FUZZY CONNECTION ADMISSION CONTROL FOR SELF-MANAGED NETWORKS

Raj Srinivasan and Jérôme Talim
University of Saskatchewan, Canada
Carleton University, Canada
raj@math.usask.ca, jtalim@sce.carleton.ca

Abstract

In this paper, we present an algorithm for Connection Admission Control (CAC) in a high speed data network. Our algorithm uses on-line measurement and off-line simulation to estimate the amount of bandwidth required to meet the request of a set of active sources and the actual use of the bandwidth. We introduce the concepts of Customer Satisfaction Factor (CSF) and Proportion of the Used Bandwidth (PUB) and identify three domains for the CAC namely, strong acceptance (with guarantee), weak acceptance (without guarantee) and rejection. The algorithm presented here is computationally efficient and does not assume any detailed statistical characteristics of the traffic sources. Sources need to declare only their mean and peak rates.

Keywords: connection admission control; fuzzy logic; quality of service; bandwidth allocation.

1. INTRODUCTION

Modern telecommunications networks are expected to transport diverse sources of traffic such as voice, data, video on demand and multimedia. Designing and managing such a single network that can perform as well as the public switched telephone network can be very challenging. Each traffic type requires different amount of network resources (buffer space and bandwidth), and its flow through the network is affected by network control functions such as queueing mechanism, bandwidth allocation, and routing protocols. Efficient and easily implementable network management algorithms are required to properly guarantee Quality of service (QoS) to all types of traffic and at the same time reduce congestion, maximize network utilization and ensure that the network is stable. The self-x framework we introduce here for the basic network management functions is part of the vision of a self-x network that can manage its functions itself without requiring human interactions. It is expected that such a network will transport diverse sources of traffic such as voice, data and video on demand, while guaranteeing individual and specific QoS to each user. It is also expected that such a network is self-installing, self-learning, self-sizing and self-healing. In order to achieve these self-aspect goals, the work should possess: (i) the self-knowledge which is typically derived from accurate on-line measurements and current network conditions, and (ii) the self-learning abilities to adapt to changes in traffic and network conditions. The self-x framework we propose here consists of three modules and addresses separately three functions namely, traffic prediction, traffic scheduling and connection admission control. See Figure 1. This paper provides the complete description of the CAC module.

In general, Connection Admission Control (CAC) can be defined as the procedure of deciding whether or not to accept a new connection. One of the fundamental aspects of CAC is to evaluate the impact of a new connection on the current traffic load. This usually involves

---

Proceedings of the 2002 IEEE Canadian Conference on Electrical & Computer Engineering
0-7803-7514-9/02/$17.00 © 2002 IEEE

determining the resources (bandwidth and buffer space) needed for a new connection with its specific quality of service (QoS) in the presence of existing connections. If the traffic generated by each connection is a deterministic function of the time, then the CAC procedure could consist of simulating off-line the traffic forwarding mechanism and in estimating the level of service provided to each. In practice, most of the time, the traffic is highly variable and can only be expected to provide its statistical characteristics. The notion of the effective bandwidth of a traffic source has been used recently to describe the effective resource requirements. This procedure is based on large deviation theory and depends strongly on the statistical properties of the traffic sources. See [3],[4],[5], [6],[7]. An alternative procedure using entropy estimators based on on-line measurement is given in [8]. Most of the methods require accurate statistical characteristics of the sources that may not always be feasible to obtain in practice. Besides, sources are assumed to be active indefinitely.

In order to design an adaptive CAC, we introduce the notion of ***Virtual Line Rate Required (VLRR)***, an estimator of bandwidth (See Figure 2). For a set of sources, we use the virtual traffic, as opposed to the actual traffic. The virtual traffic is either the predicted traffic (that is calculated by the Fuzzy Traffic Predictor, see[1]), or the approximated traffic. For a given mean rate m and a peak rate P, the approximated traffic is a sequence of random numbers between 0 and P, generated according to a uniform distribution such that the mean of the approximated traffic is close to m. The virtual traffic is then forwarded to a virtual channel using the Fuzzy Traffic Scheduler ([2], [9]) whose capacity can be fixed.

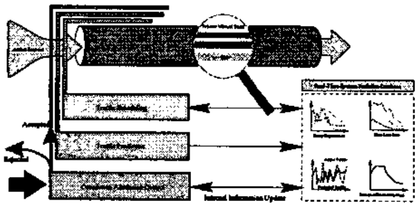

Figure 1: Main Modules of the Self-X Architecture

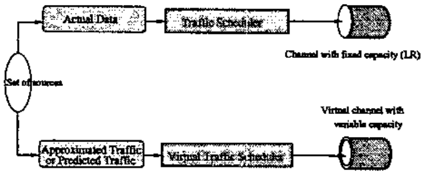

Figure 2: Virtual Line Rate Required

2. FUZZY CAC (fCAC)

In this section, we describe the CAC algorithm, implemented in the Self-X architecture. We consider a channel with a fixed capacity, denoted by LR. Throughout this section, this quantity is a constant. This algorithm integrates two main criteria:

1. The estimation of the bandwidth required to meet the request of a set of active sources; this ensures that the system will satisfy its traffic agreements.
2. The estimation of the actual bandwidth used; the goal consists in maximizing the actual use of the bandwidth, while guaranteeing the QoS requested.

We present the two aspects of the CAC structure in detail in the next sections.

2.1 Virtual Line Rate Required and Customer Satisfaction Factor

In a multiservice environment where the system guarantees a specific level of service to each user, the resources must be allocated appropriately. The Customer Satisfaction Factor (CSF) is defined as the proportion of customers whose service request is satisfied. When the system has "enough" bandwidth to process the traffic from a set of active sources, the CSF is large, ideally equal to 1. On the contrary, when the system is over-using its resources, it hardly meets its commitment. Since the VLRR is an estimation of the bandwidth needed, intuitively, one can conjecture that the CSF is decreasing when the VLRR is increasing.

Let Y be

$$Y := \frac{1 - CSF}{\frac{VLRR}{LR}},$$

with distribution function

$$F_Y(y) := P\,(Y \leq y)\,.$$

The quantity $1 - CSF$ can be interpreted as the proportion of customers who are unsatisfied with respect to their QoS request. For a set of sources, characterized by their traffic, and their QoS requests, we calculate the associated VLRR when considering the virtual traffic, and the CSF based on the actual traffic for a channel with a capacity of LR. We then estimate numerically the logarithm of the distribution function F_Y. See Figure 3(a). Using linear regression, we can determine two constants a_Y and b_Y to approximate the function:

$$\log(F_Y(y)) \sim a_Y + b_Y\ y.$$

For some relevant $\epsilon > 0$ there exists $y(\epsilon) > 0$ such that

$$a_Y + b_Y\ y(\epsilon) = \log(\epsilon),$$

which yields

$$F_Y(y(\epsilon)) \leq \epsilon \Longleftrightarrow \\ P\left(1 - CSF > y(\epsilon)\ \frac{VLRR}{LR}\right) < 1 - \epsilon.$$

For an appropriate ϵ, and two arbitrary bounds LB_1 and UB_1 (for instance $LB_1 = 30\%$ and $UB_1 = 60\%$), we can derive the following algorithm :

- IF $y(\epsilon)\ \frac{VLRR}{LR} > UB_1 = 60\%$ THEN A large proportion of customers (60%) is expected to be unsatisfied. The system may not have then enough bandwidth.
- ELSE IF $y(\epsilon)\ \frac{VLRR}{LR} > LB_1 = 30\%$ THEN There could be a reasonable proportion ([30%...60%]) of unsatisfied customers. The system may not satisfy the QoS at bursty periods.
- ELSE There should be less than 30% unsatisfied customers. The system should have enough bandwidth and guarantees the QoS.

This algorithm is part of the Fuzzy CAC implemented in the self-x project. We will present the structure of the CAC in section 2.3 .

2.2 Use of Bandwidth and Users Requirements

For a set of sources, we evaluate the Proportion of Used Bandwidth (PUB) when processing the actual data through the channel. Besides, using the source characterization, we define the coefficient of User Requirement as:

$$UR_i := \frac{\text{Peak rate}_i}{\text{Mean Rate}_i} \cdot QoS_i,$$

and

$$UR := \frac{1}{N} \sum_{i=1}^{N} UR_i,$$

for a set of sources $i = 1 \ldots N$. One can observe that $UR \geq 1$. Qualitatively, the larger the ratio $\frac{\text{Peak rate}_i}{\text{Mean Rate}_i}$ is, the more irregular the source is and a larger amount of resources is more likely to be used. The ideal situation for $\frac{\text{Peak rate}_i}{\text{Mean Rate}_i} = 1$ corresponds to a source with a constant traffic rate. For two sources characterized by the same ratio $\frac{\text{Peak rate}}{\text{Mean Rate}}$, the one with a greater QoS will be provisioned with a larger proportion of resources. Thus, the coefficient UR is an indicator of the amount of bandwidth that is required by the sources. One can conjecture that the Proportion of Used Bandwidth is increasing when UR is increasing.

Let Z be the random variable

$$Z := \frac{1 - PUB}{UR} < 1,$$

with its distribution function

$$F_Z(z) := P\,(Z \leq z)\,.$$

Note that $1 - PUB$ corresponds to the proportion of unused bandwidth. We determine numerically the function $\log(F_Z)$ which is depicted in the Figure 3(b). One can note that for Z large enough, $\log(F_Z)$ is linear. There exist two constants a_Z and b_Z such that:

$$\log(F_Z(z)) \sim a_Z + b_Z\ z,$$

for Z "large enough". Recall that $UR \geq 1$ and $1-PUB < 1$. When Z is small and close to 0, the quantity UR is large. The system is then dealing with irregular sources that request high QoS's. It is expected to use fully the bandwidth and it may not satisfy all the QoS's requested.

For some $\eta > 0$ there exists $z(\eta) > 0$ such that

$$a_Z + b_Z \ z(\eta) = \log(\eta).$$

One can observe

$$F_Z(z(\eta)) \leq \eta \Longleftrightarrow$$
$$P\,(1 - PUB > Z(\eta) \cdot UR) < 1 - \eta.$$

For an appropriate η, and for two arbitrary bounds for instance $LW_2 = 30\%$ and $UB_2 = 60\%$, the previous inequality yields the following algorithm:

- IF $z(\eta) \cdot UR > UB_2 = 60\%$ THEN there may be a large proportion (60%) of bandwidth that is unused.
- ELSE IF $z(\eta) \cdot UR > LB_2 = 30\%$ THEN Large amounts (30%) of the bandwidth could be unused at periods.
- ELSE The bandwidth is likely to be fully used.

The two bounds (LB_2 and UB_2) can be a function of the current use of the total bandwidth. We will fully discuss the adaptability of the self-x architecture in the next section.

2.3 Fuzzy CAC algorithm

As mentioned previously, the CAC function deals with (i) the evaluation of the amount of available bandwidth to process the traffic from the accepted sources, and (ii) the way the bandwidth is actually used by the traffic. On the one hand, the system aims to guarantee the quality of service to each customer; on the other, it tends to maximize the use of the bandwidth. The fuzzy CAC will integrate the two aspects, or the two algorithms presented in the previous section. See Figure 4. A situation is described by its set of sources, with their characteristics. Those characteristics are used to calculate the Virtual Line Rate Required (VLRR), the User Requirement (UR) coefficient. We define then three domains of all the possible situations:

- Domain of Acceptance with Guarantee (D1):

  When the VLRR is small (i.e. $VLRR/LR \leq LB_1$) as well as the UR (i.e. $UR \leq LB_2$), the system is facing a situation where there should be enough bandwidth and where the bandwidth is likely to be used consistently.

  The system should be in this domain as much as possible.

- Domain of Rejection (D3)

  When the VLRR or the UR is large (i.e. $VLRR/LR > UB_1$ or $UR > UB_2$), the system expects either not to have enough bandwidth, or not to use completely the bandwidth. In other words, either it cannot guarantee the QoS's, or it is going to waste too much resources.

  The system should avoid being in this domain as much as possible.

- Domain of Acceptance without Guarantee (D2):

  Any situation between the two extreme domains presented previously deals with situations where occasionally the system may not meet the requests (in particular at bursty periods) or may waste large proportion of bandwidth.

Upon an arrival of a new source, the system considers {Active Sources} ∪ {New Source}. For this set, the system determines the domain D1, D2 or D3 to which the new situation belongs. The new source will be accepted with guarantee if it falls in the domain D1 or it will be accepted without guarantee if it falls in the domain D2. Otherwise, the new source will be rejected. Sources accepted with (respectively, without) guarantee join the First (respectively, Second) Class sources.

One can note that D1, D2 and D3 are functions of the bounds (LB_1, UB_1, LB_2, and UB_2). This ensures that the self-x network can adapt its behavior to changes in traffic or workload; the bounds can be functions of the current use of the bandwidth. Thus, whenever the system is

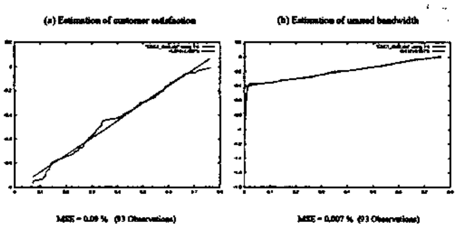

Figure 3: Estimation of Customer Unsatisfaction and Unused Bandwidth

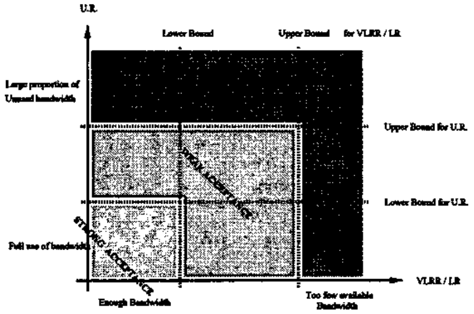

Figure 4: Structure of the *fuzzy* CAC

underusing (respectively, overusing) the bandwidth, the bounds can be increased (respectively, decreased) in order to utilize the bandwidth optimally while guaranteeing the QoS requests.

3. NUMERICAL EXAMPLE

In order to illustrate the efficiency of the Fuzzy CAC, we present the following example. The average of the arrival rates of the accepted sources (either at first or second class) are depicted in the Figure 5. In this example, the sources are heterogeneous (arrival rates can vary between several bytes to 5000 bytes). They start their activity at random moments and remain active for a random amount of time. Their QoS request is generated randomly and lies between 3 and 7, which means that the Loss Rate requested is between 10^{-7} and 10^{-3}. The Figure 6 provides the basic statistics of the example. There are 26 sources requesting service, only 15 are accepted. All first class sources are satisfied while only half of the second class sources have their QoS request satisfied. The graph in Figure 6 represents the instantaneous proportion of bandwidth that is unused.

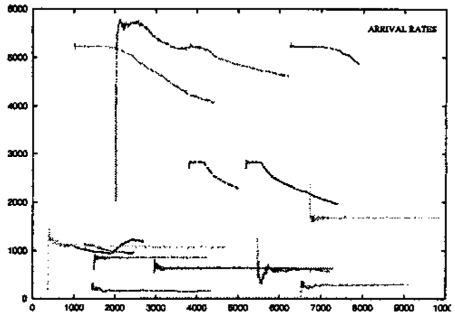

Figure 5: Average of the arrival rates of accepted sources

Total number of users considered .... : 26
Number of users accepted in First Class .... : 7 (100% Satisfied)
Number of users accepted in Second Class............... : 8 (50% Satisfied)
Number of users rejected.. : 11
Line Rate = 15,000

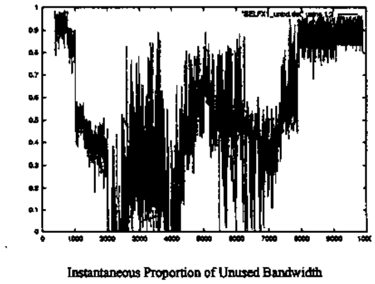

Figure 6: Statistics of the example

4. CONCLUSIONS

In this paper, we presented a computationally efficient algorithm for CAC which is based on on-line measurements and off-line simulation. Initial simulation results indicate that it provides a good trade off between users demand and the actual use of the bandwidth. The boundaries between strong acceptance, weak acceptance and rejection domains can be adaptively adjusted to reflect the changes in traffic and network conditions. The adaptive nature of our algorithm is well suited for the vision of the self-x network.

Acknowledgements

This work was funded from a grant from the Network of Centres of Excellence, MITACS and Nortel Networks. We would like to thank Don Dawson (Carleton University), James Yan and Tadeusz Drwiega (Nortel Networks) for their helpful insights throughout this project.

References

[1] D. Dawson, Q. Jiang, D. Slonowsky and R. Srinivasan "Measurement Based Adaptive Traffic Prediction Using Fuzzy Logic," *Progress Report for the MITACS Project*, Nov. 1999

[2] D. Dawson, Q. Jiang, D. Slonowsky and R. Srinivasan "Adaptive Bandwidth Allocation Using Fuzzy Logic Control and the Fuzzy Traffic Predictor," *Progress Report for the MITACS Project*, Jan. 2000

[3] A. Elwalid and D. Mitra "Effective Bandwidth of General Markovian Traffic Sources and Admission Control in High Speed Networks," *IEEE/ACM Transactions on Networking*, vol. 1, no. 3, pp. 329-343, June 1993.

[4] R.J. Gibbens, F.P. Kelly and P.B. Key "A Decision-Theoretic Approach to Call Admission Control in ATM Networks," *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 6, pp. 1101-1113 August 1995.

[5] F.P. Kelly "Effective Bandwidths at Multi-Class Queues," *Queueing Systems*, vol. 9, pp. 5-16, 1991.

[6] E. Buffet and N.G. Duffield "Exponential Upper Bounds via Martingales for Multiplexers with Markovian Arrivals," *Journal of Applied Probability*, vol. 31, pp. 1049-1061, 1994.

[7] Z. Liu, P. Nain, and D. Towsley "Exponential Bounds with an Application to Call Admission," *Journal of the ACM*, vol. 44, no. 2, 1997

[8] J.T. Lewis, R. Russell, F. Toomey, B. McGurk, S. Crosby and I. Leslie "Practical Connection Admission Control for ATM Networks Based on On-line Measurements," *Preprint*, Jan 1998.

[9] R. Srinivasan and J. Talim "Integrating Network Congestion Status and Individual Quality of Service in Dynamic Bandwidth Allocation," *Preprint*, Jan. 2002.