best path of each call can be found in real time for each call, otherwise the best routing can be determined periodically and all calls within the same time interval follow the established route.

3 Performance Simulation

The B-ISDN components (transmission links and switching nodes), traffic sources and procedures (routing and flow control) are simulated according to the model presented in reference [4], while the ATM traffic is described and simulated by the Markovian model described in reference [3]. Transmission links and switching nodes are represented by delay, error rate, throughput and buffer length. Each ATM traffic source is defined by two Markovian state space processes. The call birth of different services and users are calculated by one Markovian process with different duration and average time between call birth, in each state. The cells of each call are generated, with the appropriate parameters for each service, by the other state space process of the same traffic model. Table 1 shows the peak and the average cell rate and the burstiness of the service classes used in our simulations. The minimum interval between cells is taken equal to the simulation time unit (2.7 μs, for 53 byte cells at 155.52 Mbit/s).

Table 1: Main service characteristics

Service Class (SC)	**Cell Rate (KCell/s)**		**Burst (%)**
	Average	**Peak**	
SC_0	1.604	10.000	13.36
SC_1	3.750	5.000	50.00
SC_2	20.000	20.000	100.00

The prediction of the traffic patterns for each call has been made by a 3 layer neural network with 7 neurons in the hidden layer, and a hyperbolic tangent activation function in the internal neurons. The neural network has been trained with 3500 traffic patterns generated by timing reports produced, with an interval of 5 ms, during previous simulations of the B-ISDN operation. For training the neural network, the backpropagation algorithm was used with adaptive learning rate parameters [5] and the sum of squared errors as cost function.

Figure 1 shows the allocated bandwidth during the 20 seconds of simulation time in one node of a network loaded by calls from the services introduced in table 1. The results are normalized to the node capacity and are shown in four cases: (a) without rejection; (b) allocation based on average cell rate; (c) allocation based on the peak cell rate; (d) allocation based on the proposed technique, with the following values for the control parameters: $\alpha_j = 1.0,\ \delta_{j1} = 0.3$, and $\beta_j = \chi_j = \delta_{j0} = \delta_{j2} = 0.0,\ \forall j$. As seen in the figure if the resource allocation is made without call rejection, an overflow is reached in most of the simulation time. With the allocation by the average and the peak cell rate only the narrowest band service class (narrowest average and narrowest peak, respectively) can access the network resources, namely during the significantly loaded periods. With the proposed technique, the figure shows that all the service classes can share the available resources even when demand is higher.

Table 2 shows the effectiveness of the model to control the delay in the

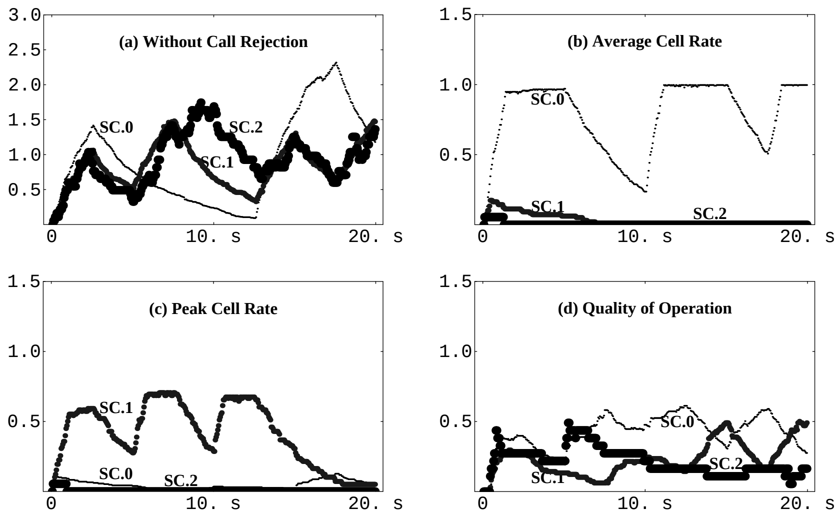

Figure 1: Normalized allocated bandwidth to each service class

network, by adjusting properly the delay control parameter. The output of the simulation is presented in tabular form for nine cases; for each one, the results shown are the normalized maximum allocated bit rate to each service class, and the average and maximum delay, normalized to the maximum buffer occupancy, obtained during the 20 seconds of simulation time. The network was loaded by calls from the services characterized in table 1, and generated with the same statistical distributions of the simulations reported in figure 1. Cases (a), (b), (c) and (d) of table 2 correspond to the simulation presented in figure 1. In cases (a) and (b), the table shows that a full buffer situation is reached (and indeed overflow!); in case (c) no overflow occurs, but the network is lightly loaded due to the peak cell rate allocation; case (d) gives good control on the delay. In case (e) and (f) the control parameter is increased with respect to case (d), which causes a corresponding reduction in the delay. Finally in the last three cases, the delay control parameter is significantly reduced for each of the three service classes. The results show that the network maximum delay found in the node is controlled in case (h) and (i), but not in (g). The explanation for this is the fact that in (g), service class 0 was lightly controlled with respect to the delay, and being a bursty service, it escapes from the network control capabilities. It is noted that other control parameters of the quality of operation function could be effective to control the delay, but in an indirect way, such as cell loss and delay variation.

Table 2: Call control: Allocated bandwidth and delay

Case	**Delay Control Parameter**			**Normalized Max. Alloc. Bandwidth**			**Normalized Delay (%)**	
	δ_{01}	δ_{11}	δ_{21}	$\mathbf{SC}_0$	$\mathbf{SC}_1$	$\mathbf{SC}_2$	**Ave.**	**Max.**
a	0.0	0.0	0.0	2.32	1.48	1.74	0.96	1.00
b	0.0	0.0	0.0	0.99	0.17	0.05	0.53	1.00
c	0.0	0.0	0.0	0.13	0.71	0.05	0.07	0.19
d	0.3	0.3	0.3	0.61	0.50	0.49	0.36	0.85
e	0.5	0.5	0.5	0.52	0.49	0.38	0.26	0.74
f	0.7	0.7	0.7	0.63	0.67	0.33	0.27	0.64
g	0.1	0.7	0.7	0.89	0.31	0.38	0.46	1.00
h	0.7	0.1	0.7	0.79	0.64	0.33	0.28	0.59
i	0.7	0.7	0.1	0.55	0.64	0.16	0.33	0.68

4 Conclusions

The Quality of Operation incorporates the total amount of traffic, the cell loss rate, the delay and the delay variation, as well as the availability of network resources related to the free transmission capacity that can be allocated to each service, and the equilibrium between the call rejection rate of different ATM services. During the operation of the B-ISDN in different traffic situations, patterns of the traffic load in the node or link are memorized to be used as training patterns of a neural network.

When a request for resource allocation to a call arrives to the B-ISDN control entity, the neural networks associated to the B-ISDN nodes and links of the path are inquired about the traffic patterns expected with and without the new connection, and the resources are allocated to the call if the overall quality of operation is expected to increase if the new connection is accepted. Simulation results show that the use of neural networks within this technique lead to a high flexiblity, allowing the control of network parameters according to specific service needs.

References

[1] ., "B-ISDN draft recommendations," in *Rec. I Series*, Geneva: CCITT, May 1990.

[2] T. X. Brown, "Neural networks for switching," *IEEE Communications Magazine*, vol. 27, no. 11, pp. 72–80, Nov. 1989.

[3] J. E. Neves and M. J. Leitão, "A markovian model for ATM traffic generation," in *IEEE-Malaysia International Conference on Communications - MICC 93*, (Kuala Lumpur), Nov. 1993.

[4] J. E. Neves and M. J. Leitão, "Modeling of B-ISDN for performance simulation of ATM services," in *International Symposium on Modular Information Systems and Networks - ICSNET 93*, (St. Petersburg), Sept. 1993.

[5] F. M. Silva and L. B. Almeida, "Acceleration techniques for the backpropagation algorithm," in *Neural Network EURASIP Workshop Proceedings*, (Springer-Verlag), 1990.

[6] E. D. Sykas, K. M. Vlakos, I. Venieris, and E. N. Protonotarios, "Simulative analysis of optimal resource allocation and routing in IBCN's," *IEEE Journal on Selected Areas in Communications*, vol. 9, no. 3, Apr. 1991.

[7] T. Takahashi and A. Hiramatsu, "Integrated ATM traffic control by distributed neural networks," in *XIII Internation Switching Symposium*, vol. III, (Stockholm), May 90.

[8] T. Troudet and S. Walters, "Neural network architecture for crossbar switch control," *IEEE Transactions on Circuits and Systems*, vol. 38, pp. 42–56, 1991.

[9] Yi-Tong, R. Chellapa, A. Vaid, and B. K. Jenkins, "Image restoration using a neural network," *IEEE Trans. on Acoustics Speech, and Signal Processing*, vol. 36, no. 7, pp. 1145–1151, July 1988.

index

Flow control; Rounting; ATM quality of service; ATM call control; Traffic prediction; Adaptive learning;