CHAPTER 1

INTRODUCTION

1.1 ASYNCHRONOUS TRANSFER MODE (ATM) NETWORKS

With the convergence of telecommunication, entertainment and computer industries, computer networking is adopting a new paradigm called Asynchronous Transfer Mode (ATM) (Black 1995, Garry 1995). ATM was selected by the telecommunication (carrier industry) as the technology to deliver the Broadband Integrated Services Digital Network (B-ISDN) carrier service. ATM is designed to handle different kinds of communication traffic (voice, audio, video and data) in an integrated way. It is the first technology to promise seamless interworking between the LAN and WAN network environments. The international standards for ATM networks are being formulated by the ATM Forum (Forum 1996) and ITU-T (www.itu.ch).

ATM uses short fixed size (53-byte packets) called “cells” which is an attractive option because: a) the transmission time per cell is fixed (which reduces the variability in queuing delays) and b) the transmission time is small (which allows building pipelined hardware architectures to process cells in switches). The resulting low mean delay, and low delay variance characteristics are the features that facilitate cell-based voice and video transmissions. However, each cell has five bytes (or 9.43% data transfer) header information which limits the maximum possible efficiency of data transmission, especially on LANs. Further, the loss of one cell results in the loss of an entire packet (which may consist of several cells). But the cell switching (as opposed to expensive packet routing) and sophisticated traffic
management technology in ATM networks allows the real efficiency to reach the maximum possible (unlike the Ethernet technology where the efficiency drops off rapidly as load increases). This feature makes ATM attractive for data communications as well.

The development of the ATM technology has also resulted in several elegant total or compromise solutions to facilitate high-speed integrated services networking. These include the use of shared switches (as opposed to using shared media), connection-oriented technology (delivery guarantee, and simplified management and control), the use of short switch-assigned labels in cell headers instead of addresses (for scalability), the development of a true QoS-based routing (PNNI) protocol, and introduction of features such as LAN Emulation (LANE) and Multiprotocol over ATM (MPOA) which has triggered off work in the field of internetworking (running technology “X” over technology “Y”) (Harry 1995).

In this thesis the focus is on the problem of supporting data applications efficiently within the integrated services framework. In addition to providing a viable solution for any one of voice, video, or data transmission in isolation, ATM allows all these applications to be supported efficiently in a single network. This is a key difference when compared with current data network technologies like Ethernet. This feature, when complemented with traffic management capabilities, allows the integrated network to be fully utilized while delivering the quality of service requested by applications.

1.2 AVAILABLE BIT RATE (ABR) SERVICE

ATM networks provide multiple classes of service to support the quality of service (QoS) requirements of diverse applications, (ATM Forum 1996). The current set of classes specified are the 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). The CBR service is aimed at supporting voice and other synchronous applications, the VBR (rt- and nrt-) services are designed to support video and audio applications (which do not need isochronous transfer), while the ABR and UBR services are designed to primarily support data applications.

![Diagram of ATM ABR and VBR Traffic Sharing a Link](image)

**Figure 1.1 ATM ABR and VBR Traffic Sharing a Link**

Typically, the CBR and VBR classes are assigned higher “priority” by the network switches and get a share of the link bandwidth first. The “left-over” capacity is used by the ABR and UBR services, with ABR typically having priority over UBR. In Figure 1.1 we show a link being shared by a “higher priority” VBR class and a “lower priority” ABR class. Note that VBR and ABR cells are queued separately.

The ABR service class includes an elaborate traffic management framework which allows the efficient handling of data traffic. On the other hand, there exists no standard method of managing traffic on the UBR service. Switches can provide proprietary traffic management mechanisms for UBR, but they cannot coordinate with other switches since a standard does not exist. In the next section, a traffic management problem is defined and discussed its role in the success of ATM as an integrated services networking technology.
1.3 TRAFFIC MANAGEMENT VS CONGESTION CONTROL

A key issue in ATM and in any other network architecture design is resource management, i.e., how to make the best use of available resources. Maintaining high utilization of resources while satisfying the users’ traffic contracts is the only way the high investment on the networking infrastructure can be recouped. However, striving for high utilization of a resource without proper allocation may lead to long queuing delays, and losses resulting in a low throughput (degradation of user-perceived quality of service).

Traffic management is a resource management problem which deals exclusively with the mechanisms required to control traffic on the network. A related problem is “congestion” which occurs when the aggregate demand for a resource (typically link bandwidth) exceeds the available capacity of the resource. In other words, congestion happens whenever the demand is more than the available capacity.

\[ \sum_i \text{Demand}_i > \text{Available Capacity} \]

There are two sets of mechanisms to handle congestion. “Congestion control” mechanisms typically come into play after the network is overloaded, i.e., congestion is detected. “Congestion avoidance” mechanisms come into play before the network becomes overloaded. i.e., congestion is predicted. “Congestion management” is a term used to denote the combination of congestion avoidance and control mechanisms (Raj Jain 1992).

Congestion management involves the design of mechanisms and schemes to statically limit the demand-capacity mismatch or dynamically control traffic sources when such a mismatch occurs. Congestion is a problem
associated with the dynamics of the network load and capacity; it has been shown that static solutions such as allocating more buffers, or providing faster links, or faster processors does not solve the problem (Raj Jain 1992,1990). In fact, the partial deployment of these static alternatives has led to more heterogeneity in the network and increased the possibility of congestion.

Observe that congestion management deals with the problem of matching the demand and capacity for a single network traffic class. Traffic management, even for a single traffic class, deals with the problem of ensuring that the network bandwidth, buffer and computational resources are efficiently utilized while meeting the various Quality of Service (QoS) guarantees given to sources as part of a traffic contract. The general problem of network traffic management involves all the available traffic classes. In ATM networks, the general traffic management problem involves the mechanisms needed to control the multiple classes of traffic (like CBR, VBR, ABR and UBR) while ensuring that all the traffic contracts are met. The components of traffic management other than congestion management schemes include scheduling mechanisms, traffic contract negotiation, admission control, and traffic policing. In this dissertation, designing traffic management mechanisms for one class—the ABR service class in ATM networks in the problem for discussion.

Historically, traditional data networks support only one class of service (data). In such networks, the term “traffic management” was synonymous with “congestion control”. In passing, also note the difference between “flow controls” and “congestion control” that is flow control deals with the control of a particular flow, whereas congestion control deals with the control of a group of flows sharing a group of network resource. It is possible to design congestion control schemes which essentially control flows individually at every hop. This makes the problem similar to flow control. An
example of such a design is the hop-by-hop flow-controlled virtual circuit (Kung 1994) or credit-based framework proposal for ATM discussed later in the dissertation.

1.3.1 Traffic Management and QoS Attributes

In order for ATM networks to deliver guaranteed quality of service on demand while maximizing the utilization of available network resources, effective traffic management mechanisms are needed. Almost every aspect of ATM network operation, from signaling requests and routing to network resource allocation and policing, contains some traffic management mechanisms.

While setting up a connection on ATM networks, users can specify the following parameters related to the desired quality of service:

- **Peak Cell Rate (PCR):** The maximum instantaneous rate at which the user will transmit. For bursty traffic, the inter-cell interval and the cell rate varies considerably. The PCR is the inverse of the minimum inter-cell interval.

- **Sustained Cell Rate (SCR):** This is the average rate as measured over a time interval.

- **Cell Loss Ratio (CLR):** The percentage of cells that are lost in the network because of error congestion and are not delivered to the destination, i.e.,

  \[
  CLR = \frac{\#\text{Lost Cells}}{\#\text{Transmitted Cells}} \times 100
  \]

  Each ATM cell has a cell loss priority (CLP) bit in the header. During periods of congestion, the network will first discard cells with
CLP = 1. Since the loss of cells with CLP = 0 is more harmful to the operation of the application, CLR can be specified separately for cells with CLP = 1 and for those with CLP = 0.

- Cell Transfer Delay (CTD): The delay experienced by a cell between network entry and exit points is called the cell transfer delay. It includes propagation delays, queuing delays at various intermediate switches, and service times at queuing points.

- Cell Delay Variation (CDV): This is a measure of variance of CTD. High variation implies larger buffering for delay sensitive traffic such as voice and video.

- Burst Tolerance (BT): This determines the maximum burst size that can be sent at the peak rate. This is the bucket size parameter for the leaky bucket algorithm that is used to control the traffic entering the network. The algorithm consists of putting all arriving cells in a buffer (bucket) which is drained at the sustained cell rate (SCR). The maximum number of back-to-back cells that can be sent at the peak cell rate is called maximum burst size (MBS). BT and MBS are related as follows:

\[
BT = (MBS - 1) \left( \frac{1}{SCR} - \frac{1}{PCR} \right)
\]

- Minimum Cell Rate (MCR): This is the minimum rate desired by a user. The first six of the above traffic parameters were originally specified in UNI version 3.0. MCR has been added recently and will appear in the next version of the traffic management document.
1.3.2 Traffic Contract

To provide a guaranteed QoS, a traffic contract is established during connection setup, which contains a connection traffic descriptor and a conformance definition. However, it is not necessary for every ATM virtual connection to have a specified QoS. The reason for this is that if only specified QoS connections are supported by ATM, then a large percentage of the network resources will be wasted. This can happen when one or more connections are not utilizing the full capacity of their QoS contracts. Unspecified QoS contracts can be supported by an ATM network on a best-effort basis. Such best-effort services are sufficient for supporting most of the existing data applications.

In general, a traffic contract specifies one of the following classes of traffic:

- **Constant Bit Rate (CBR):** This class is used for emulating circuit switching, where the bit rate is constant. Cell loss ratio is specified for cells with CLP=0 and may or may not be specified for cells with CLP =1.

- **Variable Bit Rate (VBR):** This class allows users to send at a variable rate. Statistical multiplexing is used and so there may be small nonzero random loss. Depending upon whether or not the application is sensitive to cell delay variation, this class is subdivided into two categories: real-time VBR (VBR-RT) and non real-time VBR (VBR-NRT). In both categories cell transfer delay is specified. CDV is specified only for real-time VBR. An example of real-time VBR is interactive compressed video and a non real-time VBR is multimedia email.
- Available Bit Rate (ABR): This class is designed for normal data traffic such as file transfer and email. Although the standard does not require the cell transfer delay and cell loss ratio to be guaranteed, it is desirable for switches to minimize the delay and loss as much as possible.

Depending upon the congestion state of the network, the source is required to control its rate. The users are allowed to declare a minimum cell rate (MCR) that is guaranteed to the VC by the network. Most VCs will ask for an MCR of zero. Those with higher MCR may be denied connection if sufficient bandwidth is not available.

- Unspecified Bit Rate (UBR): This class is designed for those data applications that want to use any left-over capacity and are not sensitive to cell loss or delay. Such connections are not rejected on the basis of bandwidth shortage (i.e., no connection admission control) and not policed for their usage behavior. During congestion, the cells are lost but the sources are not expected to reduce their cell rate. Instead, these applications may have their own higher-level cell loss recovery and retransmission mechanisms. Examples of applications that use this service are email and file transfer. Of course, these applications can use the ABR service, if desired.

ABR or UBR are usually specified in the traffic contract when the ATM network is providing a best-effort service. Thus, these two classes of traffic are sometimes referred to as best-effort traffic. The QoS parameters for the above classes of traffic are summarized in Table 1.1.
Table 1.1 ATM Layer Service Categories

<table>
<thead>
<tr>
<th>Attribute</th>
<th>CBR</th>
<th>VBR-RT</th>
<th>VBR-NRT</th>
<th>ABR</th>
<th>UBR</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLR for CLP=0</td>
<td>Specified</td>
<td>Specified</td>
<td>Unspecified</td>
<td>Specified</td>
<td>Unspecified</td>
</tr>
<tr>
<td>CLR for CLP=1</td>
<td>Optional</td>
<td>Specified</td>
<td>Unspecified</td>
<td>Specified</td>
<td>Unspecified</td>
</tr>
<tr>
<td>CTD</td>
<td>Specified</td>
<td>Specified</td>
<td>Unspecified</td>
<td>Unspecified</td>
<td>Unspecified</td>
</tr>
<tr>
<td>CDV</td>
<td>Specified</td>
<td>Unspecified</td>
<td>Unspecified</td>
<td>Unspecified</td>
<td>Unspecified</td>
</tr>
<tr>
<td>SCR and BT</td>
<td>Not applicable</td>
<td>Specified</td>
<td>Not applicable</td>
<td>Not applicable</td>
<td></td>
</tr>
<tr>
<td>PCR and CDVT</td>
<td>Specified</td>
<td>Specified</td>
<td>Specified</td>
<td>Specified</td>
<td>Specified</td>
</tr>
<tr>
<td>MCR</td>
<td>Not applicable</td>
<td>Specified</td>
<td>Not applicable</td>
<td>Not applicable</td>
<td></td>
</tr>
</tbody>
</table>

1.3.3 Congestion Control Techniques

Congestion control lies at the heart of the general problem of traffic management for ATM networks. In general, congestion arises when the incoming traffic to a specific link is more than the outgoing link capacity. The primary function of congestion control is to ensure good throughput and delay performance while maintaining a fair allocation of network resources to the users. For unspecified QoS traffic such as ABR service, whose traffic patterns are often highly bursty and unpredictable, congestion control poses more challenges than for other services. As described (Raj Jain 1990), one way to classify congestion control schemes is based on the layer of International Organization for Standardization (ISO) / Open Secure Interconnection (OSI) reference model at which the scheme operates. For example, there are data link, routing, and transport layer congestion control schemes. Typically, a combination of such schemes is used. The selection depends upon the severity
and duration of congestion. Figure 1.2 shows how the duration of congestion affects the choice of the method.

One method to avoid network congestion is to accept a new ATM connection during connection setup phase only when sufficient network resources are available to provide the acceptable QoS. This is called connection admission control (CAC), which is needed for connections where the QoS must be guaranteed. The “busy” tone on telephone networks is an example of CAC. Mechanisms for CAC are currently not standardized and are at the discretion of the network operators.

<table>
<thead>
<tr>
<th>Congestion Duration</th>
<th>Congestion Mechanism</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long</td>
<td>Capacity planning and network design</td>
</tr>
<tr>
<td></td>
<td>Connection admission control</td>
</tr>
<tr>
<td></td>
<td>Dynamic routing</td>
</tr>
<tr>
<td></td>
<td>Dynamic compression</td>
</tr>
<tr>
<td></td>
<td>End-to-end feedback</td>
</tr>
<tr>
<td></td>
<td>Link-by-link feedback</td>
</tr>
<tr>
<td>Short</td>
<td>Buffering</td>
</tr>
</tbody>
</table>

**Figure 1.2 Congestion Techniques for Various Congestion Durations**

The queue behavior in detail is studied first and then show how and to what degree queues can build up. This motivates the use of a mechanism that manages queue buildup in concert with the rate allocation mechanism for the operation of explicit rate based congestion control. Operate the network with small queues, to insure so as to minimum packet loss, to provide reasonably low (not a strict bound) delay, and keep the feedback delay small (especially if RM cells are processed in-band). Related work has been described in the recent past to manage queue build-up (De Prycker et al
1993). In broad terms, most of the techniques suggested achieve low buffer occupancy by keeping the utilization of the link below the full capacity (e.g., having a target utilization of 95%). Aim is to maintain the link utilization at the maximum (i.e., 100%), except when one need to reduce the buildup queue. Moreover, one attempt to do this while maintaining exact max-min fairness and the feature of constant-time computation of max-min fairness described in (Henderson 1997).

1.4 DYNAMIC ROUTING POLICIES (STATIC VS DYNAMIC)

Static routing algorithms are hardly algorithms at all, but are table mappings established by the network administrator prior to the beginning of routing. These mappings do not change unless the network administrator alters them. Algorithms that use static routes are simple to design and work well in environments where network traffic is relatively predictable and where network design is relatively simple.

Because static routing systems cannot react to network changes, they are generally considered unsuitable for today large and changing networks. Most of the dominant routing algorithms in the 1990s are dynamic routing algorithms, which adjust to changing network circumstances by analyzing incoming routing update messages. If the message indicates that a network change has occurred, the routing software recalculates routes and sends out new routing update messages. These messages permeate the network, stimulating routers to rerun their algorithms and change their routing tables accordingly.

Dynamic routing algorithms can be supplemented with static routes where appropriate. A router of last resort (a router to which all unroutable packets are sent), for example, can be designated to act as a repository for all
unroutable packets, ensuring that all messages are at least handled in some way

1.4.1 Routing Metrics

Routing algorithms have used many different metrics to determine the best route. Sophisticated routing algorithms can base route selection on multiple metrics, combining them in a single (hybrid) metric. All the following metrics have been used:

- Path Length
- Reliability
- Delay
- Bandwidth
- Load
- Communication Cost

Path length is the most common routing metric. Some routing protocols allow network administrators to assign arbitrary costs to each network link. In this case, path length is the sum of the costs associated with each link traversed. Other routing protocols define hop count, a metric that specifies the number of passes through internetworking products, such as routers, that a packet must take en route from a source to a destination.

Reliability, in the context of routing algorithms, refers to the dependability (usually described in terms of the bit-error rate) of each network link. Some network links might go down more often than others. After a network fails, certain network links might be repaired more easily or more quickly than other links. Any reliability factor can be taken into account in the assignment of the reliability ratings, which are arbitrary numeric values usually assigned to network links by network administrators.
Routing delay refers to the length of time required to move a packet from source to destination through the internetwork. Delay depends on many factors, including the bandwidth of intermediate network links, the port queues at each router along the way, network congestion on all intermediate network links, and the physical distance to be traveled. Because delay is a conglomeration of several important variables, it is a common and useful metric.

Bandwidth refers to the available traffic capacity of a link. All other things being equal, a 10-Mbps Ethernet link would be preferable to a 64-kbps leased line. Although bandwidth is a rating of the maximum attainable throughput on a link, routes through links with greater bandwidth do not necessarily provide better routes than routes through slower links. If, for example, a faster link is busier, the actual time required to send a packet to the destination could be greater.

Load refers to the degree to which a network resource, such as a router, is busy. Load can be calculated in a variety of ways, including CPU utilization and packets processed per second. Monitoring these parameters on a continual basis can be resource-intensive in itself.

Communication cost is another important metric, especially because some companies may not care about performance as much as they care about operating expenditures. Even though line delay may be longer, they will send packets over their own lines rather than through the public lines that cost money for usage time.

1.4.2 Routing Hierarchy

Unlike Routing Information Protocol (RIP), Open Shortest Path First (OSPF) can operate within a hierarchy. The largest entity within the
hierarchy is the autonomous system (AS), which is a collection of networks under a common administration that share a common routing strategy. OSPF is an intra-AS (interior gateway) routing protocol, although it is capable of receiving routes from and sending routes to other ASs.

An AS can be divided into a number of areas, which are groups of contiguous networks and attached hosts. Routers with multiple interfaces can participate in multiple areas. These routers, which are called area border routers, maintain separate topological databases for each area.

A topological database is essentially an overall picture of networks in relationship to routers. The topological database contains a collection of LSAs received from all routers in the same area. Because routers within the same area share the same information, they have identical topological databases.

The term domain sometimes is used to describe a portion of the network in which all routers have identical topological databases. Domain is frequently used interchangeably in AS. An area’s topology is invisible to entities outside the area. By keeping area topologies separate, OSPF passes less routing traffic than it would if the AS were not partitioned.

Area partitioning creates two different types of OSPF routing, depending on whether the source and destination are in the same or different areas. Intra-area routing occurs when the source and destination are in the same area; inter-area routing occurs when they are in different areas.

An OSPF backbone is responsible for distributing routing information between areas. It consists of all area border routers, networks not wholly contained in any area, and their attached routers.
In Figure 1.3, Routers 4, 5, 6, 10, 11, and 12 make up the backbone. If Host H1 in Area 3 wants to send a packet to Host H2 in area 2, the packet is sent to Router 13, which forwards the packet to Router 12, which sends the packet to Router 11. Router 11 then forwards the packet along the backbone to area border Router 10, which sends the packet through two intra-area routers (Router 9 and Router 7) to be forwarded to Host H2.

The backbone itself is an OSPF area, so all backbone routers use the same procedures and algorithms to maintain routing information within the backbone that any area router would. The backbone topology is invisible to all intra-area routers, as are individual area topologies to the backbone.

Figure 1.3 Hierarchical Routing Structures
Areas can be defined in such a way that the backbone is not contiguous. In this case, backbone connectivity must be restored through virtual links. Virtual links are configured between any backbone routers that share a link to a non-backbone area and function as if they were direct links.

1.5 PROBLEM STATEMENT

The problem of multiplexing different type of traffic in the network can be divided into two categories, namely call admission control and congestion management. Call admission addresses the issue of accepting traffic into the network in such a way as to prevent the network from reaching a state of congestion. Congestion management, on the other hand, includes pro-active traffic enforcement and reactive congestion control. The function of traffic enforcement is to police the bit rates of the users such that they do not violate the contract made at the call setup time. Reactive congestion control deals with the issues of how to react when the network is already congested.

(a) Goal for Data Traffic

- Data should not affect the other sources (voice and video)
- High throughput
- Near zero cell loss
- Fairness
- Adaptive algorithm allowing overuse

Data traffic should not affect the quality of voice and video traffic especially the delay jitter requirement. For data users, the throughput should be maximized. For voice and video, the issues of near zero cell loss and fairness are also applicable to data. The data sources should be allowed to exceed their nominal bit rates to make effective use of the excess bandwidth
available, in order to maximize link utilization. The allocation of the excess capacity should be in the ratio of the nominal (supposed) capacities, and furthermore, cell discarding should be biased towards users that exceed their nominal bit rate.

As part of the congestion management, a proactive, Adaptive queue rate-based source control scheme is proposed to allow non real-time data sources to increase their bit rates to effectively utilize most of the bandwidth available. Finally controlling the dynamic bandwidth allocation to transmission of data to overflow the queue management introduced a dynamic routing Multi-Source Virtual Routing Algorithm to transient the data from higher priority queue through the alternate path which is selected from route cache through route cache policy mechanism, which are already stored all possible route paths from source to destination or congested place to destination in route cache. It is crucial that the behavior of the data sources does not adversely affect the quality of service requirements for real and non real data.

(b) **Overall Goals for the Solution:**

- High network utilization
- Low implementation complexity
- Standards Compatible
- Work well over satellite links
- Algorithms and parameter values independent of network size and link types.

ATM Private Network to Network Interface routing relies on a dynamic link state based routing protocol that computes routes at the source node based on information contained in its topology database. A default model for computing the route is defined in the ATM Forum PNNI specification (ATM Forum 1996). This involves on-demand routing at every
node wherein the following operations occur every time a routing request is made.

1. a graph of nodes and links, representing the network topology, is constructed and populated with information from the nodes database.

2. a Generic Call Admission Control is carried out on this graph based on the call connection requirements. This process, also called pruning, eliminates all those links in the graph that cannot support the call requirements.

3. Dijkstra’s Single Shortest Path algorithm is run on the pruned graph so as to find a path from the source node to the destination node. This path minimizes a cost metric that is specified in the routing policy for the node and also satisfies the QoS requirements for the call.

4. a Designated Transit List (DTL) is created for the route and transported in the setup message that then goes out to the next hop.

The ABR flow control mechanism provides considerable flexibility to designers in deciding ABR switch operation and resource allocation for a given end system behavior. Different switches can use different mechanisms and still interoperate in the same network. But using different switch mechanisms leads to significant differences in terms of (performance queue length, cell loss, fairness of bandwidth allocation, throughput and link utilization, complexity and finally cost. In order to design an ABR switch mechanism one needs to understand the different flavors or options available and their impact on performance. Thus, for a given set of requirements and evaluation criteria, the switch engineering task is made easier. Classify the
ABR switches broadly into three categories based on the type of feedback, the congestion criteria and the queuing service used. This research work proposed a new rate based feedback mechanism to provide a quality of service.

The scope of the present work includes the study of following topics:

(a) To provide Virtual Source\Virtual Destination provision in the ABR traffic management framework can potentially improve performance of bursty traffic through the Segment-to-Segment feedback control mechanism.

(b) To reduce the buffer requirement in ATM Switches through the Doubly Finite Queue (DFQ) Management.

(c) To regulate the input rate to avoid the data loss through the Adaptive Queue Rate control Mechanism (AQRCM).

(d) To prevent the congested networks, setting the static multi-threshold in Doubly Finite Queue to monitor the network traffics and rerouting the transmission data.

(e) To study the new dynamic routing algorithm called as Multi-Source Virtual Dynamic Routing (MSVDR) to find out the shortest path between congested places to destination node in hierarchical networks.

(f) To study the performance of the DFQ using MSVDR in different metrics like cell delay variance, queue delay, queue size, link utilization, cell delay, route recovery time in various traffic loads.
In this thesis work, hybrid terrestrial and satellite ATM networks are considered: discarding cells in the case of congestion, namely buffer overflows inside the ATM switches, is not desirable, as the users’ quality of service requirements are not met. Therefore, a cell loss ratio close to zero is desired. Furthermore, the performance of the network should be optimized from both the users as well as the networks point of view: for users, this refers to maximizing the throughput, and satisfying the delay and jitter requirements of a specific traffic type, as discussed above, whereas for the network, the link utilization should be maximized. In order to achieve these performance goals, effective congestion management mechanisms are needed.

The goal of this thesis is to propose a new congestion management mechanism evaluated via simulation the effect of these congestion management mechanisms on network throughput and utilization, and users quality of service. An access policing and control mechanism together with doubly finite queue management is proposed. The goal of the access policing is to protect the network from badly behaving sources, whereas the goal of the queue management is to promote fairness among the users by biasing cell discarding towards the users that violate their average transmission rate contract with the network.

In summary, the network utilization should be maximized as that relates to the profit (Maximum utilization) of the network. A low implementation complexity is desired, as that reduces both the hardware and software costs. In addition, the solution should be standards compatible and take advantage of the network structure and the ATM cell structure. As discussed before, the solution should work well in a satellite environment, in which the propagation delays are significant. Finally the algorithms and parameter values should be designed so that they are not dependent on the network size and the link types.
1.6 PROPOSED CONTENTS OF THE THESIS

This thesis has been arranged in eight chapters. A brief description of each chapter is given below.

Chapter 1 provides an introduction to the Asynchronous Transfer Mode (ATM) Networks, Available Bit Rate (ABR) Service, and Traffic Management versus Congestion Control, Traffic Management for the ABR Service, basic information of Queuing Algorithm, Dynamic routing Policy, Problem Statement, Proposed contents of the thesis.

Chapter 2 provide a brief review of literature about the Virtual Dynamic Routing Algorithms, Queue Management, VS/VD Switching Design, Resource Management Cell, Threshold Techniques that are presented and discussed, and also provide the scope of the research work.

Chapter 3 discusses the ABR Traffic Management Frameworks, ABR Parameters, In-Rate and Out-of-Rate RM Cells, Forward and Backward RM cells, RM Cell Format, Source End System Rules, Destination End System Rules, and Switching Behaviors of the proposed systems that are represented and also discusses the Virtual Source/Virtual Destination Switching Design Consideration, A Non-VS/VD Switch, A VS/VD Switch, A VS/VD Switch with Unidirectional Data Flow, Bi-directional Data Flow, VS/VD Switch Design Options, Measuring the VC’s Current Rate, Measuring the Input Rate at the Switch, Effect of Link Congestion Actions on Neighboring Links, Frequency of Updating the Allocated Rate, VC Rate Measurement Techniques, Input Rate measurement techniques, Combinations of VC rate and input rate measurement options, Effect of Link Congestion Control Actions, Link Congestion and Allocated Rate Update Frequency Variable Options, Performance Evaluation of VS/VD Design Options, Metrics that are followed in this work. Finally it discusses the Route
Caching Policy, how Pre-computed route cache be filled, Cache Capacity, Cost of route cache pre-computation, Route selection from the cache, Appropriate responses to a cache miss, Cache entry invalidated, Cache replacement policy, Route Cache Update Heuristics, Crank back Indicator, Timer based Method, PTSE Count Method, Combined Heuristics and Limitations of the model.

Chapter 4 discusses and implements the PNNI Protocol Design, Multi-Source Virtual Dynamic Routing Algorithm, Doubly Finite Queue Management, Quality of Service for ABR, Traffic Consideration, Doubly finite Queuing Analysis, Source Behavior, Switching Behavior and Destination Behavior.

Chapter 5 presents the details of Simulation Configuration, Network Model, Call Block Ratio Vs Traffic Load, Time delay Vs Message Size, Link Utilization Vs Time Slot, Queuing delay Vs Time Slot, Link Utilization Vs ATM Node, Throughput % with and without MSVDR, Mean delay variance, Route Recovery Time Vs Failure level, Queuing Delay Vs Queue Management, Link Utilization Vs Queue Management, Queue delay time Vs Queue Length and these are compared with other reported algorithms.

Chapter 6 The conclusions from the different simulations and theoretical studies are summarized in this chapter. The scope for further work is also outlined.