Bandwidth allocation, traffic control and topology design in ATM overlay networks*

José Augusto Suruagy Monteiro†, Shivi Fotedar† and Mario Gerla†
Departamento de Informática, Universidade Federal de Pernambuco, Brazil†
Computer Science Department, University of California, Los Angeles†

Abstract

The Virtual Path (VP) feature of ATM allows the implementation of "virtual" overlay networks which can be used for various applications. The design of these overlay networks poses new challenges in the areas of bandwidth allocation, congestion control and topology design. In this paper we examine two specific examples of overlay networks: Broad-band Virtual Private Networks (BVPNs) and Connectionless Server (CLS) networks. For BVPNs, we attack the problem of bandwidth allocation and congestion control, and compare two different policing strategies. For CLS nets, we pose the problem of topology design, and propose a heuristic solution.

1 Introduction

Bandwidth allocation and, more generally, topology design of a wide area ATM network is quite different from that of a conventional packet network. In a packet network, topology design is a "one step" problem: given the requirements (traffic, performance) and the trunk options (capacity and cost), we configure the topology so as to minimize a well defined measure (cost or delay), within given constraints. In an ATM network, the topology design must be carried out in multiple steps and is closely coupled with bandwidth allocation.

The first step involves "embedding" ATM trunks on a network of fiber facilities. These facilities are interconnected by digital cross connect switches (DCS), thus allowing the establishment of circuits between any pair of sites. The second step requires establishing VP subnetworks in an ATM network. A Virtual Path (VP) can be viewed as a "pipe" into which several user data streams (VCs or datagrams) are multiplexed. The VP facility in the ATM network allows us to define overlay VP subnetworks called Virtual Private Networks (VPNs). A VP link in a VPN is typically established end to end between two sites with peak bandwidth assigned to it. Such a Virtual Private Network is referred to as an End-to-End Virtual Private Network (EEVP).

The simple and most common VPN solution today is end-to-end VP networking (EEVP). Unfortunately, EEVPs, do not scale very well to configurations with large numbers of customer sites. This is because the bandwidth is fragmented among a number of VPs which grow with the square of the number of customer sites. To overcome this problem, Bellcore recently proposed a new type of VPN, called Broadband Virtual Private Network (BVPN) [WA92]. In this case, the end points of the VP links are not limited to customer sites, but also include ATM nodes (Virtual Channel Cross Connects). Thus, the BVPN topology is a mesh topology, with links consisting of VP pipes and, nodes which include customer sites and ATM cross connects.

Another important example of VP subnetwork embedding is the Connectionless Server (CLS) overlay network. The CLS network consists of CLS datagram switches installed at some of the ATM sites. The CLSs are connected by VP pipes derived from the underlying ATM network. The CLS network carries all the connectionless traffic submitted to ATM, while the ATM network carries exclusively connection oriented traffic. Thus, the CLS overlay network is embedded in the ATM network, but operates independently from it, except for sharing ATM trunk bandwidth.

It should be clear from the above description that the topology design and bandwidth allocation problem in an ATM network must be solved at various different levels. In this paper, we focus on the issues involved in the design and bandwidth allocation of ATM overlay networks. More specifically, we address the problem of bandwidth allocation and congestion control in BVPNs, and discuss the topology design of CLS overlays. For each case, we propose solution approaches and illustrate them with simple examples.

2 Bandwidth allocation in BVPNs

As we mentioned before, the BVPN is a virtual subnet embedded in ATM, which consists of VP links connecting user sites and ATM nodes as shown in Fig. 1. As in the EEVP strategy the user is allocated peak bandwidth on the VPs. However, the VCs are not confined to a single end to end VP, but may traverse several VPs. During the data transfer phase a temporary association of bandwidth to the VCs of the logical network is realized. Thus, in this case, the net-
work provides not only VP cross connection, but also VC cross connection. The BVPN solution offers considerable advantages to users: first, each traffic stream, between different source/destination sites, can statistically share a common VP when their paths overlap; secondly, traffic pattern changes can be more easily absorbed by a well designed VP topology, without requiring dynamic reallocation of bandwidth, as in the EEVP strategy.

In this approach, the ATM node must perform some extra work: first, the ATM node, an ATM Cross Connect (XC) for instance, must cross connect not only VPs, but also VCs. So, VC routing maps must be maintained. Another important concern is VP traffic policing which is required at each VP segment. Furthermore, VP policing cannot protect the ATM XC from internal congestion caused by changes in traffic patterns. For instance, consider the ATM XC "A" in Fig. 1. Assume that initially the traffic into and out of node A is balanced, and therefore bandwidth allocation is the same value for VP 1, 2, 3 and 4. Accordingly, the input VP policing parameters are set to the same peak value. Because of traffic fluctuations, it is possible that during a period all traffic entering from VP 1 and 2 goes out to VP 4. Clearly this overload will cause congestion in VP 4. Yet, "output" policing on VP 4 can not protect the node from congestion (i.e., internal buffers may overflow, thus, degrading QOS guaranteed to other ATM users). Thus, more elaborate traffic control policies are required. Two different schemes have been suggested in [Croc+94] to solve the problem of traffic congestion, namely Fast Resource Management (FRM) and Intra-Node Policing (INP). These two schemes consider a public network, which consists of ATM XC nodes and of an ATM network manager, and a private user which has at each site a Local Customer Manager (LCM) and, optionally, a Central Customers Manager (CCM) to coordinate the operations of the LCMs.

2.1 Fast Resource Management for BVPN

In the FRM strategy the user requests for peak bandwidth allocation for each incoming burst of traffic or connection. When a burst arrives at the user's LCM, a fixed reservation cell is issued on the proper VC. The first ATM XC on the path verifies that the request can be accepted on the incoming and outgoing VPs (i.e., it keeps track of user bandwidth allocation on VP links). Then, it forwards the fast reservation request to the next node along the path, which in turn, verifies that the request can be accepted on the next VP. In this first phase a bandwidth booking procedure is performed at each node. If the reservation is not blocked, the last node in the path returns an acknowledgment to the origin, reserving bandwidth on the way back and updating the input VC policer in the ingress node. The source is notified of acceptance which then issues the burst. More detailed examples of protocols can be found in [Croc+94] [TBRM92].

This scheme clearly protects the ATM XC from congestion, since it polices all the VCs individually and thus assures that the allocated VP bandwidth is never exceeded. It also supports dynamic user bandwidth allocation. However, the round trip delay and processing overheads degrade efficiency. Further, the procedure is non transparent to ATM (i.e., the ATM XC must keep track of user bandwidth allocation). Lastly, additional hardware and software is required in the ATM XC.

2.2 Intra-Node Policing for BVPN

Consider Fig. 2 and assume that the input traffic from VP 1 splits into rates R13 and R14 directed to VP 3 and VP 4 respectively. Similarly, VP 2 splits into R23 and R24. The condition to avoid internal node (i.e., ATM XC) congestion is that R13 + R23 < peak bandwidth allocated to VP 3. Ideally, we would like to jointly police the sum R13 and R23. Unfortunately, this is not possible since these two traffic streams enter from two different exchange terminations.

Policing requires counting cells in real time, and therefore it can be performed only on the exchange termination on which the traffic stream is entering. Therefore, a more realistic solution entails the policing of R13 and R23 separately. This policing strategy, which we call Intra-Node Policing (INP) is rather easy to implement. It simply requires keeping a common policing counter (on VP1, say) for all the VCs which are routed from VP 1 to VP 3. If R13 and R23 are chosen so that R13 + R23 < peak bandwidth allocated on VP 3, nodal congestion is prevented. Thus the scheme is fail-safe. However, from the user performance standpoint, this scheme is rather conservative, since it does not allow statistical multiplexing of the streams corresponding to R13 and R23. Conceptually, it is as if artificial "Intranode-Virtual Path" (I-VP) links were introduced between the original internodal VPs. These I-VPs clearly pose further constraints on user bandwidth allocation, and may, in some cases, become undesirable bottlenecks, unless properly chosen and updated. This may be corrected by allowing the user to change the I-VP policer parameters.

2.3 Bandwidth allocation example

A quantitative comparison of the above mentioned bandwidth allocation strategies was carried out for several representative virtual topologies. For the sake of conciseness, we report here only the experiments on an "NSFnet-like" mesh topology (Fig. 3). More extensive results are found in [Mon+94].

Traffic requirements were assumed to be uniform between all node pairs. More specifically, each node pair has a session with peak bandwidth Bp, burstiness b and mean burst length L. Burst length and interarrival time were assumed to be exponentially distributed, with the mean burst length, L, equal to 100 cells. Quality of service (QOS) was defined to be the burst loss probability Pb, which for our purpose is more meaningful than the conventional cell loss rate. In our examples, we defined Pb to be less than or equal to 10⁻⁴. In the NSF-like topology, we assume that all the ATM nodes can act as ATM XCs for the BVPN overlay. The perfor-
Performance measure used to compare the various schemes is the aggregate bandwidth of all the VPs over the entire network. The evaluation was carried out both analytically [Croc‘93] and through simulation (see Fig. 4).

3 Topology design in CLS overlays

The need for connectionless service in ATM stems primarily from the interconnection of local and metropolitan area networks (respectively, LANs and MANs) across a wide geographical area. The Switched Multi-megabit Data Service (SMDS) [B91] [ETSI93] was specifically introduced to support this high speed interconnection. The SMDS/ATM solution, however, is not straightforward. The main challenge stems from the fact that ATM is a connection-oriented network whereas LANs, MANs and SMDS run connectionless protocols. This mismatch can be overcome by providing a connectionless service data “on top of ATM”.

One of the schemes for providing connectionless service in ATM is the Connectionless Virtual Overlay Network. Special purpose switches called Connectionless Servers (CLSs), are in charge of storing, routing and forwarding datagrams across the ATM network. The connectionless servers (CLSs) are installed at various ATM local exchange nodes as well as at ATM VP cross connect nodes. IWUs are connected to one or more CLSs via dedicated VPs. The CLSs, in turn, are interconnected with each other by a “virtual overlay network” consisting of several permanent VP with preallocated bandwidth [LP91]. The overlay network does not necessarily include a VP for each CLS pair (i.e., full mesh). Thus, datagrams may traverse several CLSs on the path from source to destination.

Parameters related to topology and bandwidth allocations, in a CLS network are the number and location of CLS facilities, and overlay topology connectivity. The objective of the topology design is to minimize average cell delay in the CLS overlay, with the constraint that the aggregate CLS VP bandwidth on an ATM trunk does not exceed residual bandwidth.

In the next section, we attack the problem of topology design of the CLS overlay network. As we shall see, this problem also includes bandwidth allocation (to VPs) and routing (of VPs as well as of traffic on multiple VPs) as subproblems. The CLS node location problem is not specifically addressed. This problem, however, can be solved heuristically by solving a sequence of topology problems, each with a different set of nodes, as shown in the example below.

3.1 CLS topology design example

In this section we present an example which illustrates the application of our solution technique and tools to ATM-VP allocation (CLS topology design).

Let us consider the ten node ATM network shown in Fig. 5, with traffic requirements given in Table 1. In our example, each ATM trunk is assumed to have 150 units of capacity which is entirely available for CLS traffic. Without loss of generality, both traffic requirements and trunk capacities are expressed in the same units.

In this ATM network, we can define a number of CLS overlay networks. One of the key design decisions is the selection of the number and locations of CLSs. With fewer CLSs, we can use VPs with larger bandwidth to interconnect them, and therefore, achieve a higher multiplexing gain. On the negative side, fewer CLSs imply longer IWU to CLS paths, and higher datagram traffic at each CLS. By increasing the number of CLSs, on the other hand, we reduce the multiplexing gain, and incur a higher cost. A good compromise (in the number of CLS nodes) can be found by examining several cases with an automated design tool.

As we mentioned before, the reduction in the number of CLSs increases path lengths, and therefore, increases total network traffic as well as overall trunk utilization. We can express this overall trunk utilization by the inverse of the utilization of the most saturated VP, which we call Maximum Throughput increase factor. In other words, if we increase traffic by an uniform factor, this is the factor which brings the first VP into saturation. A related measure is the Congestion Measure (average VP queueing delay). As one can see in Table 2, three experiments were run, with 2,3 and 5 CLS nodes respectively. In addition, a 5 CLS node “multiple homed” experiment was run, where each IWU can be connected to multiple CLS nodes, for load balancing (whereas in the previous runs each IWU is connected only to the nearest CLS). Fig. 6 shows that 5 CLS single homed topology. The results in Table 2 clearly indicate that an increase in number of CLS nodes reduces congestion and increases throughput. As expected, the multiple home solution performs better than the single home, because of load balancing. Clearly, an increase in CLS nodes leads also to higher costs. Thus, a more general optimization (which goes beyond the scope of our study) should consider a tradeoff between cost and delay.

4 Conclusions

ATM overlay networks offer novel solutions for the support of specialized applications (for example, connectionless traffic) or specialized traffic patterns (for example, Virtual Private Networks) over a public ATM network. Congestion control in an ATM overlay network is greatly simplified due to the fact that the traffic in the overlay network is segregated from the remaining traffic (that is, it is transported on separate, peak allocated VPs).

The topology design of overlay networks poses novel problems. One important variable is the number of internal nodes (that is, terminations of VP pipes) used in the virtual topology. We have described a combination of automated tools and heuristics which can efficiently solve this problem.

References


Fig. 4. Link Bandwidth versus burstiness in NSF like topology. (Bp = 10Mbps; N = 13).

Table 1: Traffic Requirements

<table>
<thead>
<tr>
<th>Nodes</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>5</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>-</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>3</td>
<td>-</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>-</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>-</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>-</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>-</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>-</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>-</td>
<td>5</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 2: Summary of Results.

<table>
<thead>
<tr>
<th>Case</th>
<th>Congestion Measure (Avg. delay)</th>
<th>Max. Throughput increase factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple homed</td>
<td>698.25</td>
<td>5.134</td>
</tr>
<tr>
<td>5 CLS</td>
<td>849.58</td>
<td>3.375</td>
</tr>
<tr>
<td>3 CLS</td>
<td>867.90</td>
<td>2.547</td>
</tr>
<tr>
<td>2 CLS</td>
<td>992.38</td>
<td>1.891</td>
</tr>
</tbody>
</table>

Figure 5: 10-node ATM network.

Figure 6: 5 CLS network.