On Fault Recovery Priority in ATM Networks

Bjørn Jæger *
Department of Informatics
Molde College, Britvegen 2-N-6400, Molde, Norway

David Tipper †
Department of Information Science and Telecommunications
University of Pittsburgh, Pittsburgh, PA 15260

Abstract

In this paper we propose a priority scheme for reconnection of virtual circuits (VCs) in ATM networks that have been disrupted by a failure. ATM networks offer several service categories each designed to handle applications with specific traffic characteristics. A failure typically results in a large number of disrupted connections in each category, all of which must be restored simultaneously. A critical issue in the restoration is the order in which the service categories are processed, and the order of processing the connections and routing within each category. The proposed priority scheme aims at minimizing the impact of a failure on the network.

1 Introduction

A basic challenge in ATM networking is meeting user demand for reliability and fault tolerance in a cost effective manner. The topic of fault recovery in ATM networks has received special attention recently with work on physical layer, logical layer (i.e., virtual path (VP) layer in ATM) and VC layer restoration appearing in the literature [1, 2, 3, 4, 5, 6].

One aspect of fault recovery at the VP and VC layers that has been ignored is the ordering of VPs and VCs for rerouting. This problem can be seen by considering a wide area ATM network which uses source node routing of the VCs. In source node routing, each network node maintains a database of the network topology and determines the route through the network for all VCs originating at the node. The PNNI routing scheme developed by the ATM Forum for switched VC routing within a peer group adopts this approach [9]. In the event of a failure the VCs using the failed device are disrupted and need to be reconnected if possible. The source nodes for the VCs that were traversing the failed device are responsible for the restoration of the affected VCs, as discussed in the ATM Forum. A failure typically results in *several* nodes being sources for failed VCs with each having *many* VCs to *simultaneously restore*, possibly on the order of tens of thousands. The way in which the VCs are processed will determine in part, if the connection is restored, the delay in reconnection and the QoS provided after restoration.

In this paper we propose a restoration priority scheme based in part on the ATM service classes, which aims at minimizing the impact of a failure on the network while providing users the best possible service. The scheme involves both a priority for reconnection among ATM service classes and a rule for ordering and routing VCs within a service class. This is based in part on our earlier work on ATM fault recovery routing [5]. The proposed scheme is formulated within the context of switched VC routing but is applicable to VP restoration as well.

Note, that after a failure all cells are lost for the duration of the restoration period causing service disruption ranging from msec to several seconds and beyond depending on the type of failure and the restoration strategy. How long an acceptable disruption period is varies with the type of application, the protocol used and the protocol settings [12]. It has been shown that the dominant factor on network performance immediately after a failure is the transient congestion due to retransmission of dropped cells [13]. Studies have shown [13] that this additional load is the reason for buffer overflow at both the source node switch and at the buffers of other nodes along the path to the destination. Note that a dropped cell may trigger a packet retransmission at a higher level causing the number of retransmitted cells to be larger then the number dropped. How traffic sources handle retransmissions is therefore a primary issue to address when developing a priority scheme. Hence, the priority traffic restoration technique proposed here is based in part on minimizing the number of dropped cells that need retransmission, thus reducing the transient congestion.

The rest of the paper is organized as follows. In Section 2, we present a discussion of ATM service categories and retransmission of lost cells. Section 3, discusses the timing of the fault recovery process. In Section 4 the ordering of VCs for processing at a source node is discussed. Section 5 presents the proposed priority fault recovery scheme and an evaluation of its performance. Lastly, Section 6 summarizes the paper.

*funded by Norwegian Research Council Project: 110912/410

†supported in part by NSF grant NCR-9506652

0-7803-4788-9/98/$10.00 © 1998 IEEE.

2 Service Categories

The five service categories supported by ATM networks are [8]: Constant Bit Rate (CBR), real-time Variable Bit Rate (rt-VBR), non-real-time Variable Bit Rate (nrt-VBR), Available Bit Rate (ABR), and Unspecified Bit Rate (UBR). Considering the ATM service classes, in terms of retransmission of dropped cells the following general behavior is found.

CBR and rt-VBR applications

The CBR and rt-VBR categories have identical timing requirements, and are suited for real-time applications. Under normal operation it is assumed that the cell-loss-rate (CLR) is sufficiently small to be ignored by the application. Applications in these categories are not likely to retransmit lost cells since they will be of little or no value for the application.

nrt-VBR applications

The nrt-VBR category has cell-transfer-delay (CTD) guarantees but no guarantees regarding cell-delay-variations (CDV) making it suitable for response-time-critical data applications. For these applications all cells are vital, and lost cells are retransmitted. Under normal operation this class guarantees a low CLR, which means that retransmission of these cells does not place significant additional burden on the network. In case of a failure all cells lost during the restoration period must be retransmitted. Since this additional traffic load is due to a network failure (the user is not at fault) it behooves the network provider to carry this traffic in a timely fashion.

ABR applications

Under normal operation, the CLR is low since the sources adjust cell flow in response to network abilities provided by control information. Lost cells are expected to be retransmitted causing no significant extra burden on the network. Upon a network failure, the sources are notified of the network degradation by the missing control-cells. The application is then expected to temporarily stop transmitting until the restoration procedure is completed.

UBR applications

Both under normal and failure operation lost cells are expected to be valuable for the application and retransmitted. However, the UBR service category is expected to be run in a private environment which is not a multi-service environment (e.g., LAN) and we ignore this category.

3 Timing Considerations

A network failure causes a service outage until the failed component is fixed or until the traffic is restored around the failure. In studying the outage time it is useful to consider a single VC traversing multiple ATM switches. The source switch is the first access switch next to the calling party, and the last switch of a VC is the destination switch since it is next to the called party. Consider a link failure between two intermediate switches along the VC path. We denote the neighboring switches of a failed link as the upstream and downstream switches, where the downstream switch is the one closest to the destination switch. In general, the outage time consists of the following components: Detection, Notification, Route Selection and Rerouting. We briefly discuss each in turn.

Detection Time: The detection time varies considerably depending on the type of failure occured (line card, software, etc.). For the specific case of a physical layer failure, the failure is detected by the downstream switch and reported to the local ATM Layer Management. The reporting of a physical layer failure to the ATM Layer Management occurs within an *unspecified vendor specific* time period [10]. There is a variation in the detection time at the physical layer as well. For example, the ATM Forum calls for detection of loss-of-signal within 100 *μsec* and for loss-of-frame within 3 *msec*.

Notification Time: Upon detecting a failure a switch notifies the other switches along the path towards the destination using the dedicated Fault Management operation-and-maintenance (OAM) cells called alarm-indication-signal (AIS) cells. According to ATM Forum specifications, within 250 *μsec* after the destination switch has received the AIS cells, the switch generates a remote-defect-indication (RDI) OAM cell which is sent upstream to the source switch. This notification is used to trigger the dynamic rerouting process in the source switch and other failure dependent operations like stoping billing. Typically the RDI signal must persist for a certain *vendor specific* time length before the source switch considers a failure to have occurred. Note that each failed VC within a source node receives the RDI-cells at different points of time since the distance from the source to the destination generally varies for each VC, and the signalling might use connections of varying length and speed. A simple method for reducing the notification time is that if once the first failed VC RDI cell is received, the source node checks if any of its other orginating VCs are also affected by the failed device in order to start restoration immediately rather than wait on the RDI cells.

Route Selection Time: When a source switch is notified of a failed VC it releases all resources allocated to the failed VC and tries to select a new route for the VC. In ATM PNNI routing [9] each switch maintains a topology database containing the cost of using each link in the network. Also, to speed up the routing computation each node maintains a precomputed set of alternative routes between each source destination pair that is determined from the network topology (usually restricted by a hop count limit). The link cost database is updated periodically with each node notifying the other nodes of its current cost using a flooding approach. In addition, asynchronous updating of the link cost occurs whenever a link utilization changes by an amount exceeding a predefined threshold. Based on the cost of each individual link, the path cost can be calculated. There is no standard routing algorithm specified and the cost of a path can be calculated in several ways. In [5] we studied the behavior of four different routing algorithms for fault recovery in ATM and their effects on call blocking, demand restored, route selection and the transient network congestion. The routing schemes considered were:

1. Minimum Delay (MD) routing uses a link cost proportional to the derivative of the link queueing delay based on an M/M/1 model as in [14].

2. Minimum Hop (MH) routing uses a load independent constant link cost. The MH approach results in the number of nodes directly handling rerouted connections being minimized and the restoration paths are as short as possible. This method attempts to isolate and restrict the physical area of congestion occurring due to rerouting.

3. Load Distribution Among Paths (LDAP) uses the negative of the residual capacity as link cost, resulting in selecting the route with the maximum residual bandwidth. The LDAP approach distributes the potential congestion among the links in the network in order to not degrade the performance at any one link.

4. Load Distribution among Primary Links (LDPL) focuses on VC source nodes and attempts to balance the load on outgoing links. The link cost on the outgoing links of the source node is inversely proportional to the residual capacity on the link, downstream link costs are a load independent constant as in MH.

It was shown that each algorithm has advantages over the others depending on the network load and which of the performance parameters (length of congestion period, % demand restored, call blocking, etc.) one wants to emphasize. The computational time used by a source node to select new paths is expected to be on the order of μsec per VC, based on the measurements given in[11].

Rerouting Time: The rerouting of a VC is done using the same mechanisms as when setting up a new VC. A setup message containing the destination address, the desired traffic and QoS-parameters is sent along the minimum cost path. The Connection Admission Control (CAC) procedure at each switch along the path allocates resources for the call, forwards the message to the next switch if sufficient resources were available, or sends a crank-back message back towards the source switch in order to try to find another route. Upon receiving a crank-back message, each intermediate switch release previously allocated resources. The call will be rejected if no alternative route with sufficient resources can be found. If the call is accepted by the network and the destination, a connection acceptance message is sent back to the source and the VC begins working. The VC setup time is a largely a function of the number of hops traversed. Let $t(H)$ denote the time it takes to reconnect a VC over a path with H switch-hops. In our calculations we use the measurements from [11] where t(2)=395, t(3)=535, t(4)=619, and t(5)=669, all in msec.

4 Ordering of Connections

The manner in which a source node processes the VCs for restoration will affect; the routes selected, the connection blocking, the time to reconnect a specific VC, and the length of the total restoration period. Basically there are two ways to process the failed VCs either sequentially or in parallel. In the sequential method, at each source node, each failed VC is connected/rejected before the next one is processed. Whereas, in the parallel approach each source node sends out setup messages for it's failed VCs as quickly as possible.

If the sequential method is used the order in which the VCs are processed can be done in several ways, here we consider two: 1) sorting the VCs by increasing amount of bandwidth (i.e., demand), and 2) sorting the VCs by decreasing amount of bandwidth. The first approach is expected to give the lowest call blocking rate, and the second approach is expected to maximize the demand restored. This is confirmed by our results reported in [5]. The drawback of the sequential approach is the long restoration time which is nearly a linear function of the failed VCs at a source node.

In order to reduce the total recovery time the parallel method can be adopted whereby the setup messages for all failed VCs are distributed as fast as possible, the source switch then handles the corresponding replies from the destination switches in a FCFS order. A similar approach for restoration at the DCS cross connect level in circuit switched networks is discussed in [12], where it is noted there are many practical implementation hurdles in parallel recovery schemes. Note, that under heavy loading conditions many failed setup attempts occur since the source nodes have inaccurate link cost information, extending the restoration process possibly to the extent that no gain in restoration time is achieved compared to the sequential approach.

Here, we use the sequential approach since in current ATM standards the signalling for parallel VC setup is not available. Also, the sequential approach will have better call acceptance and route selection behavior since link cost updates may occur during the processing.

5 A Priority Restoration Scheme

Given the preceeding discussion we propose the following simple priority scheme: 1) strict ordering among service categories; 2) specific routing scheme for restoration within each category; and 3) specific ordering of VCs for processing within each category. Note that all categories are handled in a strictly sequential order. We revisit the ATM service categories to specify/justify the scheme.

<u>nrt-VBR:</u>

Since the nrt-VBR category is designed for applications requiring short response time, low cell loss and retransmission of lost cells, this category is given the highest restoration priority among the service categories. Within this category, the MH-scheme is suggested for rerouting since it has the shortest Rerouting Time, which will minimize the number of cells dropped. Further, in order to restore the maximum number of failed VCs, the VCs are restored in increasing order of bandwidth.

<u>CBR</u>

The CBR category is designed for real-time applications which do not retransmit lost cells. It is assumed that a short service interuption is acceptable while the nrt-VBR connections are being restored giving CBR applications the second highest restoration priority among the categories. Within this category, the MD-scheme is suggested for rerouting, since it is optimal for steady state network conditions. Moreover, in order to restore the maximum amount of demand the VCs are restored in a decreasing order of bandwidth. Note that, since CBR is expected to support telephony with its relatively low service outage requirement [12] it is given priority over rt-VBR.

rt-VBR

This category is designed for real-time applications which do not retransmit lost cells, and the same reasoning as for CBR applies. This category is given the third highest priority among categories and MD rerouting with decreasing bandwidth ordering is used.

ABR

The ABR category is designed for applications which can adapt to changing network performance. Hence, the number of dropped cells should be small. Further, since it has no real-time requirements it is recommended to have the lowest restoration priority among service categories. The LDAP-scheme is used for rerouting, since this scheme distributes the congestion over the set of available paths and results in lower call blocking. The VCs are restored in increasing order of bandwidth to the maximze number of restored VCs.

Table 3 summarizes the priority scheme. In order to evaluate the effectiveness of the priority scheme we studied a sample network and calculated outage times and cells lost. To provide a basis for comparision under a common set of *simplified* assumptions we studied using no priorities (random ordering) versus using the priority scheme. The ten node, $n = 10$, network with $|L| = 42$ links shown in Figure 1 was studied. The connectivity of the topology as measured by the average node degree is 4.2, which is in the range of many existing networks. A C program model of the network was developed. The routing algorithms were implemented in the network model in a distributed fashion with each node maintaining a local database of the link cost with periodic cost updates. A hop count limit of five links was used in the model to restrict the number of feasible paths and speed up the route selection. In the simulation the capacity of each link was OC-3 rate. A maximum steady-state link utilization threshold of 0.9 was used for all links and calls are blocked if no route can be determined without exceeding the utilization threshold.

A sample set of results with no ABR traffic are reported here for a lightly loaded network with an average link utilization of 0.4 before the failure. Since the network was lightly loaded all of the failed VCs were restored. A total of 100 experiments were conducted, in each of them enough VCs were setup to create the average link utilzation of 0.4 ($\approx$ 500 VCs on average). The VC were set up with random source - destination selection, random service category selection and random demands selected from the following ranges: nrt-VBR: 9.6 - 155 Kbit/s, CBR: 9.6 - 15500 Kbit/s, and rt-VBR: 9.6 - 3100 Kbit/s. The ranges were picked based on the rates of typical applications intended for these categories. A MMPP model was used for nrt-VBR traffic and a Possion model for rt-VBR.

After the VC were set up, link 2-4 was failed which on average caused approximately 30 VCs to fail. The time to detect the failure at node 4 and notify the source nodes was a constant of one second. The Route Selection time was set to 250 μsec per VC as in [11]. For each failed VC we calculated: the new path, the outage time, and the number of cells lost. The results were then grouped by service category and averaged over the 100 experments. This is summarized in Tables 1 and 2 for the random ordering and priority schemes. The Outage Times in the tables represent the mean outage time of all VCs in that category averaged over the 100 experiments. The Cells Lost listed in the tables is the total number of cells lost by VCs in that category averaged over the 100 experiments.

Category	Outage Time	Cells Lost	Routing Alg.
nrt-VBR	32.4	26,703.9	MD
CBR	32.8	557,343.0	MD
rt-VBR	35.3	121,239.2	MD

Table 1: Summary for no priority restoration

Category	Outage Time	Cells Lost	Routing Alg.
nrt-VBR	20.8	5,271.9	MH
CBR	34.9	555,329.9	MD
rt-VBR	52.5	171,591.5	MD

Table 2: Summary for priority scheme

In comparing the results in Tables 1 and 2 we can see that the priority scheme results in a short outage time and smaller number or cells lost for the nrt-VBR category VCs. Specifically, the total number of lost cells for the nrt-VBR connections is 5272 under the priority scheme in comparision to 26703 lost cells under the random case. Since these lost cells will need retransmission the reduction in the number lost should significantly improve network performance after the failure. This is at a cost of increasing the average outage time for the CBR and rt-VBR connections from 32.8 to 34.9 seconds, and from 35.3 to 52.5 seconds respectively. Additional numerical results, illustrating similar behavior (including ABR traffic) are given in [15]. A more detailed simulation study is underway and results including transient network congestion metrics and normalization times will reported in later versions of this paper. Lastly, we note that in a real network each source node might have hundreds or even thousands of failed connections to restore. Thus the number of lost cells will be more significant indicating that congestion will be a important issue. For these situations it generally may not be possible to restore all nrt-VBR connections and still satisfy the recommended restoration time constraints on the CBR and rt-VBR connections. Therefore it might be better to use a scheduling scheme to give a time slot to each category. For example, using the principle of weighted fair queueing common in processor sharing algorithms, the time spent for each category can be set proportional to the number of failed VCs in the category. This is currently under study.

6 Summary

In this paper we presented a priority scheme for restoration of connections in an ATM wide area net-

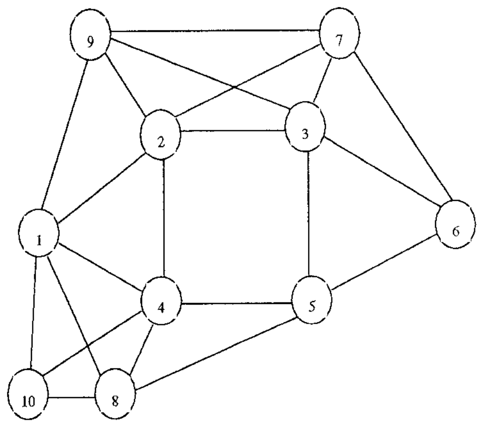

Figure 1: Ten node network topology

Service category	retrans-mission	priority	routing	ordering
rt-VBR	no	3	MD	decreasing
CBR	no	2	MD	decreasing
nrt-VBR	yes	1	MH	increasing
ABR	maybe	4	LDAP	increasing

Table 3: Priority scheme

work. Our numerical results show that the priority scheme proposed significantly reduces the amount of cells needing retransmission after a failure, thereby reducing network congestion, at a cost of longer restoration times for lower priority service categories.

References

[1] Special Issue, "Integrity of Public Telecommunications Networks," *IEEE Journal of Selected Areas in Communications*, Vol. 12, January 1994.

[2] K. Murakami and H. Kim, "Virtual Path Routing for Survivable ATM Networks," *IEEE/ACM Transactions on Networking*, Vol. 4, No. 1, Feb., 1996.

[3] T.H.Wu, *Fiber Network Service Survivability.* Artech House, Boston, Mass., 1992.

[4] K.R. Krishnan, et. al., "Improved Survivability with Multi-Layer Dynamic Routing" *IEEE Communications Magazine*, July 1995.

[5] W.P. Wang, et. al., "Fault Recovery Routing in Wide Area Packet Networks, *Proceedings of 15th International Teletraffic Congress*, Washington, DC, June 1997.

[6] D. Hsing, et. al., "A Restoration Methodology Based on Pre-Planned Source Routing in ATM Networks," *Proceedings IEEE ICC '97*, Montreal, Canada, June, 1997.

[7] ITU-T Rec. I.371 "Traffic Control and Congestion Control in B-ISDN,".

[8] ATM Forum, *Traffic Management Specification Version 4.0.* ATM Forum Document af-tm-0056.000, April 1996.

[9] The ATM Forum Technical Committee, *Private Network-Network Interface Specification version 1.0 (PNNI 1.0).* ATM Forum Document af-pnni-0055.000, March 1996.

[10] The ATM Forum Technical Committee, *BISDN Inter Carrier Interface (B-ICI) Specification Version 2.0 (Integrated).* ATM Forum Document af-bici-0013.003, December 1995.

[11] D. Niehaus, et. al., "Performance Benchmarking of Signaling in ATM Networks," *IEEE Communications Magazine*, August 1997.

[12] J. Sosnosky, "Service Applications for SONET DCS Distributed Restoration," *IEEE Journal of Selected Areas in Communications*, Vol. 12, January 1994.

[13] D.Tipper, et. al., "An Analysis of the Congestion Effects of Link Failures in Wide Area Networks," *IEEE Jrnl. Sel. Areas Comm.*, vol. 12, pp. 179-192, 1994.

[14] D. Bertsekas and R. Gallager, *Data Networks*, Prentice Hall, 1992.

[15] B Jaeger and D. Tipper, "Fault Recovery Priorities in ATM Networks," Technical Report, University of Pittsburgh, Feb., 1998 anonymous ftp to *violet.tele.pitt.edu*.