TECHNICAL RESEARCH REPORT

On QoS Provisioning in ATM Networks

by Anubhav Arora, John S. Baras

CSHCN TR 2002-7
(ISR TR 2002-14)

19 Center for Satellite and Hybrid Communication Networks 91

On QoS Provisioning in ATM Networks

Anubhav Arora, John S. Baras
Department of Electrical and Computer Engineering
University of Maryland at College Park
College Park, MD 20742

Abstract

ATM is representative of the connection-oriented resource provisioning class of protocols. An ATM network is expected to provide end-to-end QoS guarantees to connections in the form of bounds on delays, errors and/or losses. Performance management involves measurement of QoS parameters, and application of control measures (if required) to improve the QoS provided to connections, or to improve the resource utilization at switches. QoS provisioning is very important for real-time connections in which losses are irrecoverable and delays cause interruptions in service. Most scheduling disciplines provide static allocation of resources at connection setup time. End-to-end bounds are obtainable for some schedulers, however these are precluded for heterogeneously composed networks. The resource allocation does not adapt to the QoS provided to connections in real-time. In addition, mechanisms to measure the QoS of a connection in real-time are scarce.

A novel framework for QoS management is proposed in this paper to provide QoS guarantees to real-time connections. It comprises of in-service QoS monitoring mechanisms, a hierarchical scheduling algorithm based on dynamic priorities that are adaptive to measurements, and methods to tune the schedulers at individual nodes based on the end-to-end measurements.

Index Terms

ATM, QoS Provisioning, Performance Monitoring.

I. Introduction

IN any connection oriented QoS provisioning protocol, a contract is defined between the network and the user introducing a flow in the network at the connection setup phase. It includes the QoS requirement of the user and also the expected behavior of the user traffic. The network provider may police the user traffic upon entry into the network and the network is expected to furnish the contracted QoS to the flow until the termination of the call. An important point to note is that the QoS requirements are always end-to-end in nature, i.e., the performance metrics are defined for the complete path of the call. Real-time connections demand strict control on the end-to-end delay and jitter of the packets. We consider the problem of QoS provisioning for real-time connections in connection oriented networks.

Traffic Engineering deals with the control of data-flow inside the network and the issues of guaranteeing performance to connections. It is an important constituent of protocols like ATM, MPLS, and Diff-Serv. The principles of traffic engineering remain similar for these protocols. ATM is the classical representative of the concepts behind these protocols, with mature and precise standards defining it. Henceforth, this paper would be confined to the subject of traffic engineering for real-time flows in ATM networks.

Specifically, the problem is to provide guaranteed end-to-end delay bound, delay-jitter bound, and a bound on the cell loss rate to a real-time connection in an ATM network. The standards pertaining to ATM [1], [2], [3] do not specify any mechanism for this problem, and the implementations, if any, are vendor specific. The algorithms that provide control over QoS on a network element are the queueing and scheduling discipline. A considerable amount of research has been dedicated to the study of queueing and scheduling in one network element. Though there are results to calculate bounds on end-to-end delays for certain schemes, there are hardly any mechanisms to provide the required

This material is based upon work supported under contract number QK9931.

end-to-end delays to connections. It is also important to note that due to the heterogenous composition of the network, switches could be running different queueing and scheduling algorithms. In this case the calculation of end-to-end delay bounds becomes exceedingly difficult.

Moreover, in every queueing and scheduling mechanism the resources are assigned at the setup phase to a connection and there is no procedure to change these resources in response to changing traffic conditions. It is also necessary for the network and the user to be able to monitor and measure the performance metrics in real-time connections. The standard pertaining to Operations and Maintenance in ATM [4] provides inadequate methods for monitoring.

A framework for QoS provisioning is proposed, with a description of of the various modules required [5]. This framework does not assume the existence of same scheduler at every node, rather only a few basic requirements from the schedulers. On the subject of performance monitoring, a new protocol is proposed in order to provide an accurate and efficient measurement scheme. For switches, a hierarchical scheduler is presented that is capable of providing different types of local QoS bounds.

II. Earlier Work

A. Monitoring

The Operations, Administration and Maintenance (OAM) standard for ATM from ITU-T [4] specifies fault management and in-service performance monitoring mechanisms. One way delay measurement requires that the clocks at the source and the destination be synchronized, thus only round trip delays are accurately measurable using the optional timestamp field and loopback of cells at the destination. Moreover, the timestamp field is optional and is not implemented by most applications currently. Round trip delay measurements do not provide the capability to pinpoint the bottleneck links/segments in the circuit as well. The standard at no point mentions the objective of using the minimum overhead of OAM cells or any related algorithms. The precision with which the measurements are to be made, both the precision of each measurement and the interval between measurements, is not addressed in the standard. Thus it is observable that the OAM standard for ATM does not specify many of the performance monitoring objectives.

The authors of [6], [7] propose that the one-way cell delay can be accurately measured by segmenting it into delays experienced at each switch. Thus, a management cell accumulates the delay along the path, as a sum or as distinct fields. The delay field at the destination gives a sample of the cell transfer delay which does not suffer from the clock synchronization problem, as the differences are taken from the same clock. This scheme requires new processing capabilities at the switches to modify cells on ingress and egress. Also note that this technique requires new implementation in all existing switches, which is nearly an impossible task.

The technique proposed by Roppel in [8] relies on estimation of the one-way cell transfer delay by analyzing the properties of the remote clock and correcting the time-stamp of the destination for the offset. In this method, the switches do not need any new processing capabilities. This method however requires that the minimum delay along both directions to be the same. The number of samples required to converge to correct clock parameters may be large, thus the time for convergence for low bit rate links can be very high. These disadvantages can prohibit the use of this scheme for a large class of networks and connections.

B. Scheduling

The CAC assigns resources to new connections while providing protection to existing connections, and the scheduling controls the delay of cells at a switch. New connections that can potentially cause congestion in the switch are not accepted. The scheduling disciplines can be broadly divided in the following categories (see [5] for a comprehensive review):

1) Priority scheduling
2) Generalized Processor Sharing based bandwidth sharing
3) Round Robin based bandwidth sharing
4) Delay based scheduling
5) Traffic shaping (including Rate-Controlled schemes)

The scheduling schemes cater to very specific objectives, for example, fair allocation of bandwidth or maximum delay guarantees or bounds on CDV etc. Each of them is suitable only to a specific class of service in order to optimize different parameters. Also the disciplines either tightly couple delay and bandwidth together or look at only delay requirements. In order to provide end-to-end guarantees, a homogenous composition of the network needs to be assumed. However in a real ATM network, there are multiple classes of traffic on a switch with many widely different QoS objectives ranging from a fair share of bandwidth, to delay and loss guarantees, to both together.

III. Framework Description

A. Assumptions

- **Single ISP Domain:** Suitable mechanisms exist to delegate QoS requirements between service providers. In this paper, the provisioning of QoS in one domain would be considered.
- **Policing at the Edge:** Every connection is policed for its UPC contract parameters at the network access points.
- **Output Queueing in Switches:** The focus of this paper is on the output queued switch architecture.

B. Requirements

A multiservice network like ATM requires tools for performance management in order to provide end-to-end QoS for real time connections. The following are some of the requirements for the design of such a framework:

1) **Measure:** Means to accurately measure the QoS of real time connections are needed.
2) **Adaptation:** Mechanisms to change the resource allocation during the life of a connection should be present.
3) **Local bounds:** The provisioning of QoS to the cells of a connection on a single node, given the knowledge of the required performance bounds, is not an easy task either because the requirements imposed vary widely in nature from bandwidth guarantees to bounds on delay, jitter and loss. As noted above, there is a significant body of literature devoted to the subject but, the design of each scheduling discipline is to optimize on a single objective.

The theme of the proposed framework is to perform in-service QoS monitoring for real-time connections using special (OAM) cells, and correct the resource allocation parameters at the intermediate nodes, if required, in response to the periodic measurements. The adaptation protocol should also have the flexibility to improve the resource utilization of switches in case the measured QoS is significantly better than the contracted parameters. Thus the following components are needed in the framework: ability to accurately and efficiently monitor the QoS of a real-time connection (using the in-service monitoring methods) to measure the end-to-end delay, delay-jitter and loss rate, flexible QoS provisioning at intermediate nodes (i.e., correctable scheduling and CAC parameters on switches), and algorithms to regulate the QoS of a connection in real-time in order to either improve the QoS, or the resource utilization at nodes.

IV. Functional Components

A. Monitoring

QoS Monitoring of a connection is accomplished using specialized devices called *Performance Management (PM) devices* placed immediately prior to the network access equipment. The function of the PM device is to insert and extract OAM performance monitoring cells, and calculate the statistics of various QoS parameters on a per connection basis. The PM device can be capable of initiating control measures as well. The functionality of the PM device can be present in a seperate device [9], or in the ATM access switch. Note that this requires a new functionality only at the network edge and not in the core cloud. Based on the performance objectives for delay and loss parameters, a new protocol based on *pattern cells* is proposed to verify the conformance with the delay bounds. This new protocol has the flexibility to trade bandwidth overhead with the precision of measurement of delay.

For a particular connection time is considered to be slotted according to the the PCR of the circuit. By definition, the source will not inject cells with spacing less than $\tau_p = \frac{1}{PCR} sec/slot$. A cell belongs to the slot in which the larger fraction of it was transmitted. Pattern cells reflect the profile of user cell transmissions since the last pattern cell was transmitted, in the form of a bit sequence that furnishes information about the relative time between cell transmissions.

Each bit represents one or more time slots, depending upon the OAM bandwidth overhead that can be incurred on the connection.

In the 45 octet payload of the forward monitoring cell, it is proposed to use forty bytes representing the timing pattern. Let the number of time slots represented by a bit be n ($n \,\epsilon\, \mathcal{I}^+$) and the desired OAM overhead be h ($h \,\epsilon\, (0, .., 1]$) . Then,

$$n = \left\lceil \frac{PCR}{h * 8 * 40 * SCR} \right\rceil . \tag{1}$$

Now let i be the least integer such that $\frac{2^i - 1}{i} \geq n$. Thus, a block of i bits represents $n * i$ time slots, and the number of cell transmissions in these slots can be encoded using the i bits. The relative time of transmission of a cell from the time of transmission of the first cell can always be known within $n * i$ PCR slots. In the standard I.610, the minimum time for sending OAM cells is every 128 cells, thus giving the best case accuracy of determining a cell's relative position to be within 128 SCR time slots on an average. Consider an example: let $h = 0.5\%$, $\frac{PCR}{SCR} = 10$. Thus, $n = 7$ and $i = 6$. So the accuracy of this protocol is to be able to relatively place a cell's departure time within 42 PCR time slots equivalent to 4.2 SCR time slots, and the accuracy of OAM protocol (inserting OAM cells after every 128 cells) would be 128 SCR time slots. In this case, the protocol proposed has a bandwidth overhead of 0.5% and standardized OAM protocol has an overhead of 0.77%. For more accuracy, n can be decreased to 2, $i = 3$ and the relative position can be determined within 6 PCR slots or 0.6 SCR slots. In this case, the overhead is $h = 1.56\%$. The extreme case would that be of $n = 1$ and $i = 1$, where the positioning can be determined accurately, with the overhead about 3.13%. The overhead gets worse for streams with higher $\frac{PCR}{SCR}$ ratio (more burstiness).

The destination can take the time of arrival measurements by its own clock, calculate the time of departure of a cell using the relative position of the cell and the time of arrival of the first cell. An absolute time error T_0 in the measurement of CTD of a cell can be assumed because the clocks at the source and destination will never be synchronized. This error would be reflected in the mean of the delays and thus would not appear in the peak to peak CDV, or variance estimates. Cell losses can also be measured by the difference in the number of cells received and the number of cells transmitted (as known from the pattern cell). Cell transfer delays are measured within an additive constant. This constant can be determined by sending timestamped cells at periodic intervals using the techniques proposed in [6] or [8].

OAM cells are bandwidth overhead for the data stream. The idea of adaptive sampling is to analyze the trend in the gathered data and adapt the sampling rate to the minimum rate required for the purpose. In the protocol described above, the overhead being incurred can be reduced by increasing the value of n and thus i (which in turn makes accuracy worse). Further improvements can be obtained by inserting the parameters n and i in the cell itself, and varying these according to the source rate to incorporate the ideas of adaptive sampling mentioned in [10]. Thus using this trade off between bandwidth overhead and accuracy, monitoring can be adapted for either better accuracy or less overhead.

It is also proposed to standardize a high priority cell that is capable of jumping a queue, i.e., it gets queued to the head of the queue rather than the tail of the queue. A number of schemes can be derived using this cell to measure a variety of things, notably: the delay in a particular queue (node) or set of queues. This can be done by sending two cells at the same time, one of them jumping selected queues in the path. The difference in arrival times of these cells gives a reasonable estimate of the queueing delay of the queues that were jumped.

B. Scheduling at one node

The framework poses two requirements on the schedulers of the nodes: all schedulers should be able to provide various QoS guarantees, and the local QoS requirement of the connections should be correctable (in some schedulers, if not all). This is a large class of schedulers, and hence this framework would be appropriate for networks that have nodes running different schedulers.

Based on the QoS requirements define the following different classes of service in a switch [1] besides the best-effort service:

1) **Rate based and real-time QoS insensitive:** requiring an average bandwidth over a period of time.
2) **Maximum delay sensitive:** requiring bounded delay in every switch.
3) **Jitter sensitive:** requiring (approximately) the same departure pattern as the arrival profile of the connection. Also may require a bound on the maximum delay in a switch.
4) **Loss sensitive:** requiring a certain amount of buffer in the switch in addition to any of the above requirements.

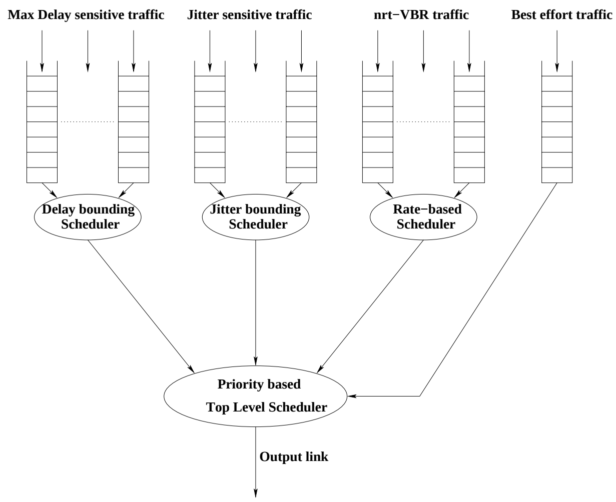

Fig. 1. Proposed hierarchical scheduler.

Each of the per-VC queue i has an associated urgency $U_i(n)$ as a function of time slots n, defined as the urgency of the head-of-the-line cell of that queue to exit the queue at slot n. The urgency is an abstraction for quantifying relative priority of connections demanding different QoS. The calculation for urgency is based on the QoS demand of the connection and other observed parameters. The urgency of the lower-level scheduler at a particular time is the urgency of the queue that is chosen to transmit at that time. For each slot on the output, the lower level schedulers contend (figure 1) and the slot is assigned to the highest urgency scheduler (ties are resolved randomly). Three lower level schedulers are as follows.

Rate based: The rate based QoS insensitive traffic is switched by a round robin or WFQ scheduler. The urgency of the head cell of a queue to grab a slot at the output rises as the connection gets less bandwidth on the output and vice versa. A moving average of the bandwidth provided to the connection can be maintained and the urgency reflects its difference with the traffic contract.

Delay based: The deadline of each cell j of queue i is the local delay bound in the switch (δ_i) added to the arrival time of the cell (A_i^j). The cells can be scheduled using an EDF scheduler which is known to be an optimal scheduler [11]. The relative delay (i.e., the ratio of the waiting time to the local delay bound) of the chosen cell can be used as the urgency of the scheduler.

Jitter based: The design criterion for this scheduler is that the cell should be transmitted very close to its deadline. Thus the urgency of a cell at the head of the queue should rise to maximum very sharply at the time of the deadline. A delay bounding scheduler with very short delay bounds can be used for the purpose, otherwise any jitter bounding scheduler can be used with an appropriate definition of urgency.

In order to provision for the cell losses, a buffer sharing policy is required which is beyond the scope of this paper. If none of the schedulers require the slot, the UBR queue is chosen if it is not empty. The notion of urgency introduces the concept of comparison between different QoS classes. The definition of urgency for various classes also introduces fairness between them. The concept of fairness in this case extends to comparing connections with different requirements.

C. Adaptation

In every switch the scheduling discipline needs to be adaptive to the end-to-end QoS measurements on real time connections. The correction in the scheduling parameters of a particular connection is in response to the end-to-end measurements performed by the PM device. The control measures can be initiated in two distinct ways - **Centralized:** The PM device taking measurements and calculating the statistics would also evaluate whether the QoS is in conformance with the guarantees, and detect any patterns of deterioration. Subsequently this device would also to identify the switch responsible for the violation or deterioration in QoS, and send a message addressed to that switch to take corrective measures and increase the priority of the connection. **Distributed:** The PM device would send out the measurements or functions of the measurements (for example alarm states) in a special cell to all the switches in the path. The switches after reading the information would respond to it by changing the priority of the connection (if required) according to their resource availability.

However, methods to identify the switch causing the most deterioration in QoS (bottleneck identification) do not exist. Identification would also take one Round Trip Time after congestion has been detected in the circuit. Also, in the case when the measured QoS is much better than the required, there is no mechanism for the PM device to relax the requirements at some switches. Thus the centralized scheme is inefficient and nearly unimplementable. In the distributed scheme, cells could be periodically sent out to indicate the measurement results to the switches in the path. If there is a significant difference between the measurements and requirements, the first switch in the path (that can accommodate new scheduling parameters to alleviate the difference) takes the corrective measures. It also marks its action on the forwarded cell to inform the downstream nodes. Therefore if a corrective action is required, all switches in the path would contribute as much as they can afford. Similarly, if the measurements are much better than the contracted QoS, the switches on the path that are experiencing congestion may increase the local delay bound of the connection. The delay in control action after the measurement is performed is approximately one round trip time in the usual case. This is a simple heuristic algorithm to adapt the local schedulers at switches.

In [12], the authors propose a similar scheme for IP networks, where the PM device changes the resource allocation of the connection in all the switches of the path. According to their simulations, a conservative adaptation scheme "gracefully adapts to the state of the network and converges to a stable state fairly responsively." The proposed heuristic algorithm above, as well as the adaptation of the schedulers requires further study, however the positive results obtained in [12] indicate that an efficient distributed scheme can be designed for ATM networks also.

V. Summary

This paper proposes a new QoS provisioning architecture to guarantee end-to-end QoS for real-time connections. The scheduling at one switch is an important constituent of this framework. The proposed hierarchical scheduler is capable of providing bandwidth to connections, and bounded delay or jitter guarantees to real-time connections independent of their bandwidth requirement. In order to guarantee end-to-end performance the delay and jitter statistics are measured on a per connection basis using in-service monitoring mechanisms. A scheme for accurate measurement of delay and jitter using pattern cells are presented in the paper. Messages are sent to all nodes by the PM device if QoS violations or deteriorating patterns are detected. Periodic updates of QoS are also sent in order to facilitate switches to change parameters for better resource utilization. For QoS sensitive connections, the parameters controlling the urgency can be changed based upon end-to-end QoS measurements.

A complete understanding of the framework requires more answers. The problem of communicating the measurements to the switches in the nodes, and initiating distributed corrective measures in the schedulers of the switches is an open problem in the context of ATM networks. The effect of control actions of one connection on the network performance and on the performance of other connections (including a study of stability) is a topic for further research. In addition to the updates algorithm, the choice of the scheduling parameters of connections at connection setup also needs to be determined. In the study of the hierarchical scheduler, the optimal scheduler for the rate shaped connections and the corresponding definition of urgency is a topic for future research. The notion of urgency also needs to be defined for any other scheduler that can be used as a delay or jitter bounding scheduler.

REFERENCES

[1] ATM Forum, *Traffic Management Specification Version 4.1*, Mar 1999.

[2] ITU-T Recommendation I.356, *B-ISDN ATM layer cell transfer performance*, Mar 2000.

[3] ITU-T Recommendation I.371, *Traffic Control and Congestion Control in B-ISDN*, Mar 2000.

[4] ITU-T Recommendation I.610, *B-ISDN Operations and Maintenance Principles and Functions*, Feb 1999.

[5] Anubhav Arora, *Performance Management in ATM Networks*, Ph.D. thesis, University of Maryland at College Park, May 2002.

[6] T. M. Chen et al., "Monitoring and control of ATM networks using special cells," *IEEE Network*, pp. 28–38, Sept/Oct 1996.

[7] Thomas M. Chen and Stephen S. Liu, "Monitoring and control of strategic and tactical networks using control packets," in *Proceedings ATIRP Consortium*, Feb 1999, pp. 183–187.

[8] C. Roppel, "Estimating cell transfer delay in ATM networks using in-service monitoring," in *Proceedings Globecom*, 1995, pp. 904–908.

[9] RAD Data Communications - White Paper, *Customer Premises ATM NTUs: Cost Effective End-to-End Management of ATM Services*.

[10] Anubhav Arora, John S. Baras, and George Mykoniatis, "Delay monitoring in ATM networks," in *Proceedings ATIRP Consortium*, Feb 1999, pp. 259–263.

[11] Leonidas Georgiadis, Roch Guerin, and Abhay K. Parekh, "Optimal multiplexing on a single link: Delay and buffer requirements," in *Proceedings of Infocom*, 1994, pp. 524–532.

[12] Thyagarajan Nandagopal, Narayanan Venkitaraman, Raghupathy Sivakumar, and Vaduvur Bharghavan, "Delay differentiation and adaptation in corestateless networks," in *Proceedings of IEEE INFOCOM*, Tel Aviv, Israel, March 2000.