An Adaptive Multipath Routing Scheme for Connectionless Traffic in an ATM Network *

Josep Sole-Pareta * Debapriya Sarkar [†] Jörg Liebeherr [†] Ian F. Akyildiz ^{††}

* Departament d'Arquitectura de Computadors Universitat Politecnica de Catalunya Barcelona, Spain

> [†] Computer Science Department University of Virginia Charlottesville, VA 22903

^{††} School of Electrical and Computer Engineering Georgia Institute of Technology Atlanta, GA 30332

Abstract

This paper addresses the problem of routing connectionless traffic through an ATM network. A solution is proposed based on a perpacket adaptive multipath routing scheme which is added to the routing algorithm implemented at the Inter-Working Units. A scheme is presented that distributes packets among multiple Virtual Paths (VPs) according to the utilization of the links on these VPs. The utilization of the VPs is determined by a periodic feedback mechanism. Simulation studies show the effectiveness of the proposed adaptive multipath routing scheme.

^{*}The work of JSP was supported in part by a CIRIT (Generalitat de Catalunya) grant (expedient number EE92/2-338), and in part by the National Science Foundation under Grant No. INT-94033646. The work of IFA was supported in part by the National Science Foundation under Grant No. INT-94033646.

1 Introduction

To support connectionless data traffic in B–ISDN one must consider the following three aspects:

Architecture. Three different architectures have been proposed for implementing connectionless data transport services on top of ATM Networks [9]:

- (i) In the On-Demand architecture, a virtual channel connection (VCC) is established for each packet¹.
- (ii) The Semi-Permanent Connections architecture, illustrated in Figure 1(a), assigns a Semi-Permanent Virtual Path for each pair of Inter-Working Units (IWU). The result is a fully meshed network of virtual paths (VPs) that connect the IWU switches, forming a virtual network of VPs on top of the ATM network.
- (iii) The Connectionless Virtual Overlay Network architecture, shown in Figure 1(b), employs Connectionless Servers (CLS) that are connected by Semi-Permanent VPs. The CLSs provide routing functions for connectionless traffic.

Congestion Control Policy. A congestion control policy must be implemented for efficient traffic management in each of the three architectures. The objective of this policy is to maximize the bandwidth allocation for connectionless traffic without violating the service guarantees that have already been made to connection-oriented traffic. Most approaches for bandwidth allocation fall into two categories: bandwidth reservation methods [1, 2, 8] where IWUs allocate a fixed amount of bandwidth for connectionless traffic, and bandwidth regulation methods [3, 4, 7, 11] where the bandwidth available to connectionless traffic is controlled by an adaptive feedback scheme that involves the ATM switches and the IWUs. A recent study proposes a scheme that combines bandwidth reservation and regulation [5].

¹We use the term **packet** to refer to a *Connectionless Protocol Data Unit* or CL_PDU.

The Fast Reservation Protocol (FRP) [2] is a widely discussed bandwidth reservation method where bandwidth is allocated only when IWUs have a packet to transmit. FRP works as follows: The source IWU indicates to the network the required peak rate using a reservation message which is sent to the destination IWU. Switching elements along the path to the destination IWU allocate the peak bandwidth. If sufficient bandwidth is available at every link on the route, an acknowledgment is returned, otherwise a negative acknowledgment is returned. In the former case, the packet is allowed to be transmitted, while in the latter case the packet is blocked at the source IWU. As an alternative solution to FRP, the bandwidth advertising scheme [3] proposes an on-the-fly transmission of packets. The transmission is based on the knowledge of the available bandwidth on the path through a 'bandwidth advertising' as follows: sources are allowed to transmit cells in excess of the available bandwidth; but the extra cells are marked as 'low priority' by setting the CLP bit in ATM cell headers. If the network experiences congestion it may selectively drop the low priority cells. Two different regulation schemes are presented in [4] and [11]. In both schemes, the traffic rates are dynamically adjusted based on end-to-end feedback information on the network congestion. An alternative to bandwidth regulation based on end-to-end feedback is the hop-by-hop approach [7]. Note that the bandwidth regulation methods in [4, 7, 11], are presented as general solutions for congestion control in ATM networks. But they are potentially also applicable for designing bandwidth allocation strategies for connectionless data traffic. Schemes combining both bandwidth reservation and bandwidth regulation mechanisms can also be found in the recent literature. For example, a combination of FRP with the adaptive peak rate control is proposed in [5].

Routing. In the context of virtual VP networks where VP endpoints are viewed as nodes and VPs are viewed as links, two routing methods have been considered for improving the performance of the ATM bearer service, namely *alternate paths routing* and *multipath routing*. In the former, traffic is sent primarily on a single path, and alternate paths are only selected for

transmission when the primary path experiences heavy congestion. In the latter, multiple paths can be used simultaneously. The multipath schemes can be further divided into two categories based on the granularity of the routing schemes:

Per-VCC-Routing: Routing is done on a per connection basis by setting up a VCC through a VP that can cope with the desired traffic load.

Per-Packet-Routing: A routing decision is made for each packet, i.e., each packet is routed through any of several alternative VPs.

The performance of both schemes is compared in [6] in the context of a single hop network with two nodes. For this simple scenario, it was shown that it is possible to obtain better performance with per-packet bandwidth allocations, especially if traffic is highly bursty. For bursty traffic, better performance is obtained despite the effects of the additional delays incurred due to resequencing of packets.

We study a per-packet multipath routing scheme which couples congestion control with the routing method. The first scheme that considered a coupling of congestion control and routing is proposed in [10]. The scheme provides each pair of IWUs with several parallel VPs, initially without any bandwidth allocation. When a burst must be sent, a broadcast request process is initiated on all paths. All paths independently try to allocate the bandwidth requested. The packet is transmitted along one of the VPs that have sufficient bandwidth available. The rest of the VPs release their bandwidth. However, bandwidth reservation on a burst-by-burst basis incurs control overhead and may result in excessively long delays if the propagation delay between the source-destination pair is long.

We address the problem of connectionless traffic management in ATM networks, considering both architectures shown in Figure 1. We propose a per-packet adaptive multipath routing mechanism that achieves efficient statistical multiplexing of connectionless traffic by incorporating a feedbackbased congestion control method. We present a scheme that distributes packets among multiple VPs according to feedback information on the traffic load on each VP. The remaining parts of the paper are organized as follows. In Section 2 we introduce the adaptive multipath routing scheme. In Section 3 we study the performance of the new method through extensive simulation experiments. In Section 4 we conclude the paper.

2 Adaptive Multipath Routing of Connectionless Traffic

We consider an architecture with Semi-Permanent Connections where each IWU can choose between different VPs for transmitting a packet. Cells of the same packet are transmitted on the same VP. The steps of our method are outlined as follows:

Step 1: Multiple VP Set-Up. We assign several semi-permanent VPs, placed on different routes, for each IWU-IWU pair that interconnects connectionless sources.

Step 2: Bandwidth Reservation for Connectionless Traffic. To prevent connection-oriented traffic from affecting the bandwidth available to connectionless traffic, a fixed amount of bandwidth is reserved for exclusive use by connectionless traffic. In our case, this amount of bandwidth is split among the semi-permanent VPs assigned to the source IWU-destination IWU pairs. The bandwidth reservation should be pessimistic to account for the burstiness of connectionless traffic. We propose to use the estimation method presented in [15].

Step 3: Adaptive Multipath Routing. At a source IWU, the arriving packet is probabilistically routed on any one of the multiple routes. The probability for choosing a particular VP depends on the relative utilization of the links on the route of the VP. The probabilities for the VPs are dynamically updated using the feedback mechanism suggested in [4].

The multipath routing algorithm is executed in the AAL layer of the IWUs. This means that each IWU maintains a table with all possible destinations, each associated with a VPI, in this layer. This table is used to set the VPI field of the cells into which a packet is segmented. In using the adaptive multipath routing an IWU takes the following actions:

- The IWU updates the table that associates each VP with the corresponding destination every time a packet is completely delivered to the ATM layer.
- The probabilities of the VPs are re-calculated periodically by feedback cells generated by the destination IWU. A feedback cell traverses the route of a VP in reverse order, collects information on the load the links on a VP, and makes the information available to the source IWU.

We can assume that these actions will take place after the segmentation, and before setting the VPIs of the cells of a new packet. The process of updating the VPI consists of simply checking the SAR Type (Segmentation And Reassembly Type) of the SAR_PDU header if AAL 3/4 is being used, or the PTI (Payload Type Indicator) of the ATM cell header if AAL 5 is being used [8]. The update procedure is executed when the last cell of a packet is delivered to the ATM layer.

Figure 2 shows a flow-chart describing the process of updating the VPI for the general case of n VPs per IWU–IWU pair. Assuming that the current VPI value in the routing table is α for a given destination D, as a result of running the algorithm, the new VPI value may remain α or change to β according to the current values of the n probabilities p(i) and the current value of a random variable r, which is uniformly distributed between 0 and 1.

Feedback Mechanism. For updating the routing probabilities we require that the IWUs receive congestion information from the network. Our metric for network congestion is the one suggested in [4], i.e., the maximum buffer occupancy at a switch encountered by a packet on its route. We propose a feedback procedure where control cell that indicate the congestion of a route are sent by the destination at periodic intervals. The payload of the control cells contains the maximum average utilization of the buffers at any switch on the route of a VP, where the averaging interval is given by the time interval between two consecutive control cell arrivals. When receiving a control cell, a switch compares its buffer utilization with the current contents of the control cell, and writes the greater value into the control cell.

3 Performance Evaluation

The efficiency of the multipath scheme has been investigated in a set of simulation experiments. A simulator has been implemented to model an ATM network and the traffic that it carries. We first give a detailed description of the simulated network model and then discuss the experiments that have been performed.

The simulated network has a single source IWU, a single destination IWUand $k \cdot n$ intermediate ATM switches. The IWUs and ATM switches are connected by n(k+1) links with a capacity of 150 Mb/s each. The source and the destination IWUs are linked by n parallel VPs, each with k intermediate switches. The VPs do not share any switches except the source and the destination switches. In the simulations, we consider two cases, a *two-paths* scenario with n = 2 parallel VPs and a *three-paths scenario* with n = 3parallel VPs. The VPs are denoted by VP1 and VP2 in the two-paths scenario, and by VP1, VP2, and VP3 in the three-paths scenario. In our simulations the number of intermediate switches, k, is set to k = 1, and each switch with an output buffer size of 50 cells. We distinguish two types of network traffic:

- *CL Source Traffic*: This class comprises traffic generated by individual connectionless sources. There is a fixed number of connectionless sources. Each connectionless source generates packets with an exponentially distributed size and an average size of $N_{CL} = 100 \ cells$. At the source switch a packet is fragmented into 53-byte ATM cells and
 - 7

allowed into the network at a rate of $R_{CL_{peak}} = 10 M b/s$. The interarrival time between two packets generated by a single connectionless source is exponentially distributed such that the overall traffic load of the source is $R_{CL_{ave}} = 50 K b/s$.

• Cross Traffic: This class represents the aggregate traffic from both connection-oriented as well as connectionless sources on a single link that connects two ATM switches or an ATM switch with an IWU. The cross traffic at each link is statistically independent of the traffic at any other link and is represented by a two-state model (HIGH state, LOW state). In the HIGH state cells are generated with an exponentially distributed interarrival time, such that the average traffic load is equal to the high average rate $R_{X_{high}} = 40 Mb/s$. The duration of the HIGH state is exponentially distributed with a mean of $T_{X_{high}} = 1.06$ secs. In the LOW state, cells are generated with an exponentially distributed interarrival time at an average traffic load equal to the low average rate, $R_{X_{low}} = 30 Mb/s$. The duration of the LOW state is exponentially distributed with a mean of $T_{X_{low}} = 1.41$ secs.

The destination IWU is responsible for initiating the feedback mechanism. It generates *feedback cells* after periodic time intervals, T_f , which traverse the VPs in the reverse direction, gathering information on the normalized queue lengths of the switches on the VPs as described in the previous section. The cross traffic parameters of the links of a single VP are assumed to be identical. We will refer to the cross traffic parameters of the two links collectively as the cross traffic parameters of the VP. For the experiments, a link whose *high average rate* is set to $R_{X_{high}} = 40 Mb/s$ is considered to be an uncongested link while a congested link has its *high average rate* set to $R_{X_{high}} = 120 Mb/s$. A VP with congested links in its path is considered to be a *congested VP*, and a VP without any congested link on its path is an *uncongested VP*.

Here, we show the results of three sets of experiments. The experiments compare the performance of the *single path option* against the *multipath*

$R_{CL_{peak}}$	$10 { m ~Mb/s}$
$R_{CL_{ave}}$	$50 { m ~Kb/s}$
N_{CL}	100 cells

Table 1: Connectionless Source Parameters.

	Uncongested Link	Congested Link
$R_{X_{high}}$	$40 \mathrm{~Mb/s}$	$120 \mathrm{~Mb/s}$
$R_{X_{low}}$	$30 { m ~Mb/s}$	$30 { m ~Mb/s}$
$T_{X_{high}}$	1.06 seconds	1.41 seconds
$T_{X_{low}}$	0.35 seconds	1.41 seconds

Table 2: Cross Traffic Parameters.

option. In the single path option, the network statically assigns each connectionless source to one of the n parallel VPs. All cells generated by the connectionless source must be routed along that dedicated VP. In the multipath option, the network routes all cells comprising a single packet probabilistically through one of the n parallel VPs. The performance measure in our experiments is the cell loss ratio (CLR) of the connectionless sources which is the proportion of connectionless cells that are dropped in the network. The simulation parameters used in our experiments are summarized in Tables 1 and 2.

3.1 Experiment I

In this experiment, we measure the sensitivity of the cell loss ratios to changes of the number of connectionless sources. The cross traffic load on the VPs is not identical. VP1 is on a set of links which are all congested and all other links are assume to be 'uncongested'. The number of connectionless sources is varied between $S_{CL} = 18$ and $S_{CL} = 36$ in the two-paths scenario, and between $S_{CL} = 18$ and $S_{CL} = 66$ in the three-paths scenario. The CLR values of the multipath option versus the single path option is

plotted for both the two-paths scenario and the three-paths scenarios. The results are shown in Figure 4. We observe that the multipath option consistently performs better, i.e., has a lower CLR value than the single path option. In the three-paths scenario, both options have similar CLR values initially, but with the increase in the number of connectionless sources, the multipath option is clearly seen to be superior. In the two-paths scenario, the multipath option always has a lower CLR value. The reason for this phenomenon is as follows. The average cross traffic load on the links in the path of VP1 is 120Mb/s when the links are in the HIGH state. The links on the paths of the other VPs, which are uncongested, have an average cross traffic load of only 40Mb/s in the HIGH state. Therefore, the traffic on VP1is more likely to face congestion than the traffic on the other VPs. In the multipath option, the network will periodically receive feedback information on the congestion levels at the VPs and update the probabilities of each VP. As long as congestion persists in VP1, the network will route cells from the connectionless sources through the uncongested VPs, thus avoiding the congestion on the path of VP1. In the single path option, the connectionless sources assigned to VP1 have to send cells through VP1 even when it is congested. Therefore, the single path option will suffer a higher cell loss rate compared to the multipath option.

Another point of interest is the different behavior of the two- and the three-paths scenario for the same S_{CL} values. We note that for $S_{CL} = 18$, for instance, the single path and the multipath options have the same CLR values in the three-paths scenario. But in the two-paths scenario, the single path option has a much higher CLR value as compared to the multipath option. The relatively higher CLR values of the single path option in the two-paths scenario, half of the bandwidth available for transmission of connectionless packets is subject to congestion, while in the three-paths scenario, only one-third of the bandwidth available for the transmission of packets experiences congestion. In the case for $S_{CL} = 18$, for instance, the two-paths scenario allocates nine sources on the congested path, whereas the three-paths scenario allocates

cates only six sources on the congested path. Thus, when considering the single path option in the two-paths scenario, a greater proportion of packets are subject to congestion as compared to the three-paths case, leading to a higher cell loss rate. The difference is not as pronounced in the multipath option because the feedback mechanism routes packets away from the VP in which congestion is experienced.

3.2 Experiment II

In the second set of experiments, we measure how the CLR changes with varying cross traffic at the links. In this experiment, the number of connectionless sources is fixed at $S_{CL} = 30$. The traffic rates of VP1 is set to $R_{X_{high}} = 120 M b/s$ and $R_{X_{low}} = 30 M b/s$. In the three-paths scenario, the values of $R_{X_{high}}$ of the other VPs, VP2 and VP3 is increased in steps from $R_{X_{high}} = 40 M b/s$ to $R_{X_{high}} = 120 M b/s$. The other parameters are as given in Tables 1 and 2. The results of the experiment are shown in Figure 5. We observe that as the cross traffic levels of VP2 and VP3 rise, the relative advantage of the multipath option over the single path decreases, i.e., the difference between the CLR values decreases. At $R_{X_{high}} = 120 Mb/s$, when the congestion levels of the VPs are equal, the two options have almost similar CLR values. The same effect is observed in the two-paths scenario. This experiment indicates that the multipath option is superior to the single path option only if the VPs have different congestion levels. In this case, the multipath option is able to route connectionless source traffic away from the VP where there is high congestion and into the VPs where there is low congestion. If the VPs are almost equally congested, then the multipath option cannot route connectionless source traffic away from congested VPs and the multipath option does not show any advantage over the single path option.

3.3 Experiment III

In this experiment, we study the transient behavior of the multipath option in order to demonstrate that the algorithm is inherently stable and does not cause an unbalanced network. We plot the probabilities of each of the three VPs as calculated by the feedback mechanism over a period of time. The setup for this experiment differs substantially from that of Experiments I and II. Firstly, the simulation is performed only for the multipath option using the three-paths scenario. Secondly, all VPs have identical rates of $R_{X_{high}} = 120 Mb/s$ and $R_{X_{low}} = 30 Mb/s$. Thirdly, the state transitions of the cross traffic on the links are now deterministic. Initially, all links are in the LOW state. In the experiment all links of a particular VP undergo state transitions simultaneously. The order of transitions is as indicated in Figure 6. The parameters for the connectionless sources are as given in Table 1. The number of connectionless sources is fixed at $S_{CL} = 42$. The simulation time is set to 0.35 seconds. The simulation results are shown in Figure 6:

- t = 0.00 sec: The cross traffic at all links start in the LOW state. As the traffic load on each of the VPs is the same, the feedback mechanism sets the routing probability of every VP to be equal to 0.33 on the average.
- t = 0.10 sec: The cross traffic at the links of VP2 goes to the HIGH state. This raises the traffic load on VP2, leading to congestion. The feedback mechanism detects the congestion and causes the routing probability of VP2 to be at 0 for most of the interval, and those of VP1 and VP3 to be at 0.5. The routing probabilities do not stay constant but are seen to fluctuate. This is due to the random arrival of cells at the switches and partly due to the adaptive nature of the multipath scheme, which diverts traffic away from the congested VP.
- t = 0.15 sec: The cross traffic at VP3 changes from the LOW state to the HIGH state. Now, both VP2 and VP3 have high traffic levels
 - 12

compared to VP1 and therefore experience congestion. The buildup of congestion at these VPs causes their routing probabilities to be at 0.17 during most of the interval, while that of VP1 is at 0.66.

- t = 0.20 sec: The cross traffic at VP1 makes a transition to the HIGH state. Now, in this interval all VPs are in the HIGH state and are equally likely to suffer from congestion. Therefore, their routing probabilities are at 0.33 for the major part of the interval. The variations of the probabilities are caused by the fact that the VPs do not receive cell bursts simultaneously. However, in this interval, the routing probabilities of all VPs take values in the same range (between 0 and 0.66).
- t = 0.25 sec: The cross traffic at VP1 and VP2 return to the LOW state. This situation is similar to that at $0.10 \le t \le 0.15$ sec. The routing probabilities of VP1 and VP2 fluctuate between 0.5 and 0.33, while that of VP3 fluctuates between 0 and 0.33.
- t = 0.30 sec: The cross traffic at VP3 returns to the LOW state. All links in the network are now uncongested. Accordingly, the routing probabilities of all VPs are now at 0.33 on the average.

The short term dynamic behavior of the feedback mechanism is shown in Figure 7, which plots the routing probabilities over a the time period $0.1 \leq t \leq 0.13 \ secs$. In the figure, the cross traffic at VP1 and VP3 are in the LOW state and that at VP2 is in the HIGH state. We observe that the routing probabilities of the three VPs alternate between two different sets of values. In one case, all the routing probabilities are equal to 0.33. In the other case, the routing probability of VP2 drops to 0 and the routing probabilities of VP1 and VP2 rise to 0.5 each. The reason for this behavior is as follows. Recall that the high cross traffic rate of VP2 causes it to experience congestion whenever a burst of cross traffic arrives at VP2. Then, the feedback mechanism reduces the routing probability of VP3 to 0.5 each. The network

now routes the connectionless source traffic only on VP1 and VP3 to avoid the congestion. Reducing the traffic load on VP2, allows the congestion at VP2 to be dissipated. The feedback mechanism recognizes that VP2 is no longer congested and resets the routing probabilities of all VPs to 0.33. The network now resumes routing connectionless source traffic through VP2. As congestion builds up in VP2 again, the process is repeated. It is important to note here that the congested VP2 never has routing probability that exceeds that of the uncongested VPs, VP1 or VP3.

4 Conclusions

Our proposal to support connectionless traffic in an ATM network is based on the approach of defining a virtual network of VPs overlaid on top of an ATM network. The management of the virtual network is performed by the IWUs themselves or by CLSs. In the present proposal, the virtual network is defined in such a way that every connectionless IWU–IWU pair is interconnected by at least two independent paths. A fixed amount of bandwidth is allocated among the VPs connecting the same IWU–IWU pair. We have proposed an adaptive multipath routing scheme that will detect congestion in the VPs and adapt the traffic distribution between the multiple VPs accordingly. The connectionless packets (CL_PDUs) between the IWU–IWU pairs is probabilistically routed between the alternate VPs, based on the traffic load on intermediate links. We have performed extensive simulation studies to show that adaptive multipath routing results in a lower cell loss ratio for connectionless traffic in the face of congestion as compared to static routing.

References

- G. Bianchi, F. Borgonovo, and V. Trecordi. An Optimal Bandwidth Allocation Algorithm for Remote Bridging of FDDI Networks Across B-ISDN. In Proc. IEEE Globecom'92, pages 1623-1627, 1992.
- [2] P. E. Boyer and D. P. Tranchier. A Reservation Principle with Applications to the ATM Traffic Control. Computer Networks and ISDN Systems, 24:321-334, 1992.
- [3] M. Gerla, T.-Y. C. Tai, and G. Gallassi. Internetting LANs and MANs to B-ISDN for Connectionless Traffic Support. *IEEE Journal on Selected Areas in Communications*, 11(8):1145-1159, October 1993.
- [4] Y. Gong and I. F. Akyildiz. Dynamic ATM Traffic Control Using Feedback and Traffic Prediction. In Proc. IEEE Infocom'94, pages 91-99, June 1994.
- [5] C. Ikeda and H. Suzuki. Adaptive Congestion Control Schemes for ATM-LANS. In Proc. IEEE Infocom'94, pages 829-838, June 1994.
- [6] R. Krishnan and J.A. Silvester. Choice and Allocation Granularity in Multipath Source Routing Schemes. In Proc. IEEE Infocom'93, pages 322-329, June 1993.
- [7] H. T. Kung. Use of Link by Link Flow Control in Maximizing ATM Network Performance: Simulation Results. In Proc. of the IEEE Hot Interconnects Symposium'93, August 1993.
- [8] L. Mongiovi, M. Farrell, and V. Trecordi. A Proposal for Interconnecting FDDI Networks Through B-ISDN. In Proc. IEEE Infocom'91, pages 1160-1167, April 1991.
- [9] P. Newman. ATM Local Area Networks. *IEEE Communications Mag-azine*, 32(3):86-89, March 1994.

- [10] H. Suzuki and F. Tobagi. Fast Bandwidth Reservation Scheme with Multilink and Multipath Routing in ATM Networks. In Proc. IEEE Infocom'92, pages 2233-2240, May 1992.
- [11] N. Yin and M. G. Hluchyj. On Closed-Loop Rate Control for ATM Cell Relay Networks. In Proc. IEEE Infocom'94, pages 99–108, June 1994.



(a) Semi-Permanent Connections





Figure 1: Approaches to LAN Interconnection in ATM Networks.

Figure 2: Adaptive Multipath Routing Algorithm (Number of VPs per IWU is n).



Figure 3: The ATM Network Model (Three-Paths Scenario).



(a) with two parallel VPs



(b) with three parallel VPs

Figure 4: Effect of Varying Source Traffic.



Figure 5: Effect of Varying Cross Traffic.



Figure 7: Transient Behavior of the ${}^{2}P$ robabilities (Enlargement from Figure 6 for the Time Interval $[0.1 \le t \le 0.13]$).