A Congestion Control Method for Real-Time Communication Based on ATM Networks

Lichen Zhang
Faculty of Computer Science and Technology
Guangdong University of Technology , 510090 Guangzhou
GuangDong Province, P.R. of China
Email: lchzhang@gdut.edu.cn

Abstract: In this paper, we present results of a study of congestion control for real-time communication based on ATM networks. In ATM networks, congestion usually results in cell loss. Based on the time limit and priority, the cells that compete for the same output line could be lined according to the character of real-time service. We adopt priority control algorithm for providing different QoS bearer services that can be implemented by using threshold methods at the ATM switching nodes, the cells of different deadline and priority could be deal with according to the necessity. Experiments show the proposed algorithm is effective in the congestion control of ATM real-time networks

1. Introduction

ATM is proposed as the next generation global telecom network. This network should support a number of communications services, including phone calls, video conferencing and computer communication. ATM is also used for a Local Area Networks (LANs). However, providing integrated services in high speed store-and-forward networks like ATM is difficult because of the wider range of traffic patterns and quality of service (QoS) requirements to support. Real-time communication services such as video & audio conferencing, video-on-demand, and remote medical services in an integrated network pose serious challenges in meeting their stringent QoS requirements such as bounded cell-delivery delay and cell-loss ratio, while handling the burstiness of their traffic. Real-time communication can be classified into two categories according to QoS requirements: deterministic and statistical. In the former, QoS requirements are specified in deterministic terms and no cell losses or deadline misses are allowed. In order to satisfy its absolute QoS requirements, each deterministic real-time connection requires resource reservation based on the worst-case source traffic-generation behavior, thus resulting in severe underutilization of network resources when source traffic is bursty. In order to make more efficient use of network resources, statistical real-time communication specifies QoS requirements in statistical (instead of deterministic) terms, thus tolerating a certain percentage of cell losses and deadline miss. Such a specification allows for overbooking network resources and, at the same time, enhancing the multiplexing gain. Statistical real-time communication is useful to those applications(i) that can tolerate a portion of cell losses and deadline misses and (ii) whose traffic is bursty. The statistical multiplexing gain is substantial, especially in Variable-Bit-Rate (VBR) applications such as MPEG-coded video.

Usually the most effective way to counteract congestion is to mainly base on preventive approaches such as call admission control algorithm and reduces the traffic at the entry points to the network. This can be achieved by feedback signaling from the network to the sources of information. The general idea is to adjust traffic flow into the network in such a way as to optimize network performance. A number of rate-based control schemes have been proposed [1][2][3][4]. Yet the ATM equipment providers aim at the network utilization and the increasing demand of large capacity. In a commercial network where pricing is directly related to bandwidth reservation, such a low utilization is highly undesirable. There is no way to shun the congestion control problem. When more than one cell competes for the same out-port, the congestion arbiter will queue the cells according to the priority and time limit of each cell. Then the most important cell will be transmitted and the least important cell could be discarded. Others will be add time delay then wait for retransmission or veer to a new router.

ATM implementations adopt the leaky bucket approach to avoid congestion. The bucket is a number of counters that are maintained at the UNI network side for each connection. A token generator periodically issues values named tokens that are placed into a token pool. Each cell should get a token before it was connected. When cells are sent to the network, the token pool is exhausted. In such a situation, the cell will be discarded or dropped to the UNI without further processing. Consider real time upsurge traffic arrives when the token pool is exhausted, it is obviously that the cell will be discarded. If the cell is important and can't be retransmitted because of time limit, a disaster wound happen.

In this paper, we adopt priority control algorithm for providing different QoS bearer services can be implemented by using threshold methods at the ATM switching nodes. In the separate route approach, buffer management is not required at the switching nodes, since priority processing is executed at the connection level by the routing function. Priority disciplines in queuing theory can be categorized into two major types: 1). Service or delay priority disciplines, which govern the time at which cells in the buffer are transmitted. 2) Buffer access or space priority disciplines, which govern the input access, or space priority disciplines, which govern the input access of cells into the buffer.

The rest of this paper is organized as follows: Section 2 outlines the characteristics of common ATM real-time

services. Section 3 describes the model of our method. Lastly, we conclude the paper in section 4.

2. ATM Real-Time Service

2.1 ATM services

Data traffic services on ATM will be using the available bit rate(ABR) type of service defined by the ATM Forum. Network bandwidth left over after handling the real-time traffic such as the constant bit rate (CBR) and The variable bit rate (VBR) will be used by ABR or UBR (unspecified bit rate) types of service. End users connected by ABR services will be guaranteed a minimum cell rate (MCR), a peak cell rate (PCR) and a cell loss rate (CLR) that are specified during the establishment of the connection. An ABR source may transmit at a bit rate range between MCR and PCR under the condition that the negotiated CLR is satisfied. However, an ABR type of connection should be delayable and flow controllable. An ABR source may be forced to adjust its transmission rate according to the control signals sent from the network. It is recommended that the ABR service is good for high-performance, non-real-time data applications such as hypermedia document access, file transfer, image transfer, etc. UBR services are often referred to as the best-effort type of services and they require no guaranteed transmission rate, for example, the connectionless datagram services like email. Spare bandwidth left over after processing QoS guaranteed services will be used by UBR under the condition that the service quality of the throughput-or-guaranteed flow is not violated.

Congestion is defined as a condition that exists at the ATM layer in the network element such as switches, transmission links or the cross connects where the network is not able to meet a stated and negotiated performance objective.

Traffic control defined a set of actions taken by the network to avoid congestion. Traffic controls measure to adapt to unpredictable fluctuations in traffic flows and other problems within the network. Traffic control mechanics allot the limit resource according to the QoS of services.

The main purpose of common traffic control is to allot resource and reject excessive connection request. ATM traffic control mechanic assigns the resource to the request connection after it was accepted. The resource can't be shared till the connection is deleted. If there is no resource left, new connect request will be rejected no matter how important it is. It has been frequently claimed that the priority scheduling approach has superior "stability" compared with other approaches, because "essential" processes can be assigned high priorities in order to ensure that they meet their deadline.

2.2 Congestion Control

The main function of an ATM switch is routing. Considering the arrival of upsurge services, cells from different in ports need to be exchanged to the same out port at the same time. There will happen a competition. If the congestion happens in the internal of a switch, we call that internal congestion. Since the transport speed of ATM networks is very high, the congestion tends to be

deterioration and the network is forced to paralyze. It's a big problem for vendors to think of it.

The purpose of common traffic control is to allot resource and reject excessive connects request. ATM traffic control mechanic assigns the resource to the request connects after it was accepted. The resource can't be shared till the connection is deleted. If there is no resource left, new connect request will be rejected no matter how important it is. It has been frequently claimed that the priority scheduling approach has superior stability compared with other approaches, because "essential" processes can be assigned high priorities in order to ensure that they meet their deadline.

Usually the most effective way to counteract congestion is to mainly base on preventive approaches such as call admission control algorithm and reduces the traffic at the entry points to the network. This can be achieved by feedback signaling from the network to the sources of information. The general idea is to adjust traffic flow into the network in such a way as to optimize network performance. A number of rate-based control schemes have been proposed and good reviews are given in ref's [1] [2][3][4]. Yet the ATM equipment providers aim at the network utilization and the increasing demand of large capacity. In a commercial network where pricing is directly related to bandwidth reservation, such a low utilization is highly undesirable. There is no way to shun the congestion control problem. Fig.1 shows the procedure method of congestion control. When more than one cell compete for the same out-port. The congestion arbiter will queue the cells according to the priority and time limit of each cell. Then the most important cell will be transmitted and the least important cell could be discarded. Others will be add time delay then wait for retransmission or veer to a new router.

ATM implementations adopt the leaky bucket approach to avoid congestion. The bucket is a number of counters that are maintained at the UNI network side for each connection. A token generator periodically issues values named tokens that are placed into a token pool. Each cell should get a token before it was connected. When cells are sent to the network, the token pool is exhausted. In such a situation, the cell will be discarded or dropped to the UNI without further processing. Consider real time upsurge traffic arrives when the token pool is exhausted; it is obviously that the cell will be discarded. If the cell is important and can't be retransmitted because of time limit, a disaster wound happen. Here we adopt priority control algorithm for providing different QoS bearer services can be implemented by using threshold methods at the ATM switching nodes. In the separate route approach, buffer management is not required at the switching nodes, since priority processing is executed at the connection level by the routing function. Priority disciplines in queuing theory can be categorized into two major types: 1). Service or delay priority disciplines, which govern the time at which cells in the buffer are transmitted. 2) Buffer access or space priority disciplines, which govern the input access, or space priority disciplines, which govern the input access of cells into the buffer.

3. Model Based on Multi-Markov

ATM network traffic control model is usually built according to the type and usage of services. In general, the model of non real-time services is based on Poisson Model. Real time services and data flow is built as MMPP Method or Multi MMPP method. Considering the cells of the same connection can have different priorities but same demand of discard/accept sequence. The vector C_i (t_d, t_0, p_i) denotes the request priority of cell i, t_0 indicates the start time, t_d means the time deadline, so $|t_d - t_0|$ means time left for a cell to be sent; we take p_i as the priority of this cell. So the time limit and the priority can be considered at the same time.

For the input queued cells, we use a threshold buffer with finite size K (cells) for each outgoing route to switching matrix. The queuing analysis can be focused on a buffer associated with any particular outgoing link. Assume there lies a discrete-time slotted queuing system in which each slot duration is equal to an ATM cell transmission time and is assumed to be of unit length. Cells start transmission only at the start time of a slot. Let the ratio that total cells of an input line to an output line be Y_{ij} there will be:

$$\sum_{j=0}^{N} Yij = 1, N = total.cells$$

Let p_i be the probability of the cell i in the buffer at the end of a time slot. Let $\prod_{ij}$ denote the transmission probability of a cell to be transmitted from the input line i to the output j. As for the total cells, we get the following formula. So the priority and deadline of each cell are considered at the same time. The ATM arbiter could classify the cells according the QoS of each connection.

$$\Pi = \begin{bmatrix} C_0 & C_1 & \cdots & C_n & 1-\sum_{i=0}^{K}C_i \\ & Y_{11}C_0 & \cdots & Y_{11}C_{n-1} & 1-\sum_{i=0}^{K-2}\frac{p_i}{|t_d-t_0|}C_i \\ & & \cdots & & \\ & & & Y_{1n-1}C_1 & 1-\sum_{i=0}^{K-2}\frac{p_i^{K-1}}{|t_d-t_0|}C_i \\ & & & & 1-\sum_{i=0}^{K-2}\frac{p_i^{K}}{|t_d-t_0|}C_i \end{bmatrix}$$

$$Y_{ij} = \frac{p_i^i}{|t_d - t_0|}, i=0, 1, ..., K$$

For example, if there are 2 different kinds of service compete for an output line, one service has high priority while the other has low priority but limit time deadline. Thus the arbiter will compute the possible cell loss rate and decide to hold the low priority cell or the high priority cell.

According to the loss probability of each cell, all the cells could be stamped and dealt well. From the probability

matrix, the condition of all the cells can be divided clearly. So we can transmit the urgent cells, add time-delay to the important cells so that they could be transmitted later. If necessary, some less important cells are discarded to avoid congestion. With this mechanic, the arbiter work well.

4. Experiments and Conclusion

We build the experiment environment as follows: Different real-time services such as voice, video, real-time compression data or fax are adopted by the Input Control line i and exchanged to the Output Control line j. After clock recovery and SAR (segment and resemble), they are sent to an ATM Switch to exchange to the correct route. To emulate the congestion circumrotation, we set the head of those service cells in the same VPI/VCI. We use cell loss detection machine at the Output Control line to check the cell loss rate. Fig.3 shows the buffers size and the cell loss rate. From Fig3 we know that the cell loss rate is linked to the buffer size. When we queue the cells according to the model discussed in Section 3, the cell loss rate improves greatly. This method proved to be effective in cell loss at the cost of a tolerant degradation of the behavior of the network.

This paper proposes a theoretical analysis of congestion control mechanic. The cell loss probability performance of the buffer threshold scheme to implement the priority assignment arbiter is also discussed. According to this algorithm, we can order all the cells that waiting for transimission. Thus cells of different deadline and priority could be deal with according to the necessity. The ATM network could offer the best QoS. Experiments show that this mechanic is effective in ATM congestion control. Future work will concentrate on UPC control and CAC algorithm to assign the traffic of each channel.

References

[1] A.Ermedahl, et al., Response time guarantees in ATM networks, Proceedings of the 18th IEEE Real-Time Systems Symposium, 274-284, IEEE Computer Society Press, 1997.

[2] A.E.Eckberg, et Al. Meeting the Challenge Congestion and Flow Control Strategies for Broad Band Information Transport. GLOBERCOM'89, 1989, pp 1769-1773

[3] J.K.Ng et al., Integrated delay analysis of regulated ATM Swith, Proceedings of the 18th IEEE Real-Time Systems Symposium, 285-296, IEEE Computer Society Press, 1997.

[4].Saunders. ATM Forum Ponders Congestion Control Options, Data Communications, 1994, Mar, pp 55-60.

[5]Chung G.Kang, Harry H.Tan, Queueing analysis of explicit priority assignment buffer access scheme for ATM networks, Computer Communications21 (1998), ELSEVIER, pp 996-1009
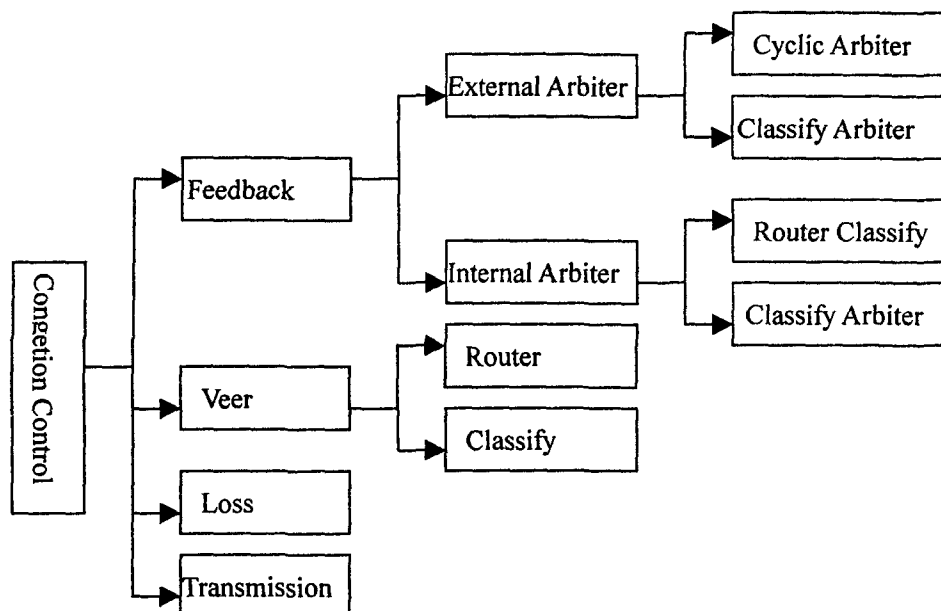
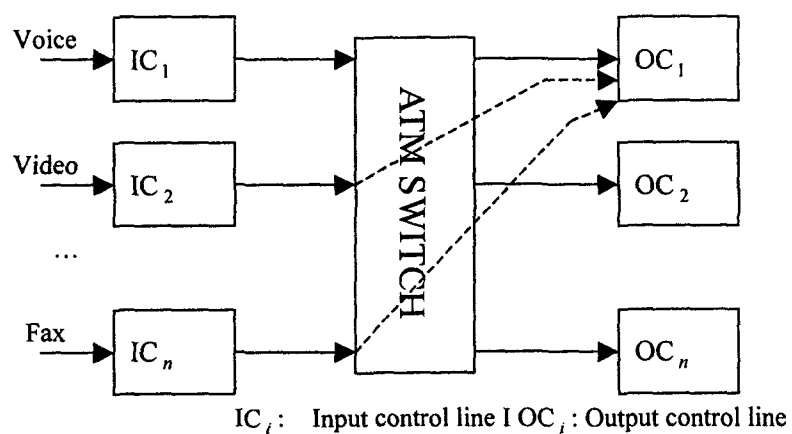Fig1. ATM congestion control procedure

IC_i : Input control line I OC_i : Output control line
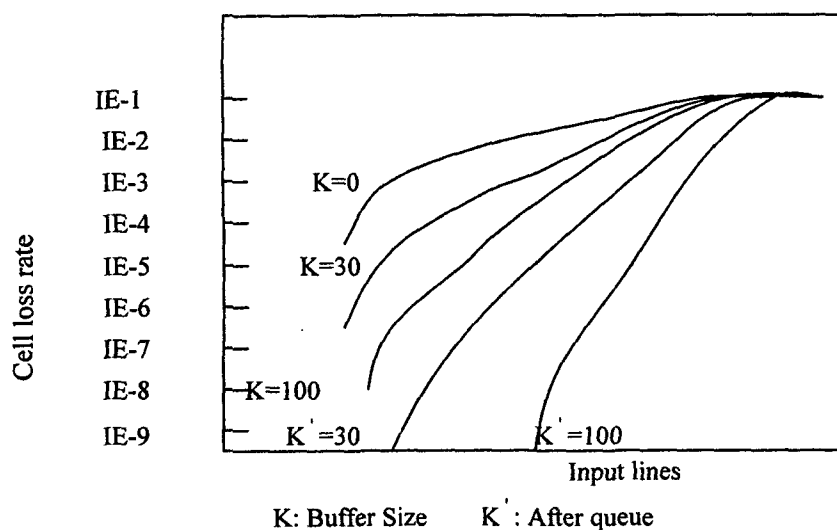
Fig2. Experiment and Device

K: Buffer Size   K': After queue

Fig3. Cell Loss Rate and Buffer Size