Connection Admission Control in ATM Networks using Neural Networks: Research Directions and Issues in Commercial Exploitation

Paul Cullen and Brian Carse

Faculty of Engineering
University of the West of England
Coldharbour Lane, Bristol BS16 1QY, UK
pcullen@newbridge.com, Brian.Carse@uwe.ac.uk

Abstract

A number of attempts have been made recently (roughly since 1990) to implement connection admission control (CAC) in ATM networks by means of neural networks. These attempts use various methods and have met with varying levels of success. They all try to solve some very serious networking problems related to the inadequacies of conventional algorithmic computing. Given that there has been little or no adoption of neural network solutions for CAC within the commercial world to date, there are serious questions as to whether or not research scientists are solving the problems which commercial R&D are experiencing. This paper discusses these issues and their relevance to current commercial development.

Introduction

Manufacturers of telecommunications equipment are investing heavily in the research and development of ATM switching elements. This is due to a perceived need of the major network operators in the early 21st Century. At the time of writing there are signs that industry has yet to fully understand the benefits of a dynamic CAC scheme based on intelligent computing methods and that the currently best perceived methods are proprietary ones based on traditional algorithmic techniques. To some extent existing CAC algorithms require a certain amount of manual intervention in setting up the connection's tolerance to the cell loss ratio (CLR). Given that the current level of traffic over existing ATM networks is relatively light, manufacturers have been able to tolerate link occupancy of less than 80% through conventional statistical multiplexing. As traffic increases the problem may be solved in the interim by providing more buffer space at the ingress and egress queues of the network elements. As the next millennium progresses, the size of ATM networks will increase dramatically and sizes in excess of 5,000 nodes are expected. Traffic levels will obviously be higher and manufacturers will be looking for competitive advantage in the way they offer improved link occupancy to network operators. At this time they will more than likely be receptive to the notion of intelligent control methods for CAC.

The ATM Forum

Interoperability of network equipment manufactured by diverse vendors is a prime requirement of network operators. To achieve vendor parity the ATM Forum (Onvural, 1995) was convened to ensure that a timely set of international standards was achieved. The ATM Forum sets de facto standards which are then taken up and modified slightly by such bodies as the International Telecommunications Union - Telecommunications division (ITU-T) and the International Standards Organisation (ISO). To facilitate congestion management, a number of parameters have been defined for the purpose of source characterisation and quality of service (QoS). This paper focuses upon statistical multiplexing which is most applicable to a category of service known as variable bit rate (VBR). VBR caters for both real time (voice and video) and non-real-time (data) traffic. The traffic parameters applicable to VBR services are peak cell rate (PCR), sustainable cell rate (SCR) and maximum burst size (MBS). PCR may be thought of as maximum bit rate while SCR may be thought of as mean bit rate. MBS is the greatest number of cells which may be sent during the SCR measurement period in a single batch. QoS parameters are six in number but only three may be negotiated at the time a connection is made: cell loss ratio (CLR); cell delay variation (CDV); and peak to peak cell transfer delay (ptopCTD). All are relevant to CAC, but the latter two are usually omitted by researchers in favour of CLR. This preference is perhaps puzzling, especially as ATM is expected to carry both real-time and non-real-time traffic. It can only be surmised that the preference for CLR is due to a desire to perfect the transfer of non-real-time traffic first. This coincidentally follows a trend among network manufacturers to provide networks which cater for transfer of computer data. It is easier and more cost effective to concentrate on one challenge at a time.

Defining the CAC Challenge

The function of CAC is to estimate the statistical bandwidth of each signal as it requests a connection to a link. The CAC function then attempts to fit the signal bandwidth into the available link bandwidth. If this available bandwidth is less than the estimated signal bandwidth, the connection is rejected. The most efficient use of available bandwidth occurs as the statistical bandwidth approaches the mean bit rate (SCR) of the source.

QoS has different meanings for different types of traffic. At the time a connection is made a contract has to be negotiated which takes into account the nature of the joining signal. A non-real-time signal will be carrying data which has little immediacy attached to it such as a computer file or a still visual image. It is important to send all of the data with no errors but the delay applied to the data in transit is immaterial. Cell loss ratio is therefore important to non-real-time signals. Real-time data such as telephone or video traffic have different needs. The eye and ear are capable of detecting delays as short as 20ms. It is therefore important that the signal providing audio or video output suffers minimum delay. Lost data is likely to be detected only momentarily by the user who will notice a slight interruption of sound or vision. Real-time signals have more need of low delay and constant delay variation and less need of high cell loss ratio. An ideal multiplexing scheme is achieved when there is 100% link occupancy and the QoS of each connection is maintained.

Selecting Input / Output Parameters for NN Solutions

More traditional algorithmic CAC solutions tend to use the inputs proposed by the ATM Forum while aiming to meet a maximum CLR. The output is less important because that has not been specified by the international bodies. Earlier NN solutions tended to use inputs which appealed to the researchers. These include: outlet circuit occupancy, number of connected calls per type, traffic volume fluctuation per type, mean volume of traffic per type (Ogino and Wakahara 1994); average cell rate, coefficient of cell interarrival times (Aussem 1994); allocated bandwidth, free transmission capacity, connection rejection rate, cell loss ratio, cell delay, cell delay variation (Neves, Leitao and Almeida 1994); number of type 1 links, number of type 2 links, link bandwidth (Hiramatsu 1994). While illustrating the thought and consideration given to solving the CAC problem, the methods using these inputs are unlikely to contribute greatly to future commercial solutions. They simply came too early and without proper knowledge of the way in which the ATM Forum would shape the international standards and influence their adoption.

Table 1 illustrates a number of the early solutions and their diverse system inputs. It has been agreed by Necker, Renger and Kroner (1994) and Youssef, Habib and Saadawi (1995) that it is difficult to measure small levels of cell loss rate in real time given that an acceptable CLR may be as low as 10^{-11}. While Necker, Renger and Kroner (1994) tried to improve upon a conventional method, the Convolution Method, by means of an NN, Youssef, Habib and Saadawi (1995) tried something new. They defined the cell count method as a means of source characterisation. There was, in their minds, good reason for this. They identified that the conventional parameters of PCR, SCR and MBS were inadequate because they were "incapable of representing the burstiness and correlations of multimedia traffic". Subsequently an entire traffic control strategy was proposed on the basis of the cell count method of source characterisation (Tarraf, et al, 1995). This is unfortunate because the ATM Forum members appear to be adamant in their specification of source characterisation parameters and some promising research is very likely to be ignored. The fate of the Hybrid Histogram method is likely to be the same. This scheme (Khalil and Ali, 1995) addresses the problem of adequately representing video traffic. Such traffic has few if any bursts of silence which is in complete contrast to the model of data traffic that is expected to have many silent periods. Although ATM is expected to carry video traffic with similar ease to the way in which it carries data traffic, the ATM Forum parameters are also expected to serve for all traffic types. A more recent approach (Cheng and Chang, 1997) preprocesses the standard parameters, PCR, SCR and MBS using the well known equivalent capacity algorithm (Guerin, Ahmadi and Naghshineh,1991) implemented in fuzzy logic (Cheng and Chang, 1996a) before using its output, the available bandwidth, with system statistics to provide an accept / reject decision. The equivalent capacity algorithm estimates the available bandwidth of the link by means of a summation process which relies upon knowing: the probability of each source being active; the peak bit rate of the connection; average duration of source activity; and the size of buffer available to each source.

Neural Network Training

One of the most challenging aspects of developing NN based systems is the derivation of learning data. Intuitively it may be said that the best data comes from the system to be controlled and the NN may be used to its fullest potential by means of on-line learning. All proponents of on-line learning admit that there is a less than satisfactory convergence time inherent in this method. Hiramatsu (1994) has attempted to overcome the problem of long convergence experienced in his earlier attempt (Hiramatsu, 1990) by using live simulation to speed up the learning process.

A sudden change of network conditions may cause a long convergence delay while new data is learned (Ogino and Wakahara, 1994). Use of small distributed neural networks may also speed up learning time (Aussem, 1994). Table 1 illustrates the large differences that abound in the literature concerning training data. It will be difficult to convince a traditional network element designer to try intelligent methods unless researchers can provide enough detail to allow third party simulation. Table 2 summarises the traffic models employed, comparisons made, and experimental results/conclusions of these works.

A Possible ATM CAC Controller

More recent research (Cheng and Chang, 1997) has produced a possible ATM CAC controller which allows adherence to the ATM Forum source characterisation parameters while using a neural network to achieve improved link utilisation with good QoS performance. It may be said that this is a second generation NN solution as it sets out to improve upon earlier attempts at intelligent CAC such as that proposed in (Hiramatsu, 1990), (Cheng and Chang, 1996a) and (Cheng and Chang, 1996b). Figure 1 (Cheng and Chang, 1997) illustrates the proposed architecture. A call set up request causes the bandwidth estimator, using the equivalent capacity algorithm, to calculate the bandwidth required by the connection. The network resource manager calculates the available bandwidth on the link accounting for all channels on the link. A system statistics calculator produces the cell loss probability from monitoring the state of the buffer queue while the congestion controller computes a value for the congestion status. Congestion on the virtual channel may be controlled by means of varying the coding rate of a video source or varying the transmission rate of all other source types.

A call accept/reject decision is made by the NNCAC controller according to the available bandwidth, the cell loss probability and the congestion status. The NN used in the controller is a three layer feed forward device which is trained to undertake pattern recognition. There are three neurons in the input, seven in the hidden layer and one in the output layer. The NN may be thought of as a universal logic element with eight possible input combinations. It is possible to train the universal logic element to provide a prescribed output in response to a given input pattern. This is achieved by using pre-determined data to train the NN with the back propagation algorithm described by Rumelhart, Hinton and Williams (1986). The number of neurons in the hidden layer is a compromise between the need to train the NN as quickly as possible and the need to cope with comparatively complex input/output patterns. In the case of this application seven neurons in the hidden layer was deemed to be the optimum number. A particularly lucid explanation of the back propagation neural network may be found in (Chester, 1993). Both the bandwidth estimator and the congestion controller are implemented in fuzzy logic (Cheng & Chang, 1996a). Results of simulations show that a link utilisation of better than 90% may be achieved. This occurs through use of a pre-installation off-line learning stage followed by on-line learning to reduce error through changes in the network traffic patterns. The extra level of sophistication provided by the cell loss probability and congestion status parameters also contribute to the increased performance.

Conclusions

It is only recently that CAC solutions have been converging on commercial requirements. Vendor parity is a tremendously important issue for telecommunications equipment manufacturers as deviation from globally accepted standards may seriously affect sales. For this reason a large amount of very good research will be ignored as it does not fit well with the ATM Forum recommendations. A solution is now in reach which helps to improve link utilisation while allowing conventional switch designers to adhere to the ATM Forum parameters. The ideal statistical multiplex of 100% link occupancy with guaranteed QoS may not be achievable in practice, but it has been shown by Cheng and Chang (1997) that an occupancy in excess of 90% may be reached. This is a significant result for future network designers. While the generation of training data remains somewhat a mystery (and therefore difficult for commercial switch designers to embrace the technology), the (Cheng and Chang, 1997) paper and others provide excellent descriptions of the training process and the traffic used for the simulations. Future work in this area will do well to consider the problems as perceived by commercial switch designers and provide solutions which address those problems in a way which may be relatively easily implemented.

Approach/Authors	NN Used	Training	Inputs	Outputs
Cell Transfer State Monitoring. (Ogino & Wakahara 1994)	BPN 10-10-1	1000 data patterns, 100ms learning cycle, on line learning	Outlet circuit occupancy, No. of connected calls per type, traffic volume fluctuation per type, mean volume of traffic per type	Accept/reject threshold (-1 accept, +1 reject).
Queuing System Modelling by Random Neural Network. (Aussem 1994)	Random, 3 element	Learning- distributed on-line.	Average cell rate, coefficient of variation of cell interarrival times	Network occupancy, average end to end cell delay
Quality of Operation. (Neves, Leitao & Almeida 1994)	x-10-y	3500 traffic patterns over 2500 epochs	Allocated BW, free transmission capacity, connection rejection rate, CLR, cell delay, CDV	Comparison between QOO with and without new connection
The Virtual Output Buffer Method. (Hiramatsu 1994)	BPN 3-10-1	On-line from node status, CLR, virtual BW	No. type 1 VCs No. type 2 VCs link bandwidth	CLR
Bitrate Management Using a Recurrent Neural Network. (Necker, Renger & Kroner 1994)	Recurrent	Off-line using backprop with mean square error provided by convolution method	Peak bit rate, mean bit rate	Required bit rate
Equivalent Capacity Refined by Neural Networks. (Ali et al 1994)	BPN 4-x-1	Data set collected from simulations	Peak cell rate, mean cell rate, burstiness, bandwidth	Buffer overflow probability.
The Count Process. (Youssef, Habib & Saadawi 1995)	BPN 20-40-1		1 cell count per frame time per frequency band	Effective bandwidth
Hybrid Histogram/ Neural Network. (Khalil & Ali 1995)	BPN x-y-1	Trained with many samples	Link state vector, CLR	Approx. CLR, or accept/reject.
Dynamic Bandwidth Allocation. (Moh et al 1995)	BPN 1-5-1	200 data elements, 60000 cycles	Bit rate at time n	Bit rate at time n+1
Quality of Service Prediction. (Sarajedini & Chau 1996)	BPN 2-15-1	Unsupervised, 400 samples	No. of users, cell delay	Probability of cell delay exceeding required cell delay
An ATM Traffic Controller. (Cheng & Chang 1997)	BPN 3-7-1	Data manually derived	Available BW, congestion status, cell loss probability	Accept/reject

Table 1. Neural Networks, Training & Inputs/outputs.

Approach/Authors	Traffic	Compared With	Results
Cell Transfer State Monitoring. (Ogino & Wakahara 1994)	Image- AR Model (Ogino & Ikeda 1991), Voice- MMPP, data- constant bit rate	No comparison.	Long convergence time. Threshold changes often required to suit new traffic situations.
Queuing System Modelling by Random NN. (Aussem 1994)	MMPP Five 4x4 switches, 2 I/Ps per switch	Simulation	Network occupancy kept below 80%. Converges within 15 learning cycles.
Quality of Operation. (Neves, Leitao & Almeida 1994)	MMPP	BW Allocation by Ave. cell rate, peak cell rate & equivalent capacity.	QOO gave a fairer BW allocation over time.
The Virtual Output Buffer Method. (Hiramatsu 1994)	MMPP Type 1- PBR= 150Mb/s Type 2- PBR= 1.5 Mb/s	Adaptive CAC	ACAC gave some cell loss and over admission, VOB gave no cell loss or over admission.
Bitrate Management Using a Recurrent NN. (Necker, Renger & Kroner 1994)	MMPP 10Mb/s data, 2Mb/s still picture, 64 kb/s voice	Convolution method	The NN method gives best response time to fluctuations in traffic patterns but can only provide 80% utilisation of link bandwidth.
Equivalent Capacity Refined by NNs. (Ali et al 1994)	MMPP Type 1- 5 sources @ 40Mb/s Type 2- 50 sources @ 4Mb/s	Flow approximation.	For type 1 traffic, flow approximation gives a good correlation to the ideal link usage under heavy load conditions. The NN method gave better performance for all other instances.
The Count Process. (Youssef, Habib & Saadawi 1995)	28 video channels by autoregressive Markov model	Stationary state model	NN method has close correlation to the ideal bandwidth usage.
Hybrid Histogram/ Neural Network. (Khalil & Ali 1995).	Histogram representation of VBR video traffic	Algorithmic histogram method. Effective BW method.	The NN method got best utilisation with 19 calls in a 140Mb/s link. The opposition could only make 10 calls fit the BW.
Dynamic Bandwidth Allocation. (Moh et al 1995)	MMPP	Allocation by mean BW & Ideal allocation characteristic for link utilisation, cell loss & cell delay	Constant buffer length & variable sources: NN scheme closest to ideal. Constant number of sources & variable buffer length: NN scheme closest to ideal but Ave gives lowest delay of all with high Nos. of sources.
Quality of Service Prediction.(Sarajedini & Chau 1996)	MMPP	Other NN methods in general	Gives 3D control surface.
An ATM Traffic Controller. (Cheng & Chang 1997)	Type 1: Delay sensitive video/voice. Type 2: delay insensitive data.	Adaptive CAC (Hiramatsu 1990), FCAC, NFCAC (Cheng & Chang 1996a & 1996b)	NFCAC gives better utilisation but can not match others with CLP of type 1 traffic. NN scheme well behind FCAC & Adaptive for CLP of type 2 traffic.

Table 2. Traffic models, Comparisons made & results.

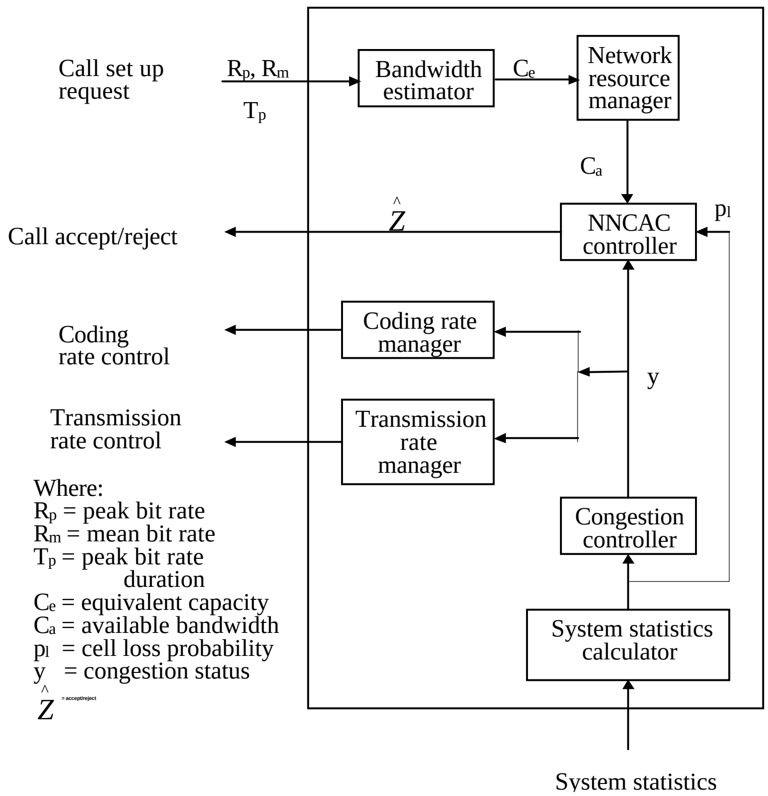

Figure 1. A Possible ATM CAC Controller (Cheng and Chang, 1997)

References

Ali, Borhanuddin Mohd; Khalil, Ibrahim; Mukerjee, MR; Bidin, Abdul Rahman 1994. Call Admission Control in ATM Networks Using Neural Networks. Proceedings of the 6th IASTED/ISMM International Conference on Parallel Distributed Computing Systems 480 - 483.

ATM Forum 1996, Traffic Management Specification Version 4.0.

Aussem, Alex 1994. Call Admission Control in ATM Networks with the Random Neural Network. IEEE International Conference on Neural Networks World Congress on Computational Intelligence. 2482 - 2487.

Cheng, R-G; Chang, C-J 1997. Neural Networks Connection Admission Control for ATM Networks. *IEE Proceedings* - *Communications* Volume 144 No.2, 93 - 98.

Cheng, R-G; Chang, C-J 1996a. Design of a fuzzy traffic controller for ATM networks. *IEEE/ACM Trans Netw*., 1996, 4, (3), 460-469.

Cheng, R-G; Chang, C-J 1996b. A neural net based fuzzy admission controller for an ATM network. IEEE INFOCOM '96, 777-784.

Chester, Michael. 1993. Neural Networks: A Tutorial. New Jersey. Prentice Hall.

Guerin, R., Ahmadi, H. and Naghshineh, M. 1991. Equivalent Capacity and its application to bandwidth allocation in high-speed networks. *IEEE JSAC*, Vol.9, 968-981.

Hiramatsu, Atsushi 1994. ATM Call Admission Control Using a Neural Network Trained With a Virtual Output Buffer Method. Proceedings of the IEEE International Conference on Neural Networks. IEEE World Congress on Computational Intelligence, 3611 - 3616.

Hiramatsu, Atsushi 1990. ATM communications network control by neural networks. *IEEE Trans Neural Networks*, 1990, 1, (1), 122-130.

Khalil, I; Ali, B Mohd 1995. A Hybrid Histogram and Neural Based Call Admission Control for VBR Video Traffic. 4th International Conference of Artificial Neural Networks, 421 - 426.

Moh, W Melody; Chen, Min-Jia; Chu, Niu-Ming; Liao, Cherng-Der 1995. Traffic Prediction and Dynamic Bandwidth Allocation over ATM: A Neural Network Approach. *Computer Communications* Volume 18 No. 8, 563 - 571.

Necker, Thomas; Renger, Thomas; Kroner. Hans 1994. Bitrate Management in ATM Systems Using Recurrent Neural Networks. Proceedings of the 1994 IEEE GLOBECOM Communications: The Global Bridge Volume 3, 1774 - 1779.

Neves, Joaquim E; Leitao, Mario J; Almeida, Luis B 1994. B-ISDN Connection Admission Control and Routing Strategy with Traffic Prediction by Neural Networks. Proceedings of Supercom/ICC '94 Serving Humanity Through Communications. New Orleans, USA. Volume 2, 769 - 773.

Neves, Joaquim E; Leitao, Mario J; Almeida, Luis B 1995. Neural Networks in B-ISDN Flow Control: ATM Traffic Prediction or Network Modelling*? IEEE Communications Magazine* October 1995, 50 - 56.

Ogino, Nagao; Wakahara, Yasushi 1995. Application of Neural Network in ATM Call Admission Control Based on Cell Transfer State Monitoring with Dynamic Threshold. *IEICE Transactions in Communications* Volume E78-B No.4, 465 - 475.

Onvural, Raif O 1995. Asynchronous Transfer Mode Networks - Performance Issues 2nd Ed. Boston. Artech House Publishers.

Rumelhart,D.E., Hinton, G.E., and Williams, R.J. 1986. Learning internal representation by error propagation. Parallel Distributed processing: explorations in the microstructure of cognition. Vol 1, Chap. 1. Cambridge. MIT Press.

Sarajedini, Amir; Chau, Paul M 1996. Quality of Service Prediction Using Neural Networks. Conference Proceedings of MILCOM '96 Volume 2, 567 - 570.

Sarajedini, Amir; Chau, Paul M 1996. Cumulative distribution estimation with Neural Networks. Submitted to the World Congress on Neural Networks.

Skelly, Paul; Schwartz, M; Dixit, Sudhir 1993. A histogram based model for video traffic behaviour in an ATM multiplexer. *IEEE/ACM Transactions on Networking* Vol. 1, No. 4, 329-343.

Tarraf, Ahmed A; Habib, Ibrahim w; Saadawi, Tarek N 1995. Intelligent Traffic Control for ATM Broadband Networks. *IEEE Communications Magazine* Oct 1995, 76 - 82.

Youssef, Sameh A; Habib, Ibrahim W; Saadawi, Tarek N 1995. ATM Call Admission Control Using Neural Networks. MILCOM '95 Universal Communications Conference Record Volume 1, 1 - 5.