University of Bern

Computer Networks and Distributed Systems

Prof. T. Braun

Project:

Differentiated Services in ATM-Networks

Alexander Dobreff, Kramgasse 10, PF, 3000 Bern 8

Differentiated Services in ATM-Networks

Content:

0. Introduction

In today’s internet there is just one ‘Best-Effort Service’ which sends the packets as fast and as reliable as possible. With the growing demand for new type of applications (like audio/video transmissions) the demand for bandwidth has grown also. The need for transmitting services that can’t adjust their bandwidth when congestion occurs will hinder these applications to go over the internet but remain on public net sides. To supply the measures for these applications the most common attempt is to enlarge the bandwidth, so that congestion can’t occur. This attempt while avoiding congestion is not always possible on an world wide net like the internet.

A different solution is to guarantee the requirements for the link over the network, so that bandwidth and other parameters are reserved for certain links which brings the benefit of guaranteed transmission, while other links still running on best-effort suffering from congestions.
The first attempts took place in small areas, but these attempts are not scaleable and therefore not capable supporting large networks.

The focus in a new working group (differentiated working group, diffserv) is on scaleable proposals that support a service differentiation. These proposals are discussed and weighted in this paper as well as the joint between ATM networks and these differentiated services. Therefore ATM is formerly described in detail.

In another section measures are described that support this service differentiation while simultaneously making the transport of these packets and flows faster.
This is done in different ways:

- Only the first packet is routed then the decision for the following packets is already made, relieving the routers (IP Switching and Tag Switching).

- To make the routes as short as possible, a protocol was invented named NHRP (Next Hop Resolution Protocol) which is already working in small areas but is not yet approved in large area networks.

- To handle the various links these are - whenever possible - clustered into one single link. This link aggregation is already done in ATM networks unburden the routers to switch every single link.

1. ATM

1.1 General Description

Asynchronous Transfer Mode (ATM) is a connection oriented cell switching network system and is often used as core (or backbone) network to support the high throughput of modern network needs. ATM has cells with a length of 53 Octets (5 Header and 48 Data Octets). Arriving packets are cached, chopped into cell size and transmitted over fast links.

Because of the reliability of the transmission (enforced through AAL, ATM Adaptation Layer) no error checking is made on link level.
Two kinds of connections can be established: Virtual Channels (VC) and Virtual Paths (VP) [1]. To avoid finding the route for every cell this is done once, while establishing a Virtual Cannel - this channel gets a unique number - to the outgoing ATM switch. The following cells are switched with the information in the cell header only making the route decision fast.

Virtual Channels having the same destination (for one, many or all hops) are aggregated to Virtual Paths - which gets also a unique number - so switching a route to the next switch is done only by a former labelled identifier in the header of the cell. That allows fast switching of cells.
A VP can have VCs with different traffic contracts that are known as Quality of Service (QoS) classes.

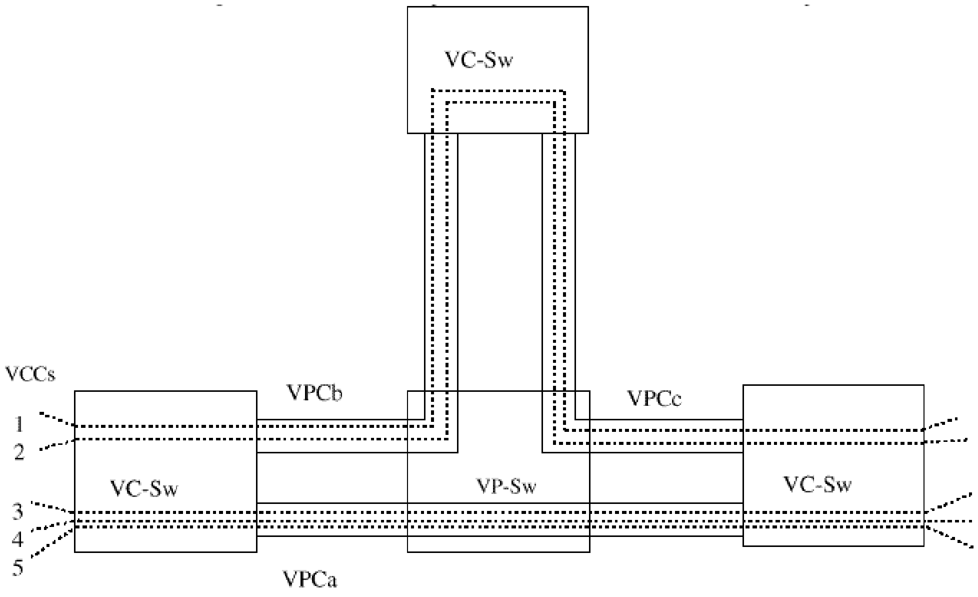


Figure 1: VCCs and VPCs

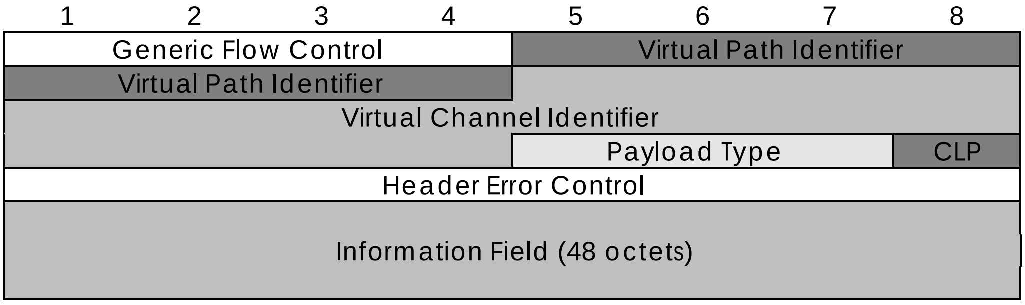
ATM has five service categories; two (CBR, rt-VBR) categories for real-time services and three for non real-time services (nrt-VBR, UBR, ABR).
Priority Control, traffic marking, policing and shaping, as well as mechanisms to avoid or recover from congestion situations in done within ATM networks. Some aspects are highlighted more closely:

1.1.1 ATM cell format

ATM cells [2] have a header with 5 octets long and a payload section which is 48 octets long. The header consists of the following sections:

- Generic Flow Control (GFC) is not yet fully used but is intended to control cell - flow.
- Virtual Path Identifier (VPI) constitutes a routing file for the network. It supports the Virtual Paths in the networks.
- Virtual Channel Identifier (VCI) is used for routing to and from the end user. Thus, it works much as a service access point.
- The Payload Type (PT) field indicates the type of information in the information field. It can distinguish between user information or not and whether there is a congestion in the net.
- Cell Loss Priority (CLP) is used to provide a guidance to the network in the event of congestion. A value of 0 indicates a cell of relatively high priority, which should not be discarded unless no other alternatives is available. A value of 1 indicates that this cell is subject to discard within the network.
- Header Error Control (HEC) is used to detect errors in the header therefore a checksum with polynomial division is calculated.

Information Field, is where the payload is situated.



<table>
<tr><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th><th>6</th><th>7</th><th>8</th></tr>
<tr><td colspan="4">Generic Flow Control</td><td colspan="4">Virtual Path Identifier</td></tr>
<tr><td colspan="4">Virtual Path Identifier</td><td colspan="4"></td></tr>
<tr><td colspan="8">Virtual Channel Identifier</td></tr>
<tr><td colspan="4"></td><td colspan="3">Payload Type</td><td>CLP</td></tr>
<tr><td colspan="8">Header Error Control</td></tr>
<tr><td colspan="8">Information Field (48 octets)</td></tr>
</table>

Figure 2: ATM cell format

1.1.2 Aggregating links

A Virtual Path Connection (VPC) provides a convenient instrument to group similar Virtual Channel Connections (VCC) together. The network provides aggregate capacity and performance characteristics on the Virtual Path, and these are shared by the Virtual Connection. There are three cases to consider:

- User-to-user applications: The network has no knowledge of the QoS of the individual VCCs within VPC. It is the responsibility of the user to assure that the aggregate demand from the VCCs can be accommodated by the VPC.

- User-to-network application: The network is aware of the QoS of the VCCs within the VPC and has to accommodate them.

- Network-to-network application: Again, the network is aware of the QoS of the VCCs within the VPC and has to accommodate them.

There are a number of alternatives for the way in which VCCs are grouped and the type of performance they experience.
If all of the VCCs within a VPC are handled similarly, then they should experience similar expected network performance, in terms of cell loss ratio, cell transfer delay, and cell delay variation. Alternatively, when different VCCs within the same VPC require different QoS, the VPC performance objective agreed by the network should be set suitably for the most demanding VCC.
In either case, with multiple VCCs within the same VPC , the network has two general options for allocating capacity to the VPC:

Aggregate peak demand

The network may set the rates on the VPC equal to the total of the peak data of all of the VCCs within the VPC. The advantage of this approach is that each VCC can be given a QoS that accommodates its peak demand. The disadvantage is that most of the time the VPC capacity will not be fully used, and therefore the network will have under-utilised resource.

Statistical multiplexing

If the network sets the capacity of the VPC to be greater or equal to the average data rates of all the VCCs but less than the aggregated peak demand, then a statistical multiplexing service is supplied. With statistical multiplexing, VCCs experience greater Cell Delay Variation (CDV) and greater Cell Transfer Delay (CTD). Depending on the size of the buffers used to queue cells for transmission, VCCs may also experience greater cell loss ratio. This approach has the advantage of more efficient utilisation of capacity and is attractive if the VCCs can tolerate the lower QoS.

When statistical multiplexing is used, it is preferable to group VCCs into VPCs on basis of similar traffic characteristics and similar QoS requirements. If dissimilar VCCs share the same VPC and statistical multiplexing is used, it is difficult to provide fair access to both high-demand and low-demand traffic streams.

1.1.3 Connection Admission Control (CAC)

Connection Admission Control (CAC) is the first line of defence for the network in protecting itself from excessive loads. In essence, when a user requests a new VPC or VCC, the user must specify the traffic characteristics in both directions from that connection. The user selects traffic characteristics by selecting a QoS from among the QoS classes that the network provides. The network accepts the connection only if it can commit the resources necessary to support the traffic level while at the same time maintaining the agreed QoS of existing connections. Once the connection is accepted, the network continues to provide the agreed-on QoS as long the user complies with the traffic contract.

The performance parameters are set to the QoS classes (see next section) accordingly.

Once a connection has been accepted by the Connection Admission Control (CAC) function, the **Usage Parameter Control (UPC)** [1] function of the network monitors the connection to determine whether the traffic conforms to the traffic contract.

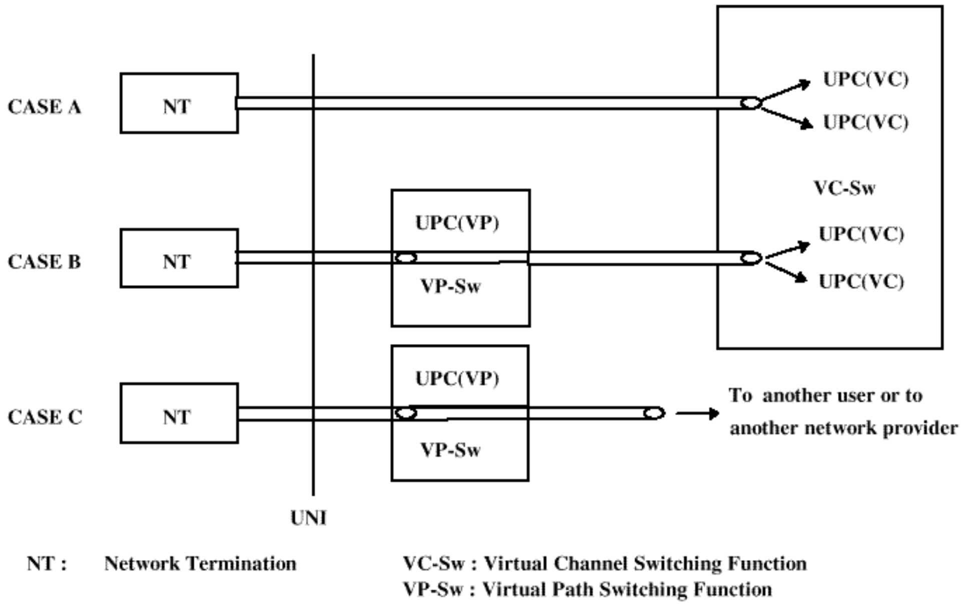

CASE A
NT
UPC(VC)
UPC(VC)
VC-Sw
CASE B
NT
UPC(VP)
VP-Sw
UPC(VC)
UPC(VC)
CASE C
NT
UPC(VP)
VP-Sw
To another user or to another network provider
UNI
NT : Network Termination
VC-Sw : Virtual Channel Switching Function
VP-Sw : Virtual Path Switching Function

Figure 3: Usage Parameter Control (UPC)

The main purpose of UPC is to protect network resources from overload on the connection that would adversely affect the QoS on the other connections by detecting violations of assigned parameters and taking appropriate actions. Usage Parameter Control can be done at both the Virtual Path and Virtual Channel level. This task is accomplished with the Peak Cell Rate Algorithm and the Sustainable Cell Rate Algorithm that both use the Generic Cell Rate Algorithm (GCRA) to perform it's task.

1.1.4 Priority Control

Priority Control comes into play when the network, at some point beyond the UPC function, discards CLP=1 cells. The objective is to discard low-priority cells in order to protect the performance for high-priority cells. Note that the network has no way to discriminate between cells that were labelled as low-priority by the source and cells that were degraded to low-priority by the UPC.

1.1.5 Traffic policing, traffic shaping

Traffic policing occurs when a flow of data is regulated so that cells (or packets) that exceed a certain performance level are discarded or given a lower priority. Traffic shaping is used to smooth out a traffic flow and reduce cell clumping. This can result in a fairer allocation of resources and reduced average delay time. It may be desirable to supplement a traffic policing policy with a traffic shaping policy.
When traffic policing and shaping is done at the border to ATM networks, less unexpected congestion is occurring and less traffic policing and shaping within the ATM network is needed.
A simple approach to traffic shaping is to use a form of the leaky bucket algorithm known as the token bucket. In contrast to the GCRA leaky bucket, which simply monitors the traffic and rejects noncompliant cells, a traffic shaping leaky bucket controls the flow of compliant cells.

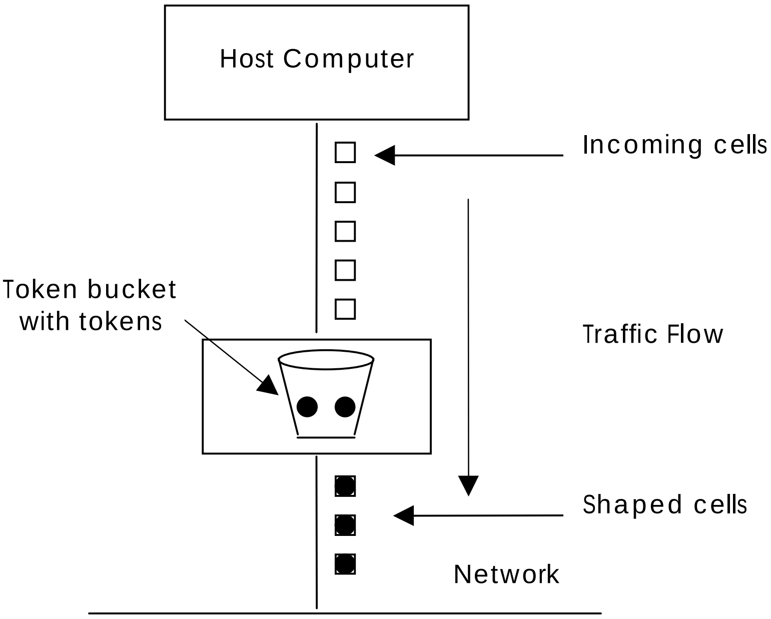


Figure 4: Token bucket

Figure 4 illustrates the basic principle of the token bucket. A token generator produces tokens at a rate of p tokens per second and places these in the token bucket which has a maximum capacity of b tokens. Cells arriving from the source are placed in a buffer with a maximum capacity of K cells. To transmit a cell, one token must be removed from the bucket. If the token bucket is empty, the cell is queued waiting for the next token. The result of this scheme is that if there is a backlog of cells and an empty bucket, then cells are emitted at a smooth flow of p cells per second with no cell delay variation until the backlog is cleared. Thus, the token bucket smoothes out bursts of cells.

1.1.6 Congestion Control

ATM congestion control refers to the set of actions taken by the network to minimise the intensity, spread, and duration of congestion. These actions are triggered by congestion in one or more network elements.
Two functions are defined:

Selective Cell Discarding

Selective cell discarding is similar to priority control. In the priority control function CLP=1 cells are discarded to avoid congestion. However, excessive cells are discarded only. Through that measure the performance objectives for the CLP=0 and CLP=1 flows are still met. Once congestion occurs, the network is no longer bound to meet all performance objectives. To recover from congestion conditions, the network is free to discard any CLP=1 cell and may even discard CLP=0 cells on ATM connections that are not complying to their traffic contract.

Explicit Forwarding Congestion Indication

Any ATM network node that is experiencing congestion may set an Explicit Forward Congestion Indication in the cell header. It indicates that this cell on this ATM connection has encountered congested resource. The application may then invoke actions in higher-layer protocols to adaptively lower the cell rate of the connection. Once the value in the cell header is set by an node, it may not be altered by other network nodes along the path to the destination.

1.2 ATM Service Categories

ATM has five Service Categories to support Quality of Service. CBR and rt-VBR are designed to support real-time services, nrt-VBR, UBR and ABR are suited for non-real-time services. A short description of each is given here. In section 1.5 ATM Service Categories Attributes the relation between the ATM Service Categories and the Traffic and QoS parameters is explained. For ATM traffic parameters see next section.

1.2.1 Constant Bit Rate (CBR)

CBR is designed to transport real-time data on a constant or maximal bitrate. The Peak Cell Rate (PCR) is defined as a constant and through this a maximum bitrate is accomplished. Late cells are considered less important and can be discarded at any time. Traffic marking, policing and shaping is important for real-time traffic support.

1.2.2 real-time Variable Bit Rate (rt-VBR)

rt-VBR is made for bursty real-time links with a tight time delay. The Maximal Burst Size (MBS) is specified in addition to all other CBR parameters. The specification of that additional parameter limits the maximum burst that is allowed. Therefore queue reservation is made to hold the bursts.

1.2.3 non-real-time Variable Bit Rate (nrt-VBR)

nrt-VBR is good for bursty non-real-time traffic whose cells are delivered with priority so that links with a critical response time can go over this service. The same traffic parameters have to be set as for rt-VBR.
Cell Loss Ratio (CLR) is guaranteed if the sender does not exceed the agreed parameters. CLR is the only QoS parameter that is specified, all other QoS parameters which would be needed to specify real-time traffic are not specified.

1.2.4 Unspecified Bit Rate (UBR)

UBR is made for traditional computer communication applications like ftp. No commitment on Cell Loss Ratio (CLR) and Cell Transfer Delay (CTD) is made; the sharing is not necessarily fair and there is no specific traffic contract, there is not even a commitment on transmitting data at all.
It is the traffic with the least Quality of Service support and can be compared to the traditional best-effort traffic. The parameters Peak Cell Rate (PCR) and Cell Delay Variation Tolerance (CDVT) are specified but no QoS agreement over QoS parameters is made.

1.2.5 Available Bit Rate (ABR)

ABR transports the same traffic like UBR but with a lower probability of congestions through flow control. Flow control is a mechanism in which the link can adjust the bitrate in accordance to the bitrate available on the network. Low Cell Loss Ratio (CLR) can be expected for stations which stay within the traffic contract and have a flow control performed through feedback from the receiver. The available bandwidth can vary from Minimum Cell Rate (MCR) - greater or equal zero - to Peak Cell Rate (PCR). The traffic contract is negotiated on both directions and the network commits fair resource share.
With these preconditions the link can expect a regular service even in congested ATM networks.

1.3 ATM Traffic Parameters

A traffic parameter describes an inherent characteristic of a traffic source. It may be quantitative or qualitative. Traffic parameters described here include Peak Cell Rate (PCR), Cell Delay Variation Tolerance (CDVT), Sustainable Cell Rate (SCR), Maximum Burst Size (MBS) and Minimum Cell Rate (MCR).

1.3.1 Peak Cell Rate (PCR)

The Peak Cell Rate (PCR) traffic parameter specifies an upper bound on the rate at which traffic can be submitted on an ATM connection. Enforcement of this bound by the Usage Parameter Control (UPC) allows the network to allocate sufficient resources to ensure that the network performance objectives (e.g., for Cell Loss Ratio) can be achieved. PCR is specified as cells per second.

1.3.2 Cell Delay Variation Tolerance (CDVT)

ATM layer functions (e.g., cell multiplexing) may alter the traffic characteristics of connections by introducing Cell Delay Variation. When cells from two or more connections are multiplexed, cells of a given connection may be delayed while cells of another connection are being inserted at the output of the multiplexer. The upper bound on this measure is the Cell Delay Variation Tolerance (CDVT).

1.3.3 Sustainable Cell Rate (SCR)

The Sustainable Cell Rate (SCR) is an upper bound on the average rate of the conforming cells of an ATM connection. Enforcement of this bound by the Usage Parameter Control (UPC) could allow the network to allocate sufficient resources, and ensure that the performance objectives (e.g., for Cell Loss Ratio) can still be achieved. SCR is specified as cells per second.

1.3.4 Maximum Burst Size (MBS)

The Maximum Burst Size (MBS) parameter specifies the burst size that is allowed in services that are explicitly supporting bursts (rt-VBR and nrt-VBR). This parameter is important to allocate the buffers size and it's also needed for the Generic Cell Rate Algorithm (GCRA) that decides whether the cells are conformant, marked out-of profile or dropped.

1.3.5 Minimum Cell Rate (MCR)

The Minimum Cell Rate (MCR) is the rate at which the source is always allowed to send at minimum. It is specified in cells per second.

1.3.6 Burst Tolerance (BT)

The Burst Tolerance (BT) can be derived from Peak Cell Rate (PCR), Sustainable Cell Rate (SCR), and Maximum Burst Size (MBS) with the Generic Cell Rate Algorithm (GCRA). In addition, the GCRA is used to test the conformance of the declared values of the traffic parameters PCR and SCR and MBS.

1.4 ATM QoS Parameters

The ATM Service Categories (CBR, rt-VBR, nrt-VBR, UBR, ABR) are specified through QoS parameters which are set differently from service to service to achieve the desired service quality.
There are two kinds of QoS parameters: Negotiated QoS parameters and not negotiated QoS parameters. The not negotiated QoS parameters are measured in the network and are compared to the traffic agreement, which is specified through the traffic parameters and the QoS parameters [1].

	QoS parameter:	Has influence on:
Negotiated	Peak-to-peak Cell Delay Variation (ptpCDV)	Delay
	Maximum Cell Transfer Delay (maxCTD)	Delay
	Cell Loss Ratio (CLR)	Dependability
Not negotiated	Cell Error Ratio (CER)	Accuracy
	Severely Errored Cell Block Ratio (SECBR)	Accuracy
	Cell Misinsertion Rate (CMR)	Accuracy

Table 1: QoS Parameters

Following a description of these parameters and how they are used with the Service Categories:

Delay Parameters:

1.4.1 peak-to-peak Cell Delay Variation CDV (ptpCDV)

Peak-to-peak Cell Delay Variation CDV (ptpCDV) measures the cell delay of each cell whether it is within a certain range or not. This range goes from a lower peak - a fixed delay for physical transmission - to a higher peak - which is part of the traffic agreement.
The lower peak is the time that a cell needs to travel over cable and wires, including the time within integrated circuits, that means the time that is fix to a cell to travel from point A to point B. The higher peak is agreed and is a probability whether cells arrive within a certain time or whether they are regarded as late or lost. The cells that arrive late cannot expect further priority treatment and can be dropped or given a lower priority.

Two measuring methods are possible:

One-point CDV:

The one-point CDV describes the variability in the pattern of cell arrival events observed at a single measurement point with reference to the negotiated peak rate. The one-point CDV for a cell at a measurement point is defined as the difference between the cell's reference arrival time and actual arrival time at the measurement point.
Positive values of the one-point CDV correspond to cell clumping; negative values of the one-point CDV correspond to gaps in the cell stream.

Two-point CDV:

The two-point CDV describes the variability in the pattern of cell arrival events observed at the output of a measurement point (MP 2) with reference to the pattern of the corresponding events observed at the input to the measurement point (MP 1). The two-point CDV for a cell between two measurement points (MP 1 and MP 2) is the difference between the absolute cell transfer delay of the cell between the two MPs and a defined reference cell transfer delay between MP 1 and MP 2.

1.4.2 maximum Cell Transfer Delay (maxCTD)

A fixed delay is given through physical parameters and switching times over components. To this the peak-to-peak Cell Delay Variation (ptpCDV) which has a probability of 1- α that the cells arrive in time (that makes a α probability for late or lost cells) is added to receive the maximum Cell Transfer Delay (maxCTD).

$$\text{maxCTD} = \text{Fixed Delay} + \text{ptpCDV}$$

A policing mechanism watches over the arriving cells whether they are within the 1 - α probability of correctly arriving cells. They are kept within a profile or they are tagged with a lower priority. Scheduling is performed with the cells which are nonconforming at every ATM-switch to assure a steady cell stream. Early cells are queued for later transmission. Nevertheless nonconforming cells can be queued or dropped.

CTD is measured at two points in different ATM switches, at the exit point of the first ATM switch and at the entry point of the next ATM switch and then compared with the agreed maxCTD parameter:

CTD = Cell entry event at Point 2 - Cell exit event at Point 1

ptpCDV and maxCTD are dependent from each other over a probability density function of CTD which specifies the probability α how much percent of cells are considered late or lost [1].

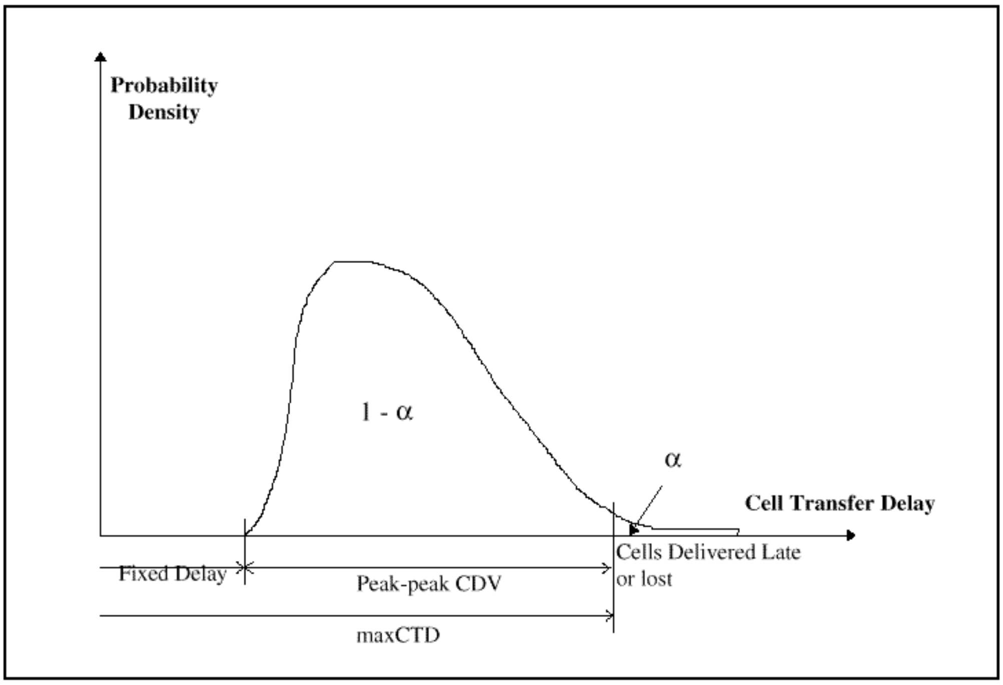


Figure 5: Dependability from ptpCDV and maxCTD

Dependability Parameters:

1.4.3 Cell Loss Ratio (CLR)

Cell Loss Ratio (CLR) is the ratio of the lost cells. It is specified and controlled over a certain amount of cells (cell blocks).

$$\text{CLR} = \frac{\text{Lost Cells}}{\text{Total Transmitted Cells}}$$

1.4.4 Cell Error Ratio (CER)

Cell Error Ratio (CER) is the ratio of cells that arrive with errors.

$$\text{CER} = \frac{\text{Errored Cells}}{\text{Total Transmitted Cells}}$$

1.4.5 Severely Errored Cell Block Ratio (SECBR)

Severely Errored Cell Block Ratio (SECBR) measures the cell blocks that are errored, lost or misinserted (reach the wrong destination) within a certain amount of cell blocks.

$$\text{SECBR} = \frac{\text{Severely Errored Cell Blocks}}{\text{Total Transmitted Cell Blocks}}$$

A cell block is a sequence of cells transmitted consecutively on a given connection. A severely errored cell block occurs when more than M errored cells, lost cells, or misinserted cell are observed in a received cell block.

1.4.6 Cell Misinsertion Rate (CMR)

Cell Misinsertion Rate (CMR) counts the misinserted cells (reach the wrong destination) over a certain period of time.

$$\text{CMR} = \frac{\text{MisinsertedCells}}{\text{Time Interval}}$$

Cell misinsertion on a particular connection is most often caused by an undetected error in the header of a cell being transmitted on a different connection.

1.4.7 Generic Cell Rate Algorithm (GCRA)

GCRA is a continuous state leaky bucket algorithm that tests every incoming cell whether it fits to the traffic contract or not. It is therefore important to do that in accordance with the above named and explained parameters.

1.5 ATM Service Categories Attributes

Finally there is a table which shows the parameters which have to be set for an specific service category [1]:

	CBR	rt-VBR	nrt-VBR	UBR	ABR
Traffic Parameters:					
PCR and CDVT	specified			specified	specified
SCR, MBS, CDVT	n/a	specified		n/a	
MCR	n/a			n/a	specified
QoS Parameters:					
ptpCDV	specified		unspecified		
maxCTD	specified		unspecified		
CLR	specified			unspec.	see Note
Other Attributes:					
Feedback	unspecified				specified

Table 2: ATM Service Categories Attributes

Note: CLR is low for sources that adjust cell flow in response to control information.

PCR	Peak Cell Rate
CDVT	Cell Delay Variation Tolerance
SCR	Sustainable Cell Rate
MBS	Maximum Burst Size
MCR	Minimum Call Rate
ptpCDV	peak-to-peak Cell Delay Variation
maxCTD	maximum Cell Transfer Delay
CLR	Cell Loss Ratio

CBR	Constant Bit Rate
rt-VBR	real-time Variable Bit Rate
nrt-VBR	non-real-time Variable Bit Rate
UBR	Unspecified Bit Rate
ABR	Available Bit Rate

2. Integration of IP and ATM

Besides the standardised specifications to run IP over ATM, like LANE [31,32,33,34] and Classical IP [35], new approaches and improvements are discussed in the following sections.

2.1 Ipsilon's IP Switching

IP (Internet Protocol) switching is a new kind of IP routing developed by Ipsilon Networks. Unlike conventional routers, IP switching routers use ATM hardware to speed packets through networks, and it appears to be considerably faster than older router techniques.

2.1.1 Introduction

A key difference between the router approach and the IP switching architecture is that IP switching allows most data between ATM ports to traverse the switch without being handled at all by a forwarding engine, whereas a router approach always requires the use of at least one forwarding engine.

In IP switching the fast ATM hardware is used, preserving the connectionless nature of IP while abandoning the end-to-end ATM connection. This has the particular advantage of not requiring end-to-end signalling, or address resolution, and requiring only the standard IP routing protocols.
Soft-state in the ATM hardware is used to cache the IP forwarding decision. This enables traffic on the same IP flow to be switched by the ATM hardware rather than forwarded by the IP software. This is an approach of IP, with the speed, capacity, and multiservice traffic capabilities of ATM. The datagram forwarding in IP requires no state to be maintained for individual connections. This has proven extremely robust in the presence of failures.

ATM has received much attention because of its high capacity, its bandwidth scalability, and its ability to support multiservice traffic. However, ATM is connection-oriented whereas the vast majority of modern data networking protocols are connectionless. This mismatch has led to complexity, inefficiency, and duplication of functionality in attempting to apply ATM technology to data communication.
IP has seen very rapid growth in the last several years also. Research suggests that IP is no less capable of supporting real-time and multimedia applications than ATM.

However, the greater complexity of IP forwarding is likely to ensure that routing remains more expensive than switching for some time.

2.1.2 Connectionless Connections

Before we can go to the core functionality of IP switching the concept of a flow has to be described.

A flow is a sequence of packets that is treated identically between a source machine (or application process) and one or more destination machines (or application processes). All the packets follow the same route through the network - besides the packets are re-routed - and receive identical service policies. The forwarding and handling of each flow is determined by the first packets in the flow. Once the flow is classified, the routing and policy decisions may be cached for further packets. These may be processed accordingly to the cache entry, without requiring the full flow classification, hence accelerating the forwarding process. Network resources may be reserved on behalf of a flow, to offer Quality of Service guaranties.

An efficient mapping of IP onto ATM must consider the characteristics of the application and the transport protocol in deciding whether to establish an end-to-end ATM connection on behalf of an specific flow.

- Flows carrying real-time traffic, flows with Quality of Service requirements, or flows likely to have a long holding time, will be handled most efficiently by mapping them into an individual ATM connection.

- Short duration flows and database queries will be handled most preferably by connectionless packet forwarding between IP routers connected via shared, pre-established ATM connections. This is particularly true for simple datagram exchanges such as DNS lookups that consist of a single packet in each direction.

Establishing an end-to-end ATM connection for **every IP packet** flow would impose a heavy load on the ATM signalling protocol, and is resulting in unnecessary delay for query-response traffic.

One of the reasons that IP scales well to large networks is due to its connectionless nature. If a router or a link fails in a network, IP simply routes around the failure. If an end-to-end connection across an ATM cloud is established, the failure of a link or router will invalidate all associated connections. This will result in a heavy load on the signalling protocol to re-establish all of the ATM connection states. Also, many connections that are not directly associated with the failed component will become sub-optimal, when the topology changes. It is also possible that routing loops can form after a topology change, until the old routing information is cleared in the address resolution servers and route servers.

It is clear that in order to take advantage of the efficiency of switching, and to offer Quality of Service guaranties, the state information must be maintained within the switches. They are typically designed to be refreshed periodically, such that many possible error conditions may be corrected by simply timing out old state. The cost of this simplicity is that the messages required to maintain the soft-state might in some cases impose a significant load on the network controllers.

2.1.3 IP Switching

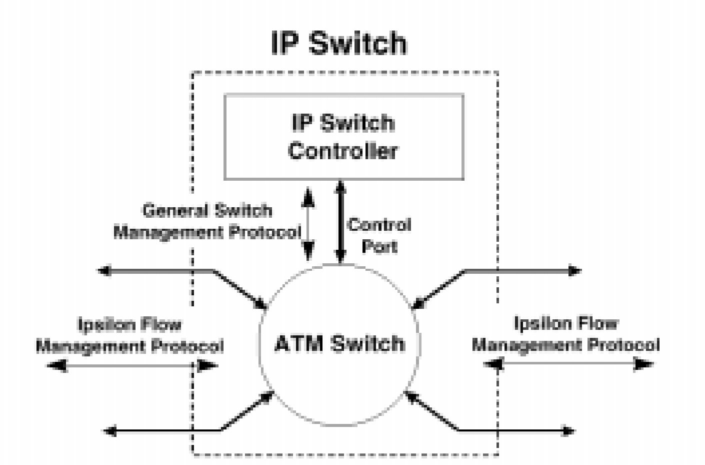

IP Switch
IP Switch Controller
General Switch Management Protocol
Control Port
Ipsilon Flow Management Protocol
ATM Switch
Ipsilon Flow Management Protocol

Figure 6: Structure of an IP Switch

To construct an IP switch (Figure 6) [3], the hardware of an ATM switch is taken, without any modification, but the software resident in the control processor above AAL-5 is completely removed. Thus the signalling, any existing routing protocol, and any LAN emulation server or address resolution servers, etc. is removed. In place of the ATM software a simple, low-level control protocol, called the **General Switch Management Protocol (GSMP)** is loaded, to give the IP switch controller access to the switch hardware. The IP switch controller is a high-end processor running standard IP router software with extensions that allow it to make use of the switching hardware. These extensions include the simple **Ipsilon Flow Management Protocol (IFMP)** to associate IP flows with ATM virtual circuits, a **Flow Classifier** to decide whether to switch each flow, while GSMP controls the switch hardware.

At system start-up, a default ATM Virtual Circuit, on a well-known Virtual Path Identifier / Virtual Channel Identifier (VPI/VCI), is established between the IP software running on the IP switch controller and that of each of its neighbors. The default channel is used for the hop-by-hop connectionless forwarding of IP datagrams. To connect IP switching networks across a public ATM network, a virtual path may be established across the public network with a configured VPI.

2.1.4 Flow Classification

An important function of the flow classification operation is to select those flows that are to be switched in the ATM switch and those that should be forwarded hop-by-hop through the forwarding engine.
Multimedia traffic like voice, image, video conferencing etc., are examples of long duration flows where there is a good probability of a high traffic volume. Many multimedia applications require multicast which makes it very suitable for switching across an ATM switch making use of ATM hardware multicast capability.
Short duration flows consisting of a small number of packets should be handled directly by the forwarding engine. Name server queries and brief client-server transactions are examples of traffic that are not worth the effort of establishing a switched connection.

Flow classification is a local decision. Two packets belong to the same flow if the values of the fields: type of service, protocol, source address, destination address, source port, and destination port are identical.

2.1.5 General Switch Management Protocol (GSMP)

General Switch Management Protocol (GSMP) is a simple master-slave, request response protocol. The master (switch controller) sends requests and the switch issues a positive or negative response when the operation is complete. Unreliable message transport is assumed between controller and switch for speed and simplicity. The link between switch and controller will either be very reliable, or broken, in which case the overhead of adding error detection and retransmission through a protocol is unnecessary. All GSMP messages are acknowledged and the implementation handles its own retransmission. GSMP runs on a single, well known Virtual Channel (VPI 0, VCI 15).

A configuration message is used by the controller to discover the capabilities of the ATM switch. The ATM switch can report the incoming Virtual Path Identifier (VPI) and Virtual Channel Identifier (VCI) ranges it can support, its interface type, the cell rate and the number of priority levels it supports in its output queue. Once the configuration of the switch has been discovered, the controller can begin issuing connection management messages. They enable the controller to establish and remove connections across the switch.

2.1.6 Ipsilon Flow Management Protocol (IFMP)

Ipsilon Flow Management Protocol (IFMP) runs independently across each link in a network of IP switches that connects IFMP peers that are: IP switches, directly attached to hosts, or IFMP capable edge routers. On ATM links it uses the default Virtual Channel (VPI 0, VCI 15).
All packets belonging to a flow that has not yet been switched is forwarded hop-by-hop between IP switch controllers using this default Virtual Channel. When a new flow arrives at an IP switch it is classified. One of the results of flow classification is the decision whether a flow should be switched or not.

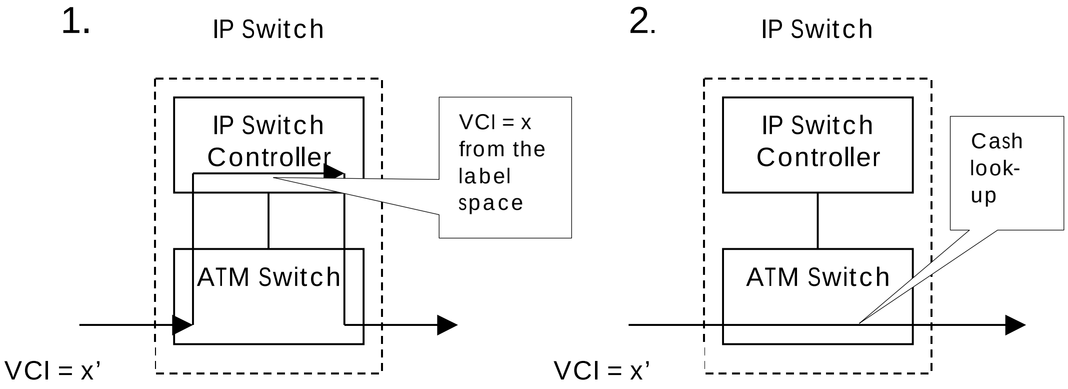


Figure 7: IP Switching

Before a flow can be switched it first must be labelled. If an IP switch controller decides that a flow should be switched, it selects a free label (e.g., VCI = x) from the label space (Figure 7. 1.). The switch driver is then instructed to map VCI = x on input port i to VCI = x' on the control port c. After making the entry in the translation table of the switch input port (via GSMP) the IP switch controller sends an IFMP redirection message upstream to the previous node. The redirection message contains the label (VCI = x), a flow identifier, and a lifetime field. The flow identifier contains the set of header fields that specifies the flow. The redirection message requests the upstream node to transmit all further packets with header fields that match the flow identifier, on the ATM virtual circuit specified by the label field (VCI = x). The lifetime field specifies the length of time for which this redirection is valid.

From this point, packets belonging to the flow will arrive at the switch controller, port c, with the ATM VPI/VCI label x'. The packets will still be reassembled and forwarded by the IP forwarding software, but the process is accelerated because the previous routing decision for this flow is cached and is indexed by the label x' (Figure 7. 2.).

The real benefit of switching is experienced when the downstream node redirects the flow to a specific VCI too. It can switch all further traffic belonging to that flow directly within the ATM hardware. The IP switch does this by instructing the switch to map label x on port i and to label y on port j. Thus, traffic on this flow is no longer processed by the IP switch controller in a store-and-forward manner, but is switched directly to the required output port.

Label	Port
: :	: :
x	i
y	j
: :	: :

Figure 8: Mapping of Labels and Ports

When traffic is swapped from hop-by-hop forwarding to switched forwarding, it is possible that packet misordering occurs. The first packet on the switched path may be delivered to the destination before the last packet on the store-and-forward path. To avoid this cause of packet loss the IP switch implementation establishes the switched path backwards from the destination through the network towards the source.

Since on ATM packets are transmitted as cells, it is also possible that the change to the switched path will occur in the middle of a packet. This will result in the loss of the whole packet. To avoid this packet loss the destination end system makes the switching decision before nodes within the network. Thus the switched path is established while traffic is still flowing over the store-and-forward path, and the source may cut the traffic over to the switched path on a packet boundary.

When an IP switch accepts a redirection message, it also changes the encapsulation it uses for packets belonging to the redirected flow. This approach is taken for security reasons. It allows an IP switch to act as a simple flow-based firewall without having to inspect the content of each packet. It prevents a user from establishing a switched flow to a permitted destination or service behind a firewall, and then submitting packets with a different header to gain access to a prohibited destination.

If one node decides to label a flow, its neighbours within the same domain will very likely make the same flow classification and switching decision. Therefore an IP switch controller can send an IFMP redirect message to see if the flow is yet labelled on the upstream link. When upstream and downstream links are both labels for a given flow, that flow is switched directly through the ATM switch.

2.1.7 Quality of Service

Quality of service may be considered as a mean of forwarding packets from different flows with controlled unfairness. Some packets receive faster service than others according to some established policy. Typically, this policy is established by a user requesting specific resources from the network using some form of signalling protocol. This is known as **contract-based quality of service**, as the user makes a contract with the network to reserve the specified resources.

An alternate approach is to use **policy-based quality of service** differentiation. In policy-based quality of service, the quality of service requirements for different flows are established by the policy within the administrative domain. This policy is established by the network manager.
Policy-based quality of service differentiation will not provide the same control granularity that is available from the contract-based approach, but it does offer the network manager significant capabilities to control the use of resources in the network. For example, interactive applications can be given higher priority than bulk transfer applications or bandwidth available to multimedia applications can be constrained to avoid overloading network links.

IP switching can support both policy-based and contract-based quality of service. Each IP switch can make a policy-based quality of service decision, according to the policy established within the administrative domain.

Each flow is classified as part of the forwarding operation and quality of service information may be included in the flow classification decision. This decision may be based upon any of the fields within the packet header, for example: the application, the type of service field, the source and destination IP addresses, etc. Each IP switch must interpret the quality of service policy according to the capabilities of the underlying ATM switch hardware.

Resources may be reserved by configuring the queuing and scheduling hardware within the ATM switch or software within the IP switch controller. Also, the flow may be policed by configuring the policing hardware in the ATM switch according to the flow specifications.

Contract-based quality of service requests for individual flows may also be supported using the Resource reSerVation Protocol (RSVP). RSVP allows an application to specify the required service and the traffic characteristics of a flow. A reservation request may be accepted or denied by each IP switch in the path using an admission control policy.

2.1.8 Other advantages of IP Switching

Robustness

It is important that the protocol is robust in the face of network or node failure. Each IP switch controller periodically examines every flow. If a flow has received traffic since the last refresh period, the controller sends another redirect message upstream to refresh the flow. If a flow has received no redirect messages for a period in excess of its lifetime, it is removed. The flow state is not deleted until an IFMP Reclaim Ack message is received to acknowledge the release of the requested label.
For flows that are labelled, but not switched, the IP switch controller can examine its own state to see whether the flow has received any traffic in the previous refresh period. For flows that are switched, the controller must query the switch hardware to discover whether a specific channel has recently been active.

If a link or a router fails, the normal process of connectionless routing will establish a new route. When routes or routing policy changes, any existing state related to a changed route will be invalidated and flushed. Affected traffic will once again be forwarded over the default channel and new virtual connections will be rebuilt from the point of failure across the new path. Thus the connections are repaired only within the locality of the failure. This is far more efficient than deleting all affected connections and newly establishing them, end-to-end, as is the case for connection-oriented networks.

One of the basic requirements of IP is that the Time To Live (TTL) field of the IP header in a packet be decremented at each node. If the TTL reaches zero, the packet must be discarded (and an ICMP Time Exceeded error message must be sent to the source of the packet).
The TTL field is included in the flow identifier to ensure that a packet exits a switched flow with the same TTL that it would have if it were forwarded hop-by-hop. This ensures that only packets with a single, specific, TTL value may be included in a switched flow. It also ensures that a packet with a TTL of zero will never be switched through a node. Thus at the end of a switched flow, the TTL of packets on that flow must be correct; the TTL field is not transmitted in the packet, but is recovered from information stored at the destination. The price of this solution is an increase in the number of flows created, as two packets, that are identical, except for the value of the TTL field, will be transmitted in two separate flows.

In order to preserve the value of the header checksum, the value of the TTL field is subtracted from the header checksum of packets at the origin of a switched flow (any node where packets arrive non-switched and depart switched is the origin of a switched flow). The header checksum is reconstructed at the end of a switched flow by adding the value of the TTL field to the checksum when the packet header is reconstructed. This operation is necessary because the number of upstream IP switch nodes is unknown at the destination of a switched flow, and may indeed change if more upstream IP switches decide to switch a particular flow. The effect of this operation - assuming no errors are introduced in the transmission path - is that the header checksum contains the value it would have contained had the packet been forwarded hop-by-hop (and the TTL decremented and the checksum updated by all the upstream nodes). Any errors introduced into the IP header along the transmission path will result in an incorrect header checksum. This will cause the packet to be rejected the next time the header checksum is checked.

While the routing protocols are converging to a consistent state, it is possible for temporary routing loops to exist. A switched flow that is established while a loop exists in the routing state may traverse the same link multiple times, using a different VCI each time. Eventually the Time To Life (TTL) will be decremented to zero and the flow will not be switched further. When the routing state becomes stable, a switched loop will be detected and removed. While a switched loop exists, it consumes VCIs that could have been used for other traffic.

Multicast

An IP switch can support IP multicast without any modification to the Internet Group Management Protocol (IGMP) or the multicast routing protocols. Flow redirection proceeds in exactly the same manner as for unicast traffic. At an IP switch, when an incoming multicast flow is replicated into a number of branches, each branch may be individually redirected by a downstream neighbour.

Simulation Results

Simulations were done with data from an FDDI ring connecting traffic from the San Francisco Bay Area to and from the Internet. The raw result is that by adding the ATM switch the routers are able to handle approximately 3.7 times more traffic [3, 6. Simulation Results, pages 8,9].

2.2 Cisco's Tag Switching

With Tag Switching [4,10] technology, networks can handle more traffic. This approach also means that ISP's and large enterprise networks can enjoy more benefits from the performance of Asynchronous Transfer Mode (ATM) switches, to provide internet and ATM services on the same platform.

By 'tagging' the first packet in a flow of data, subsequent packets of related data are expedited to the final destination. Request times and router processing are both minimised. Tag Switching uses a form of label swapping across packet or cell-based networks that involves three solution elements:

Tag edge routers

> Located at the boundaries of a network, edge routers perform network layer services and apply tags to packets. Traffic from multiple sources going to the same destination can share tags, avoiding the label explosion problem of current IP switching implementations.

Tag switches

> The ATM switches or routers within the network can switch tagged packets based on these tags. These network elements can also support full Layer 3 routing or Layer 2 switching in addition to tag switching.

Tag Distribution Protocol (TDP)

> Coexisting with standard network layer protocols including routing protocols, TDP distributes tag information between devices in a tag switching network. Since TDP decouples tag distribution from the data flows, tag switching can be used over a wide variety of media including ATM links, Packet-over-SONET (POS) links, Ethernet, Gigabit Ethernet, and others.

The tagging algorithms used by the tag edge routers provide great flexibility for network managers. Packets can be tagged for a specific destination or tagged to flow along specified routes for balancing loads on network routes. This traditional Layer 2 service can now be implemented on Layer 3 routers using Tag Switching. A third tagging alternative takes advantage of Tag Switching's ability to analyse source, destination, and other Layer 3 information. This fine-grain processing introduces the ability to control quality of service (QoS) for a specific source/destination flow of packets.

Software-Only Upgrade Expands Service Options

A Cisco software upgrade is enabling many tag switching benefits. Tag switching makes possible enhanced network services such as QoS. Network managers can use QoS as a powerful tool for distinguishing different levels of service among a single user base. By assigning tags to unique services, specialised levels of service can extended from the internet into an enterprise network, creating a business class internet.

Tag Switching Components

Tag switching consists of two components: forwarding and control. The forwarding component uses the tag information (tags) carried by packets and the tag forwarding information maintained by a tag switch to perform packet forwarding. The control component is responsible for maintaining correct tag forwarding information among a group of interconnected tag switches.

Forwarding Component

The fundamental forwarding paradigm employed by tag switching is based on the notion of label swapping. When a packet with a tag is received by a tag switch, the switch uses the tag as an index in its Tag Information Base (TIB).

If the switch finds an entry with the incoming tag equal to the tag carried in the packet, certain informations in the packet are changed. The switch replaces:

- the tag in the packet with the outgoing tag.
- the link level information (e.g., MAC address) in the packet with the outgoing link level information.

Finally the switch forwards the packet over the outgoing interface.

The simple forwarding procedure is thus essentially decoupled from the control component of tag switching.
New routing (control) functions can readily be deployed without disturbing the forwarding paradigm. This means that it is not necessary to re-optimise forwarding performance (by modifying either hardware or software) as new routing functionality is added.

Tag Encapsulation

Tag information can be carried in a packet in a variety of ways:

- as a small tag inserted between the layer and the Network Layer headers
- as part of the layer header, if the layer header provides adequate semantics (e.g., ATM)
- as part of the Network Layer header (e.g., using the Flow Label field in IPv6 with appropriately modified semantics).

Control Component

Essential to tag switching is the notion of binding between a tag and Network Layer routing (routes). At the extreme a tag could be bound to an individual application flow. A tag could also be bound to a multicast tree. The control component is responsible for creating that tag bindings, and then distributing the tag binding information among tag switches.

The control component is organised as a collection of modules, each designed to support a particular routing function. To support new routing functions, new modules can be added. The following describes some of the modules.

Destination-based Routing

Recall that with destination-based routing a router makes a forwarding decision based on the destination address carried in a packet and the information stored in the Forwarding Information Base (FIB) maintained by the router.

There are three permitted methods for tag allocation and Tag Information Base (TIB) management: (a) downstream tag allocation, (b) downstream tag allocation on demand, and (c) upstream tag allocation.

In all cases, a switch allocates tags and binds them to address prefixes in its FIB. In downstream allocation, the tag that is carried in a packet is generated and bound to a prefix by the switch at the downstream end of the link (with respect to the direction of data flow). In upstream allocation, tags are allocated and bound at the upstream end of the link. ‘On demand’ allocation means that tags will only be allocated and distributed by the downstream switch when it is requested to do so by the upstream switch. Methods (b) and (c) are most useful in ATM networks. Note that in downstream allocation, a switch is responsible for creating tag bindings that apply to incoming data packets, and receives tag bindings for outgoing packets from its neighbors. In upstream allocation, a switch is responsible for creating tag bindings for outgoing tags, i.e. tags that are applied to data packets leaving the switch, and receives bindings for incoming tags from its neighbors.

The downstream tag allocation scheme operates as follows: for each route in its FIB the switch allocates a tag, creates an entry in its Tag Information Base (TIB) with the incoming tag set to the allocated tag, and then advertises the binding between the (incoming) tag and the route to other adjacent tag switches.

The downstream tag allocation on demand scheme, operation is as follows. For each route in its FIB, the switch identifies the next hop for that route. The router then issues a request (via TDP) to the next hop for a tag binding for that route. When the next hop receives the request, it allocates a tag, creates an entry in its TIB with the incoming tag set to the allocated tag, and then returns the binding between the (incoming) tag and the route to the switch that sent the original request. When the switch receives the binding information, the switch creates an entry in its TIB, and sets the outgoing tag in the entry to the value received from the next hop.

The upstream tag allocation scheme is used as follows. If a tag switch has one or more point-to-point interfaces, then for each route in its FIB whose next hop is reachable via one of these interfaces, the switch allocates a tag, creates an entry in its TIB with the outgoing tag set to the allocated tag, and then advertises to the next hop (via TDP) the binding between the (outgoing) tag and the route. When a tag switch that is the next hop receives the tag binding information, the switch places the tag (carried as part of the binding information) into the incoming tag of the TIB entry associated with the route.

Flexible Routing (explicit routes)

One of the fundamental properties of destination-based routing is that the only information from a packet that is used to forward the packet is the destination address. While this property enables highly scalable routing, it also limits the ability to influence the actual paths taken by packets. This, in turn, limits the ability to evenly distribute traffic among multiple links, taking the load off highly utilised links, and shifting it towards less utilised links. For Internet Service Providers (ISPs) who support different classes of service, destination-based routing also limits their ability to separate different classes with respect to the links used by these classes. Some of the ISPs today use ATM to overcome the limitations imposed by destination-based routing. Tag switching, because of the flexible granularity of tags, is able to overcome these limitations without using ATM. To provide forwarding along the paths that are different from the paths determined by the destination-based routing, the control component of tag switching allows installation of tag bindings in tag switches that do not correspond to the destination-based routing paths.

Tag Switching with ATM

Since the tag switching forwarding paradigm is based on label swapping, and since ATM forwarding is also based on label swapping, tag switching technology can readily be applied to ATM switches by implementing the control component of tag switching. The tag information needed for tag switching can be carried in the VCI field. (If two levels of tagging are needed, then the VPI field could be used as well.)

To obtain the necessary control information, the switch should be able to participate as a peer in Network Layer routing protocols. Therefore, an ATM switch can support tag switching, but at the minimum it needs to implement Network Layer routing protocols, and the tag switching control component on the switch.

Quality of Service

Two mechanisms are needed for providing a range of qualities of service to packets passing through a router or a tag switch. First, a classification of the packets into different classes. Second, the handling of packets such that the appropriate QoS characteristics (bandwidth, loss, etc.) are provided to each class needs to be ensured.

Initial classification would be done using information carried in the network layer or higher layer headers. A tag corresponding to the resultant class would then be applied to the packet. Tagged packets can then be efficiently handled by the tag switching routers in their path without needing to be reclassified.

Tag Switching Migration Strategies

Since tag switching is performed between a pair of adjacent tag switches, and since the tag binding information could be distributed on a pair-wise basis, tag switching could be introduced in a fairly simple, incremental fashion. Since tag switches use the same routing protocols as routers, the introduction of tag switches has no impact on routers. In fact, a tag switch connected to a router acts just as a router from the router's perspective.

2.3 NHRP (Next Hop Resolution Protocol)

2.3.1 Description

Next Hop Resolution Protocol (NHRP) [5,6,7,8] is an address resolution mechanism that avoids extra hops in an NBMA network (Non-Broadcast Multiple Access networks, such as ATM, SMDS and X.25.), therefore constructing a short-cut route. NHRP is not a routing protocol, or replaces an existing routing protocol.

To establish an NHRP short-cut route the following steps are taken:

1. A normal IP route is established.

2. In NBMA networks a short-cut is established with NHRP. The task is to find the short-cut route either by hop or by cost.

NHRP works over inter-LIS (Logical Independent Subnet). Connections host-to-host, host-to-router and router-to-host can be established outside the inter-LIS, but a router-to-router short-cut connection can make loops and therefore this is an issue for further research.
NHRP needs to distinguish between large flows that need short-cut routing and short-time duration flows or single packets that need hop-by-hop routing. An application can to choose between these options, what is expected to be implemented as a new functionality in sockets, otherwise the router makes these decisions.

The problem with an IP packet travelling hop-by-hop over the network is that meanwhile a router can try to establish an NHRP short-cut route over the NBMA network. This can result in packet misordering. To avoid this situation, some precautions need to be taken. Therefore the following entities can form a NHRP request only to establish a short-cut path:

- The Next Hop Resolution Client (NHC) connected to the NMBA network.

- The first router on the routing path of the IP-packet, so that the next-hop router can be found through the NBMA interface of this router only.

- A policy router in an NBMA network through which a IP-packet will go through.

When an IP-packet wants to travel over an ATM network the ATM address needs to be resolved prior of sending this packet. The IP-address is sent from the ATM host at the border ATM network to an ATMARP (ATM Address Resolution Protocol) server that resolves the IP address and delivers the IP address as an ATM address semantic to the requesting ATM host. With this information a Switched Virtual Circuit (SVC) is established expanding over the ATM network.

In NBMA subnets the ATMARP is exchanged with an Next Hop Resolution Server (NHS) that reacts on both ATM and NHRP resolution requests.

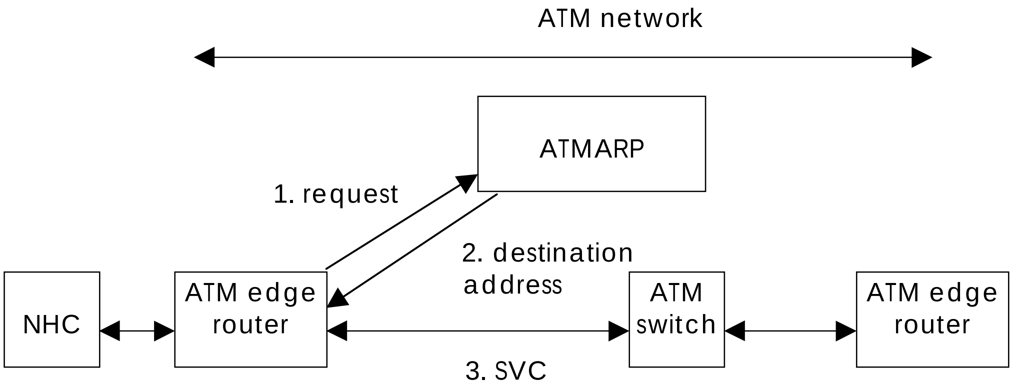


Figure 9: NHRP address resolution

The Next Hop Resolution Client (NHC) must make local / remote tests. In a local LIS the ATMARP server sends a signal back whether there is a shorter path over the ATM network (ACKnowledge packet ACK); or not (Not AcKnowledge packet NAK). Is there no shorter path, the packets are routed over the former routed path. NHRP send its request and reply packets over Internet Control Message Protocol (ICMP, see next section) messages that can be distinguished by the ICMP type in the header.
When the NBMA next-hop address is resolved the source can:

- Send packets to the destination (for connectionless NBMA) or
- Establish a connection to the destination with QoS parameters (bandwidth, delay, etc.) for connectionoriented NBMA, like ATM.

A forwarding table is constructed at every NHS, thus the decision for future packets is easy. The next-hop router doesn't needs to be resolved again.
The next NHS chosen by NHRP can depend on the QoS parameters desired (cost, shortest-path, etc.), thus constructing different routes for links with same QoS parameters is possible. In this case the links cannot be aggregated.

2.3.2 ICMP (Internet Control Message Protocol)

Before the NHRP Packet format can be described Internet Control Message Protocol (ICMP) [2,30] must be described.

ICMP echo packets are used within the internet for management functions, like ping and traceroute but also when something unexpected occurs, this event is reported by ICMP. Each ICMP message type is encapsulated in an IP-packet. About a dozen types of ICMP messages are defined and now two new messages for NHRP are defined (NHRP request, NHRP reply). Some of the most important ones are described below.

Message type	Description
Destination unreachable	Packet could not be delivered
Time exceeded	Time to live field is 0
Parameter problem	Invalid header field
Source quench	Throttle packet speed
Redirect	Teach a router the topology
Echo request	Ask a machine if it is alive
Echo reply	Machine is alive
Timestamp request	Same as Echo request, but with timestamp
Timestamp reply	Same as Echo reply, but with timestamp
NHRP request	NHRP short-cut request
NHRP reply	NHRP short-cut reply

Table 3: The principle ICMP message types

The **Destination unreachable** message is used when the subnet or the router cannot locate the destination.

The **Time exceeded** message is sent when a packet is dropped due to its counter reached zero. This event is a symptom the packets are looping, that there is enormous congestion, or the timer values are being set too low.

The **Parameter problem** messages indicates that an illegal value has been detected in a header field. This problem indicates a bug in the sending host's IP software, or in the software of a router transited.

The **Source quench** message was formerly used to throttle hosts that were sending too many packets. It is rarely used any more, because congestion control is done largely in the transport layer.

The **Redirect message** is used when a router notices that a packet seems to be routed wrong. It is used by the router to tell the sending host about the probable error.

The **Echo request** and **Echo reply** messages are used to see if a given destination is reachable and alive. Upon receiving the echo message, the destination is expected to send and echo reply message back. The **Timestamp request** and **Timestamp reply** message are similar, except that the arrival time of the message and departure time of the reply are recorded in the reply. This facility is used to measure network performance.

The **NHRP request** and **NHRP reply** messages are described in the next chapter.

In addition to these messages, there are four others that deal with Internet addressing, to allow hosts to discover their network numbers and to handle the case of multiple LANs sharing a single IP address.

2.3.3 NHRP Packet Format

ATM is treated from IP like an IP-layer. Therefore there must be always an IP router at the edge of an NBMA (here: ATM) network. NHRP has two packet formats that stay within an logical NBMA network.
These are NHRP request [6] and NHRP reply [6] that both are sent as ICMP messages.

Note that the code section indicates whether a server may answer with an cached information or it has to request the information from intermediate NHS from the path. This is called an authoritative answer.

NHRP Request

1	2	3	4
Type	Code	Checksum	
Hop Count	Unused		
Destination IP address			
Source IP address			

Figure 10: NHRP Request

Type 19

Code A response to an NHRP request may contain cached information. If an authoritative answer is desired, then code 2 (NHRP request for authoritative information) should be used. Otherwise, a code value of 1 (NHRP request) should be used.

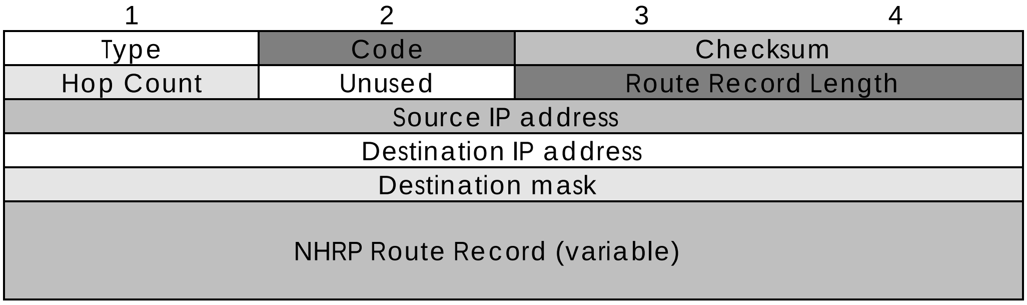
Checksum Checksum for save transmission.

Hop Count The Hop Count indicates the maximum number of NHS that a request or reply is allowed to traverse before being discarded.

Unused For future purposes.

Source and Destination IP addresses
These are the IP addresses of the NHRP request initiator and the terminal for which the NBMA next hop is desired.

NHRP Reply



1	2	3	4
Type	Code	Checksum	
Hop Count	Unused	Route Record Length	
Source IP address			
Destination IP address			
Destination mask			
NHRP Route Record (variable)			

Figure 11: NHRP Reply

Type 20

Code An NHS is not allowed to reply to an NHRP request for authoritative information with cached information, but may do so for an general NHRP request. NHRP replies may be positive or negative. An NHRP positive, non-authoritative reply carries a code of 1, while a positive, authoritative reply carries a code of 2. An NHRP negative, non-authoritative reply carries a code of 3 and a negative, authoritative reply carries a code of 4.

Checksum Checksum for save transmission.

Hop Count The Hop Count indicates the maximum number of NHS that a request or reply is allowed to traverse before being discarded.

Unused For future purposes.

Route Record Length
The length in words of the NHRP route record (see below).

Source IP address
The address of the initiator of the corresponding NHRP request.

Destination IP address and mask
If the NHRP request's destination is on the NBMA, the reply contains that destination address and a mask of all 1s. Otherwise, the responder may choose to act as the egress router for all terminals in the destination's subnet. If so, the reply contains a prefix of the requested destination IP address and the corresponding mask.

NHRP Route Record
The NHRP route record is a list of NHRP "Route elements" for NHS on the path of a positive NHRP reply. Only NHS that are willing to act as egress routers for packets from the source to the destination insert a route element in the NHRP reply. Negative replies do not carry route elements.

NHRP Route Record

An NHS may cache replies containing a route record [6]. Subsequently, when it responds to an NHRP request with the cached reply, intermediate NHS on the path to the initiator may attach route elements to the reply.

The first route element is always that of the destination terminal or, if the destination is not directly attached to the NBMA, that of the responding egress router. Each route element is formatted as follows:

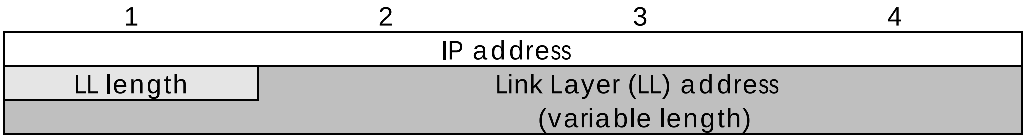


Figure 12: NHRP Route Record

IP address Destination terminal or the responding egress router.

LL length The LL length field is the length of the link layer address in bits.

Link Layer (LL) address
The LL address itself is zero-filled to the nearest 32-bit boundary.

On the reply path, an NHS willing to route packets from source to the destination should append its route element to the current route record, adjust the route record length appropriately and recompute the ICMP checksum.

If the first route element's IP address and the destination's IP address differ, the source terminal may assume that the reply was generated by an egress router.

2.3.4 Supported Features

NHRP supports a mechanism to aggregate NBMA next-hop informations in NHS cache. A certain kind of Virtual Path (VP) is established along the path to the destination with information caching in every router. This special VP is called **Switched Virtual Circuit (SVC)**.
Every NHS has an NHRP forwarding table for each QoS type and one or more cache tables for every connected network.

To receive the shortest path even when topology changes the

- Source terminal can periodically resolve the NHRP request or
- The router sends the next-hop information to the source terminal when the router has found a shorter path.

The terminals can therefore be moved to another place within an NBMA network.

To avoid asymmetric traffic flows (asymmetric means the back and forth routes are not the same) all traffic flows are mapped to one short-cut.

Proposed is:

- Map the IP packets with the DS Byte to corresponding SVC.
- Map class(es) of service to a single NBMA subaddress.
- Protocol extension, that short-cuts are used in both directions (less overhead).

Note: DS Byte: A small bit-pattern in each packet, in the IPv4 ToS octet or the IPv6 traffic class octet, is used to mark a packet to receive a particular forwarding treatment, or per-hop behaviour, at each network node.

2.3.5 Example

Now an examples for the working method of NHRP:

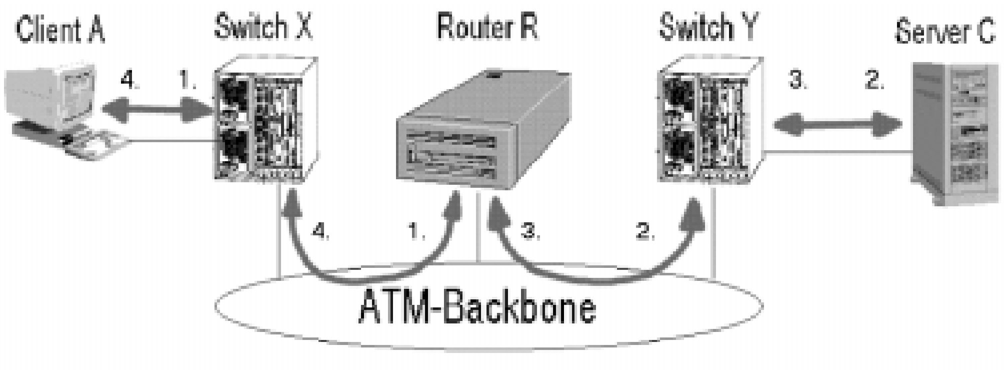


Figure 13: Hop-by-hop routing

Hop-by-hop routing is the commonly known IP routing. This takes place before NHRP established a short-cut route.

1. Client A sends data to Server C’s IP address.
2. The router R forwards data to Server C.

The same procedure comes upon in the reverse direction:

3. Server C sends data to Client A’s IP address.
4. Router R forwards data to Client A.

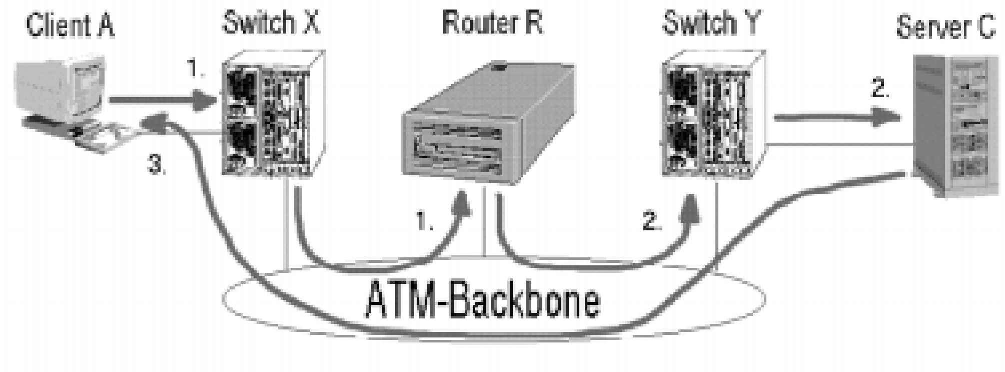


Figure 14: Short-cut Connection Setup

To set-up a short-cut connection the following steps are required:

1. Client A sends an NHRP request to Server C's IP address via router R (NHRP request contains A's MAC).
2. Router R forwards request to Server C.
3. Server C sends NHRP response directly to Client A using A's MAC address.

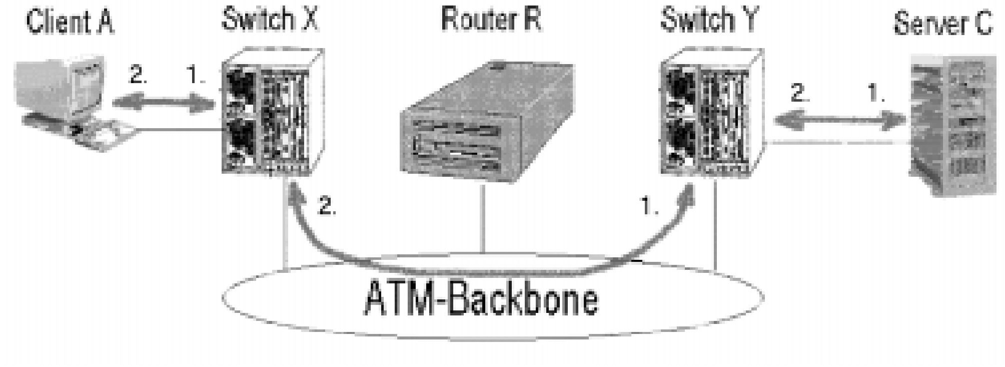


Figure 15: Short-cut Switching

To establish a full duplex short-cut switch where router R is not any longer used for this connection (Client A to Server C and reverse), the same procedure as above has to be done on reverse direction.

1. Client A sends data directly to C's MAC address.
2. As well as Client A sends data directly to Server C via C's MAC address.

3. Differentiated Services and ATM

3.1 Architecture

Differential Service [13] mechanisms allow providers to allocate different levels of service to different users of the internet. Any traffic management or bandwidth control mechanism that treats users differently is a service differentiation. However, in common internet usage the term is coming to mean any relatively simple, lightweight mechanism that does not depend entirely on per-flow resource reservation. The precise range of definition is still a matter of debates in the IETF Differentiated Services (diffserv) Working Group.

There is a clear need for relatively simple and coarse methods of providing differentiated classes of service for Internet traffic, to support various types of applications, and specific business requirements. The differentiated services approach to providing quality of service in networks employs a small, well-defined set of building blocks from which a variety of services may be built. A small bit-pattern in each packet, in the IPv4 ToS octet or the IPv6 traffic class octet, called the ‘DS Byte’, is used to mark a packet to receive a particular forwarding treatment, or per-hop behaviour, at each network node. A common understanding about the use and interpretation of this bit-pattern is required for inter-domain use, multi-vendor interoperability, and consistent reasoning about expected service behaviours in a network.

There is a new architecture required to support QoS on networks. This architecture exists of the domains that contains nodes at the edge of the domain (DS edge node) and in the interior of the domain (DS interior node). With the DS Bytes set in the packets it is possible to make reservations for flows over the domains. It is expected that the border nodes have some traffic conditioners implemented that flat-out bursts and cash packets. This has the advantage that interior nodes have to supply the reserved bandwidth and QoS parameters only therefore protecting the domain from congestion. The interior nodes have simple forwarding mechanisms implemented but do not need the whole set of traffic conditioners like edge nodes. With this architecture congestions can be effectively challenged. Most of the proposals refer to this architecture [15].

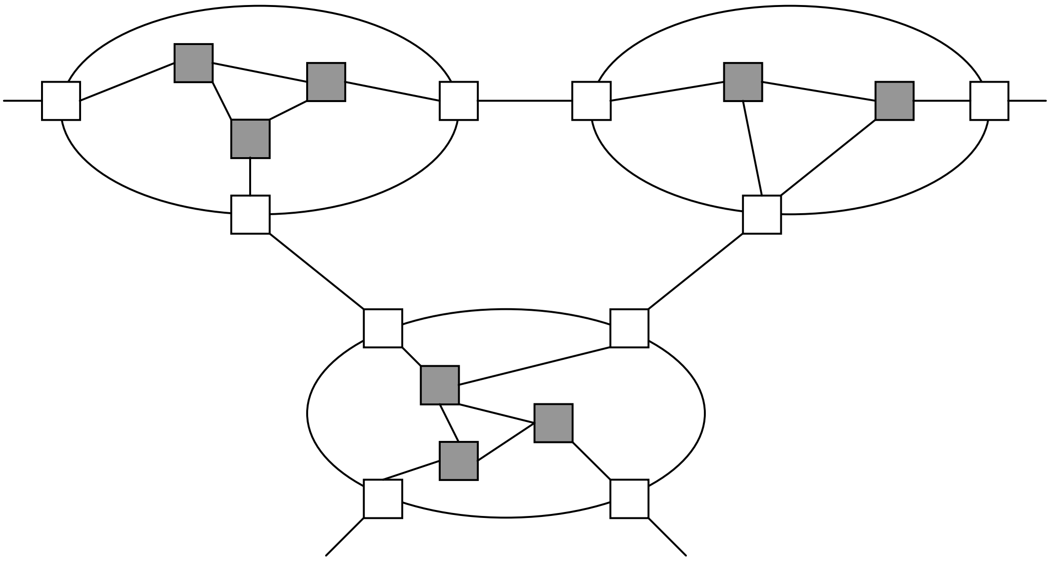


Figure 16: Differentiated Services Architecture

Note:

□ ■	DS node	A DS capable node.
□	DS edge node	A DS node that is not a DS interior node.
■	DS interior node	A DS capable node in the interior of a DS domain.
○	DS domain	A contiguous set of nodes which operate with a common set of service provisioning policies and PHB definition; it consists of DS interior nodes and DS edge nodes.

3.1.1 Classification, Marking, Policing, Shaping

Sophisticated classification, marking, policing and shaping operations need only be implemented at network boundaries or hosts. Network resources are allocated to traffic streams by service provisioning policies which control how traffic is conditioned upon entry to a differentiated services-capable network, and how that traffic is forwarded within that network. A wide variety of services can be implemented on top of these building blocks.

Traffic Conditioning

Traffic conditioning functions are performed by DS edge nodes in a DS domain to ensure that the traffic entering a DS domain conforms to the rules specified in the traffic conditioning agreement (TCA), and to prepare the traffic for the Per-Hop Behaviour (PHB) based forwarding treatment in the interior routers.

3.1.1.1 General Architecture of Traffic Conditioners

The differentiated services architecture is based on a simple model where traffic entering a network is conditioned at the edges of the network [9], and assigned to different behaviour aggregates (Behaviour Aggregate (BA) classifiers which classify only on patterns in the DS Byte). Each behaviour aggregate is identified with a single DS codepoint. Within the core of the network, packets are forwarded according to the per-hop behaviour associated with the DS codepoint.

A traffic conditioner, which is supposed to be located at the edge of a network, may contain the following elements: classifier, meter, marker and shaper. The classifier and the meter select the packets within a traffic stream and measure the stream against a traffic profile. The marker and shaper perform control actions on the packets depending on whether the traffic stream is within its associated profile.

A packet stream normally passes to a classifier first, and the matched packets are measured by a meter against the profile as defined in the TCA. The packets within the profile may leave the traffic conditioner or may be marked by the marker. The packets that are out-of-profile may be either marked or shaped according to the rules specified in the TCA. Note that discard policing can be performed by a specially configured shaper. When packets leave the traffic conditioner of a DS ingress node, the DS Byte of each packet must be set to one of DS codepoints defined by the PHB groups supported in the DS domain.

Figure 17 shows the block diagram of a traffic conditioner. Note that a traffic conditioner may not necessarily contain all four elements. For example, packets may pass from the classifier directly to the marker or shaper (null meter).

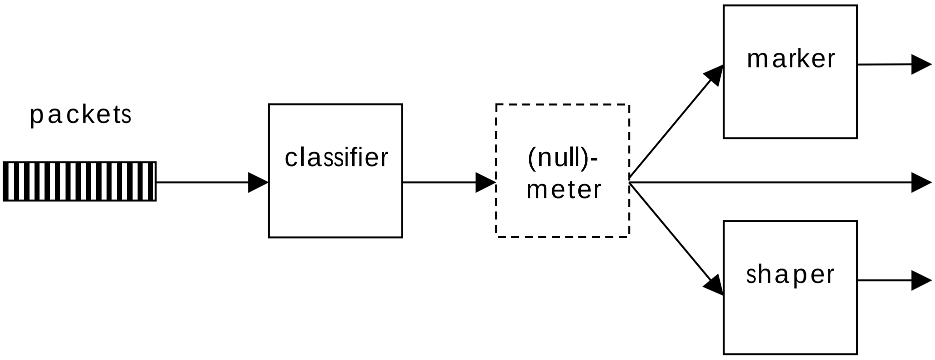


Figure 17: Logical View of a Traffic Conditioner

Traffic Conditioning Agreement (TCA)

Differentiated services are extended across a DS domain boundary by establishing a service level agreement (SLA) between the customer and provider DS domains. The SLA includes a traffic conditioning agreement which usually specifies traffic profiles and actions to in-profile and out-of-profile packets.

Traffic Profiles

A traffic profile specifies rules for classifying and measuring a traffic stream. It identifies what packets are eligible and rules for determining whether a particular packet is in-profile or out-of-profile. For example, a profile based on token bucket may look like:

codepoint = X, use token-bucket r, b

The above profile indicates that all packets in the behaviour aggregate with DS codepoint X should be measured against a token bucket meter with rate r and burst size b. In this example out-of-profile packets are those packets in the behaviour aggregate which arrive when insufficient tokens are available in the bucket. Different conditioning actions may be applied to the in-profile packets and out-of-profile packets, or different accounting actions may be triggered.

Actions to In-Profile and Out-of-Profile Packets

In-profile packets may be allowed to enter the DS domain without further conditioning as they conform to the TCA.

The actions to out-of-profile packets may include delaying the packets until they are in-profile (shaping), discarding the packets, marking the DS field to a particular codepoint, or triggering some accounting action.

3.1.1.2 Components of a Traffic Conditioner

Components of traffic conditioner [9] are classifier, meters, markers and shapers:

Classifiers

Packet classifiers select packets in a traffic stream based on the content of some portion of the packet header. The classification may be based on the DS field only (Behaviour Aggregate Classification), or on any combination of one or several fields in the packet header such as source address, destination address, DS field, protocol ID and transport-layer header fields such as source port and destination port numbers (Multi-Field Classification). Classifiers are used to steer packets matching some specified rule to another element of the traffic conditioner for further processing. Classifiers must be configured by some management procedure in accordance with the appropriate TCA.

Meters

Traffic meters measure the traffic properties of the set of packets selected by a classifier against a traffic profile specified in the TCA. A meter indicates to other conditioning functions whether each individual packet is in-profile or out-of-profile.

A null meter will identify all packets as in-profile. Such a meter may be used when the traffic profile does not specify conforming rate or burst parameters.

Markers

Packet markers set the DS field of a packet to a particular codepoint, adding the marked packet to a particular DS behaviour aggregate. The marker may be configured to mark all packets which are steered to it to a single codepoint, or may be configured to mark a packet to one of a set of codepoints within a PHB group according to the state of a meter.

Shapers

Shapers delay some or all of the packets in a traffic stream in order to bring the stream into compliance with its associated traffic profile. A shaper usually has a finite-size buffer, and packets may be discarded if there is not enough buffer space to hold the delayed packets.
Discard policers can be implemented as a special case of a shaper by setting the shaper buffer size to zero (or to a few) packets.

Location of Traffic Conditioners

Traffic conditioners may be located within a customer DS domain, and at the boundary of a DS domain. Traffic conditioners may also be located in nodes in a non-DS domain.

3.1.2 Cisco's Software Releases: Features supporting QoS

For implementation of network wide Quality of Service support multiple functions within the network are needed. Committed Access Rate (CAR) provides traffic classification and rate limiting in devices located towards the edge of network. Distributed Weighted Fair Queuing (dWFQ) and Distributed Weighted Random Early Detection (dWRED) ensure that the service guaranties are met with queuing and packet discard.

Internet Service Providers (ISP) are enabled to offer differentiated services at different tariffs. In enterprise networks these features allow for prioritisation of businesscritical applications.

Distributed Weighted Fair Queuing (dWFQ)

There are two forms of distributed Weighted Fair Queuing (dWFQ) [11]:

- Flow-based Weighted Fair Queuing (flow-based WFQ) controls the ratio of transmission bandwidth allocation among different traffic flows during periods of congestion.
  With flow-based WFQ, packets are classified by flow. Packets with the same source IP address, destination IP address, source TCP or UDP port, destination TCP or UDP port, protocol, and type of service (ToS) field belong to the same flow. Each flow corresponds to a separate output queue. When a packet is assigned to a flow, it is placed in the queue for that flow. During periods of congestion, WFQ allocates an equal share of the bandwidth to each active queue.

- Class-based Weighted Fair Queuing (class-based WFQ) allocates transmission bandwidth among different traffic flows or QoS groups during periods of congestion.
  In class-based WFQ, packets are assigned to different queues based on their QoS group or the IP precedence in the ToS field.
  A weight for each class must be specified. In periods of congestion, each group is allocated a percentage of the output bandwidth equal to the weight of the class. When the interface is not congested, queues can use any available bandwidth.

Drop policy is applied over both flow-based WFQ and class-based WFQ. WFQ keeps track of the number of packets in each queue and the total number of packets in all queues. When the total number of packets is below the aggregate limit, queues can buffer more packets than the individual queue limit.
When the total number of packets reaches the aggregate limit, the interface starts enforcing the individual queue limits. Any new packets that arrive for a queue that is over its individual queue limit are dropped. Packets that are already in the queue will not be dropped, even if the queue is over the individual limit.

Distributed Weighted Random Early Detection (dWRED)

Random Early Detection (RED) [12,26,37,38] is a congestion avoidance mechanism. By dropping some packets early rather than waiting until the buffer is full, RED avoids dropping large numbers of packets at once. TCP detects the dropped packets and starts throttling the source. Through this mechanism a congestion can be avoided. Thus, RED allows the transmission line to be used fully at all times.
In addition, RED statistically drops more packets from large users than small. Therefore, traffic sources that generate the most traffic are more likely to be slowed down than traffic sources that generate little traffic [36, section: Benefits].

Weighted RED (WRED) [11] generally drops packets selectively based on IP precedence. Packets with a higher IP precedence are less likely to be dropped than packets with a lower precedence. Thus, higher priority traffic is delivered with a higher probability than lower priority traffic.

When a packet arrives, the average queue size is calculated; it is based on the previous average and the current size of the queue.

- If the average is less than the minimum queue threshold, the arriving packet is queued.

- If the average is between the minimum queue threshold and the maximum threshold, the packet is either dropped or queued, depending on the packet drop probability.

- If the average queue size is greater than the maximum threshold, the packet is automatically dropped.

The packet drop probability is based on the minimum threshold, maximum threshold, and mark probability denominator. This denominator is a parameter to set.

When the average queue depth is above the minimum threshold, RED starts dropping packets. The rate of packet drop increases linearly as the average queue size increases until the average queue size reaches the maximum threshold.

When the average queue size is above the maximum threshold, all packets are dropped.

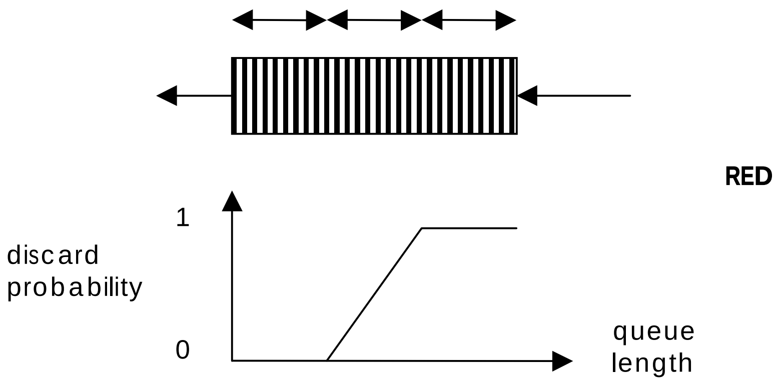


Figure 18: Random Early Detection (RED)

However, WRED is usually used in the core routers of a network, rather than the edge. Edge routers assign IP precedences to packets as they enter the network. WRED uses these precedences to determine how it treats different types of traffic.

WRED provides separate thresholds and weights for different IP precedences, allowing to provide different Qualities of Service for different traffic. Standard traffic may be dropped more frequently than traffic with a higher priority during periods of congestion.

NHRP Functionality

As described in section 2.3.

Generic Traffic Shaping

Topologies that have high speed links (e.g., a central site) feeding into lower speed links (e.g., at remote or branch sites) often experience bottlenecks at the remote end because of the speed mismatch. Generic Traffic Shaping helps eliminate the bottleneck situation by throttling back traffic volume at the source end.
Generic Traffic Shaping implements a Weighted Fair Queuing (WFQ) on an interface or sub-interface to allow the desired level of traffic flow.

Policy Routing

Where Quality of Service requirements or virtual private network (VPN) topologies dictate that traffic should be routed through specific paths, policy routing can provide the solution. By using policy routing, customers can implement policies that selectively cause packets to take different paths.

Policy routing also provides a mechanism to mark packets so that certain kinds of traffic receive differentiated, preferential service when used in

combination with queuing and congestion management techniques. These techniques provide an extremely powerful, simple, and flexible tool enabling network managers to implement policy within their networks. Policy routing allows to classify traffic based on extended access lists.
Traffic policies can be implement based on source and/or destination IP addresses, TCP port numbers, and/or packet lengths.
Internet Service Providers (ISPs) can use policy-based routing to route traffic originating from different sets of users through different internet connections across the policy routers.
Organisations can provide QoS to differentiated traffic by setting the precedence or type of service (ToS) values in the IP packet headers at the periphery of the network and leveraging queuing mechanisms to prioritise traffic in the core or backbone of the network.

3.2 Description of Differentiated Services

The Differentiated Services Working Group will standardise a common layout to be used for the ‘DS Byte’. A standards-track document will be produced that will define the general use of fields within the DS Byte (superseding the IPv4 ToS octet definitions of RFC 1349 [14]). The working group will also assign specific per-hop behaviours to a small number of particular patterns or 'code-points' of the DS Byte. The standardised code-points will only apply to per-hop behaviours already in widespread common usage within the global internet, e.g., the forwarding treatment received by best-effort traffic.

In addition to the standards-track specification document, an informational framework document will be produced. The framework document will define the differentiated services architecture and a common language for differentiated services.

The working group will also investigate the components necessary to support differentiated services, including traffic conditioners such as traffic shapers and packet markers that could be used at the boundaries of networks and finally the working group treats security threats and counter-measures.

Some proposals are already made which are following discussed in this paper.

3.2.1 Premium Service

Premium Service [15] is understood as a virtual leased line. These emulated leased lines are commonly used in telephone systems and to connect company plants over networks. Premium Service has a maximum bandwidth parameter. Therefore the bandwidth cannot be exceeded but the user or the company can let it idle or use it to the full extent of it s capacity. No influence from the presence or absence of other unsers shall be remarked by the holder of this line.

The **implementation** of Premium Service as a differentiated service (or the mapping to ATM) should get the same properties. These properties are: Peak bit rate (on flows or on aggregated flows), no bursts (only within the peak bit rate) and negligible queuing delay (for real-time applications). When Premium Service is travelling over ATM the QoS parameters that form a virtual leased line on Premium Service must be mapped to ATM QoS parameters to receive the same support of QoS over ATM.

Premium Service - as a leased line in diffserv - is a desire, because companies are still interested in virtual leased lines, so they won’t have to run and support an own network and because of the backward compatibility to existing applications (telephone).

Traffic classifying, marking, shaping and policing is needed, therefore if these functions are done in the edge routers almost no queuing in the routers inside the network (interior routers) is needed.

For Premium Service a token bucket is necessary to shape the packets to a steady packet rate. The P bit (P stands for Premium Service class) in the DS byte of the packet is reset (classification) and the packet is queued in the token bucket queue which size is n. There the packet is waiting for a token which is generated at a fixed rate. Then the P bit of the DS byte is set and the packet can travel over the net to the next-hop router again. The result of such a construction is to shape the packet stream to a steady packet rate.

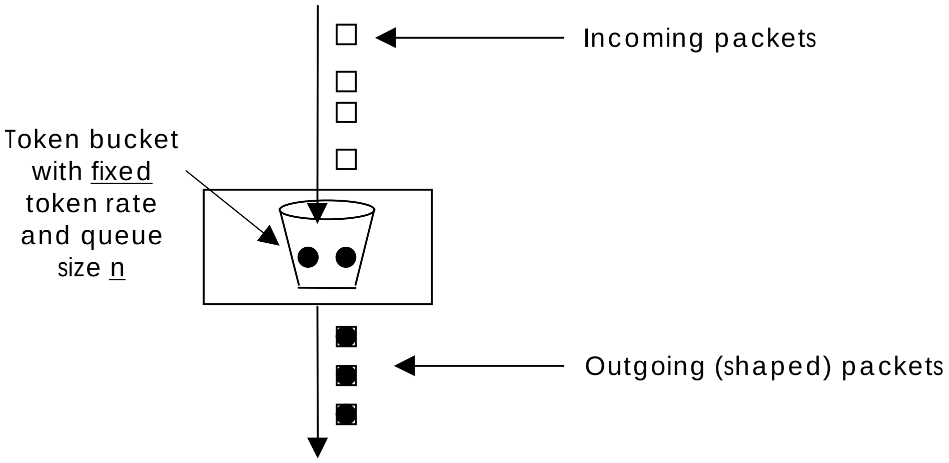

Incoming packets
Token bucket with fixed token rate and queue size n
Outgoing (shaped) packets

Figure 19: Token bucket for Premium Service

For an adequate operation the token bucket parameters need to be set appropriately. The Premium Service parameters are translated to the ATM parameters (similar to the translation of Guaranteed Service to ATM [39, section 2.5.1]). When Premium Service is mapped to CBR (which is the best mapping, see section 3.3.4) the ATM traffic parameters are set to the following values:

PCR = SCR = peak rate
MBS = 0

Who is interested in getting Premium Service: Telephone companies can still supply their customers (private users and companies) with virtual leased lines (normal telephone lines, ISDN or other future networks).
There is an additional aspect: With high speed cell switching networks, where flows are aggregated, the used capacity can better be distinguished from the unused capacity. That means, telephone calls need a very small fragment of the full capacity of the line and the lines are seldom fully used (usually one person speaks at a time). The excessive capacities can be telephone company internally used for data traffic or leased to other (even telephone-) companies for telephone or data transmission.
Internet Service Providers (ISP) can supply virtual leased lines to the customers (private or commercial).
With Premium Service Virtual Private Networks (VPNs) can be formed over ISPs. Companies with their offices scattered around the world (or just around the corner), can be connected over an ISP to form an VPN, therefore enlarging their Local Area Networks (LAN) to Metropolitan Area Networks (MAN) or Wide Area Networks (WAN). VPN ensure that the holder of this service is not disturbed from other users or traffic congestions on public nets. Security is furthermore an important aspect, that means the distinction between the lines needs to be save.

Generally this means not much difference to the now used leased lines or applications that run on such leased lines (like dial-up internet unsers), so not much needs to be changed on end-user side.

Pro: Well suited for all kinds of real-time applications. No starvation of links. No change to existing applications. VPN can be formed with Premium Service using virtual leased lines. Goes well with Assured and Best-Effort Service.

Contra: No bursts supported. No support for other kinds of applications with different needs.

3.2.2 Assured Service

Assured Service [16] guarantees no specific bandwidth but sends the packets labelled with a higher priority over the network (similar to Controlled Load on Integrated Services [17,18]).
A boundary device, probably a router, is measuring traffic and marking it either in-profile or out-of-profile. These packets are stored in a token bucket (described in section 1.1.5 Traffic policing, traffic shaping) which drops the packets which are out-of-profile first when the buffer is about to fill.
It is expected that traffic which is within the contracted rate has a very much reduced probability of being lost, while excessive traffic is more likely to be dropped. The user of this service can benefit from the available bandwidth and suffers less from congested networks, but discarded packets must be accepted occasionally.

Properties of Assured Service: No guaranteed bandwidth, but packets are labelled with high priority and the network routers will set the in-profile bit to out-of-profile when the packets are not within an agreed profile, that means the packets are early or queue overflow is encountered. Short-time bursts are supported with this behaviour.

It's agreed on contract basis how much high priority data can be sent. Otherwise - when to much high priority packets are sent - the ISP can mark the packets out-of-profile, discard them or charge the user for these excessive packets.
The delay for the Assured Service traffic is the same as for Best-Effort Service.

In congestion situations the user of Assured Service should encounter less bandwidth deficiency than Best-Effort users. With use of large buffers or smart queue management the bursts should be processed well. In backbones the traffic is multiplexed thus bursts are statistically compensated by experience. In congestion situations Best-Effort Service will starve first leaving Assured Service transmission intact. For Assured Service the lines should be estimated to hold these services even in fatal traffic congestion incidents.

For Assured Service a token bucket is necessary to shape the packets to a certain packet rate. This is done with resetting the A bit (A stands for Assured Service class) in the DS byte (classification). The token bucket rate is flexible and when a token is available the A bit is set in the packet header and the packet is put in a RIO queue (RED with In and Out [22]) with the in-profile bit set. When no token is available the packet is put in the RIO queue with the out-of-profile bit set. If the RIO queue starts to fill, packets with out-of-profile bit are discarded earlier than the packets with the in-profile bit. The result of such a construction is to drop the excessive packets (these are packets that don’t conform to the traffic agreement) earlier than packets which conform to the traffic agreement, therefore protecting the net from congestion.

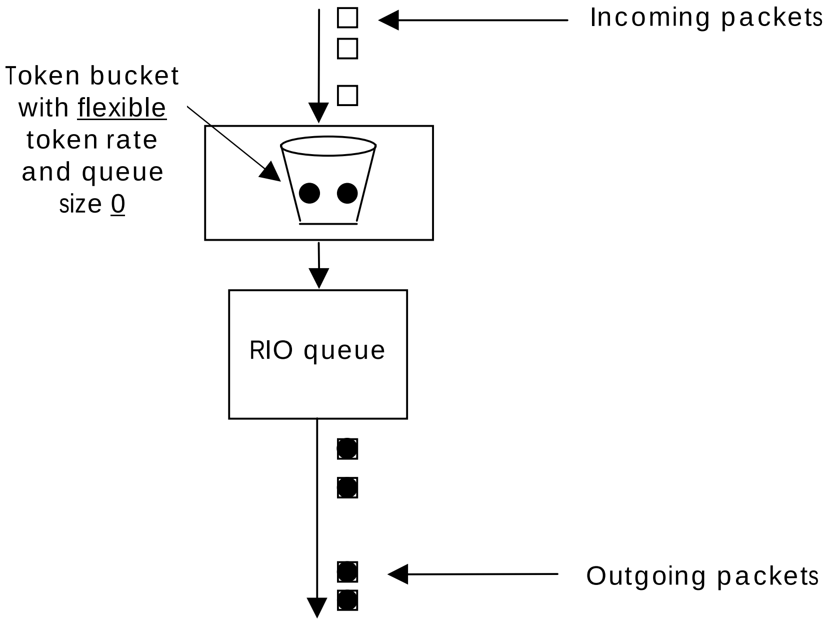


Figure 20: Token Bucket for Assured Service

For an adequate operation the token bucket parameters need to be set appropriately. The Assured Service parameters are translated to the ATM parameters (similar to the translation of Controlled Load Service to ATM [39, section 2.5.2]). When Assured Service is mapped to nrt-VBR (which is the best mapping, see section 3.3.4) the ATM traffic parameters are set to the following values:

PCR = peak rate
SCR = token bucket rate
MBS = token bucket depth

Who is interested in getting Assured Service: Banking transactions and the like are most of the time idle but when a transaction is made, priority treatment is required. Systems can gain better throughput when information is sent or required in a certain period of time with Assured Service.
Users won't always use the line, but can benefit from burst rate if they need it and getting better treatment of the packets.
Assured Service will bring advantages for many internet applications (adaptive audio and video applications), if the user is willing to pay more for this better treatment.

Pro: Well suited for applications with a critical response time. VPN can be formed with Assured Service. IP Telephone can also be realised. Goes well with Premium and Best-Effort Service.

Contra: No support for real-time applications. There is no profile (User Agreement); therefore the user agreement is enforced with policing at the border router, dropping packets until the user is in profile again.

3.2.3 User-Share Differentiation (USD)

At User-Share Differentiation (USD) [19] a minimum bandwidth is guaranteed, therefore the links will not starve. The excessive bandwidth will be distributed in a share ratio per user (like 3:2:1). This ratio will be based on per cost according to the users ISP contract. Now the user can be organised in several classes (for instance three classes: premium, basic and best-effort).

Packets may be reclassified at a new ISP. Routers need schedulers that supports proportional bandwidth sharing (CBQ [20], WFQ (see section 3.1.2). Each router needs to know the share ratio which is distributed over network management.
In congestion situations the user will be (as worst-case) degraded to his guaranteed minimum bandwidth.

Implementation will be made in bottleneck links only, so no changes in end systems is required. While the maximum granularity is per customer, ISPs can achieve different levels of aggregation by creating user classes, so the user can get one service class only. The flows can then be aggregated to a single flow at each ISP. Within the retail ISP, traffic from or to a dial-up user is aggregated whilst in the backbone only the retail ISP is visible, not the single customer link.

Who is interested in getting USD Service: Applications that can change their transmission rate and need a minimum bandwidth to transmit data. These are real-time applications with flexible (compress) ratio or applications that need an minimum bandwidth only (to avoid starvation) like ftp or airline reservation systems.
ISPs could also be interested to bring better service with minimal guaranteed bandwidth to their customers.

Pro: Fair share of resources and a guaranteed minimum bandwidth. No starvation of links. Simple implementation in bottleneck routers. No changes are required in end systems. Good solution for adaptive applications.

Contra: No support for real-time applications with need of constant bitrate. No delay rates can be guaranteed for longer routes that result from routing around bottleneck areas. Links cannot be aggregated to customer purposes.

3.2.4 Olympic Service

In this Olympic Service [21] proposal the bandwidth is shared among users that belong to a certain class. A proposition is to make Gold-, Silver- and Bronze-Classes that have a share of 60%, 30% and 10% respectively. The difference between USD and Olympic Service is that the service will be classified at the border router with setting an in-profile bit, which indicates that the traffic is within the agreed traffic specification. It’s possible, that when no Gold and no Silver services are present the Bronze Service takes all the capacity. That means the user gets always as much capacity the network can supply and the limitations take effect in congested times only. As a consequence no minimum bandwidth is guaranteed.

Implementation: Because the flows are not admission controlled shaped or policed, this Service would be good for an ISP or within a campus areas where the service provides different services to different customers or users. It’s simple to implement but it probably won’t scale well on large networks. A rate-based link share scheduler has to be implemented in every router. Proposed are also four weighted round robin queues with RIO [22] droppers in every queue. The length of the queues is not specified to let open the ability to support different applications with queue length variation.

In congestion situation the Gold Service gets more capacity than the Silver Service and even more than the Bronze Service. Starvation should not be a problem but there is no minimum bandwidth guaranteed, so that all service classes might be squeezed to an insupportable small bandwidth.

Who is interested in getting Olympic service: ISPs who want to bring service discrimination to their customers. They can vary their service profiles as needed for the customer. Olympic Service can also suit well in campus areas where different service is desired.

Pro: Simple implementation in ISPs or campus area, therefore fitted for many applications for service discrimination.

Contra: The links will not starve but if many users are present, the bandwidth of a Bronze Service will be diminished to almost zero.

3.2.5 Scaleable Reservation Protocol (SRP)

In Scaleable Reservation Protocol (SRP) [23] the path will be reserved for a specific rate and the router agrees to support this rate as long as the line is hold. SRP consists of two protocols, a reservation protocol, which makes the reservation over the path and a feedback protocol, that sends confirmation back to the sender.
The reservation protocol consists of three packet types which can be discriminated through the codepoint in the header:

Request: These packets try to reserve network resources in the routers for the following packets that come with 'Reserved' codepoints. These packets are used when no request was sent yet or the feedback hasn't reached the sender yet.

Reserved: Packets with the 'Reserved' tag must be forwarded if the reservation is established. Otherwise they can de discriminated to 'Request' or 'Best-Effort' Packets.

Best-Effort: No resource reservations for this type of traffic. The packets marked with 'Best-Effort' may be dropped in congested periods.

The flows can be aggregated to a single flow at any router, therefore it is well suited to go with ATM. The endsystems play an active role by establishing reserved links. The following Figure shows how the specific packets can be degraded to a different level through each router.
In congestion situations the 'Request' packets can be degraded to 'Best-Effort' packets and the 'Reserved' packets to 'Request' packets and also to 'Best-Effort' packets. But in these cases the source will still try to make a reservation with future 'Request' and then 'Reserved' packets [24].

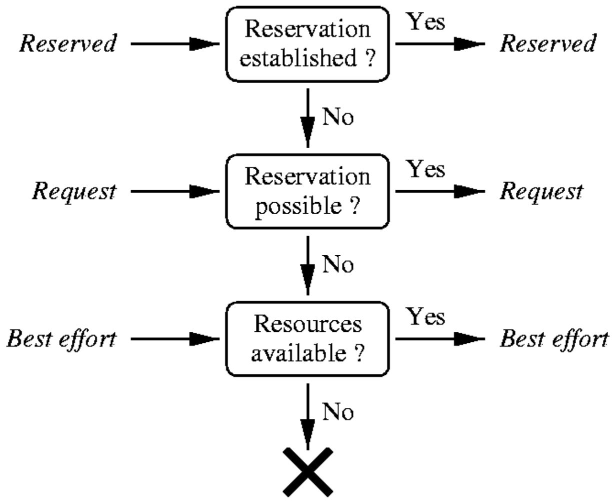


Figure 21: Packet type processing by routers

Implementation: All system routers and endsystems need to be changed. SRP works with estimators in every system part to estimate the future datastream. SRP packets use the DS Byte to classify the packets. The first block of data is sent with the first reservation packet to establish a route. If the establishing of a link would mean a degrade to other links, the new requested path will not be established in terms of admission control.

Who is interested in getting SRP service: IP Telephony as well as Telephone systems could benefit from this architecture. Telephone lines could use the reserved data stream and other data traffic could use best-effort traffic. ISPs could provide a service like a leased line to company customers.

Pro: SRP has guaranteed performance to already established links and no starvation of links is encountered, its well scaleable and well suited for large amounts of flows. The proposal has considerations on multicast, too.

Contra: Cannot be used with bursty traffic; the estimators are not reacting fast enough. If there is no datastream from the source over a longer period, the link will be cut.
Difficult and costly implementation of the estimators (they learn by example). With the different protocols travelling back and forth over the network to establish and maintaining the links, the network is heavy loaded with overhead traffic.
Not suitable for Virtual Private Networks (VPN) because of the aggregation of flows on each link in the network. The network has therefore no knowledge of individual flows.
Policing on network boundaries is another not further studied issue.

3.2.6 Best-Effort Service

Best-Effort Service guarantees no bandwidth. The user has to deal with the bandwidth available. He can profit from high bandwidth in little used times but must suffer from congestions in heavy used times, therefore it is a cheap service. The user is totally dependent from other users and their use of capacity. This is the service which is now and in future available on the internet.

No **implementation** specific issue: ‘as-is’ service. Though some improvements could be: Fair queuing of remaining bitrate (Weighted Fair Queue WFQ, see section 3.1.2) and better buffer queuing in the router (Random Early Detection RED, see section 3.1.2) could help avoid congestion.

No traffic classifying, marking, shaping or policing is made. No traffic isolation between users is possible. Anyhow Best-Effort Service can be aggregated to a single flow.

Who is interested in getting Best-Effort Service: Best-Effort Service will still be the bulk traffic on the internet, because it will not need any further implementations and if ISPs supply enough bandwidth then the user will not switch to more expensive services like virtual leased lines or high priority forwarded packets, because best-effort is cheap and most of the time it does what it supposed to do.
Even firms can run their banking transmissions backup, mirror server update, software upgrade of server/clients or ftp through cheap best-effort traffic. With the world wide effort to upgrade to even faster lines, Best-Effort Service is profiting from this effort as well.

Pro: Good for links that can work independently from the load of the network. Applications and users can profit from the high inexpensive bandwidth when it is available. Applications that can adjust to a lower bitrate when congestion occurs can use Best-Effort Service as well as any other service.

Contra: No QoS support at all (for real-time or priority based applications). Starvation of links is possible.

3.2.7 All Differentiated Services at a glance

	Premium	Assured	USD	Olympic	SRP	Best-Effort
Max. Bandwidth	Yes	No	No	No	Yes	No
Min. Bandwidth	No	No	Yes	No	Yes	No
Bursts	No	Yes	Yes	Yes	No	Yes
Queuing Delay Guaranty (for rt-Applications)	Yes	No	No	No	Yes	No
Marking (packets labelled with higher priority)	Yes	Yes	No	Yes	Yes	No
Classified at boundary / inside of network	boundary	boundary	inside	boundary	boundary	No class.
Shaping	Yes	Yes	No	No	Yes / No	No
Policing	Yes	Yes	No	No	Yes	No
Admission Control	Yes	Yes	No	No	Yes	No
Aggregation	Yes	Yes	Yes	Yes	Yes	Yes

Table 4: Characteristics of the proposed services

3.3 Mapping of Differentiated Services on ATM Service Categories

3.3.1 Criteria for a good mapping

Some basic considerations for the mapping of differentiated services on ATM service categories are made here:

- Insufficient capacity on one part does not lead to an wastage of capacity on the other part and vice versa.

- The wastage of resources is an important point for a good mapping. In some cases a good service is supplied while resources are wasted. This is an either expensive or a senseless service.

- Requirements such as real-time transmission, burst support must be supported on both - ATM and diffserv - part to make a useful coupling of ATM and diffserv.

- A minimal bandwidth should not be supported on one part while the other part knows link starvation.

- Is the service guaranteed. This means the resources can be held as long as there is a demand for, no link starvation nor bandwidth degradation or change of any other traffic or QoS parameter is encountered.

3.3.2 Some Problems with ATM in conjunction with Differentiated Services

How can an appropriate Quality of Service be established?

A router needs to know what kind of application to send over which connection. This can be presented to the router from the application which selects a specific socket functionality to get the appropriate service for the application or the router can distinguish the sort of application by the used port.
The forwarding behaviour needs to be selected, the parameters need to be set accordingly and propagated to lower layers and ATM Service Categories. The resulting problem of classifying packets at routers to support the appropriate QoS treatment should not be underestimated.

Can a new traffic condition be negotiated when there is a congestion?

Let's assume that an application uses Best-Effort Service in not congested situations, because it is cheap and sufficient, while this application switches to a better Quality of Service when congestion occurs for transmission safety reasons. This behaviour can or cannot be supported according to the traffic contract and admission control with the Internet Service Provider ISP.
In the paper Adaptive Packet Marking [25] such an proposal is already made. In this case the end-to-end signalling and enforcing of traffic profiles is eliminated. A Packet Marking Engine (PME) is monitoring the packets with the different ToS bits. If the throughput falls under a minimum target rate, the PME starts to prioritising packets - with marking - until the desired target rate is reached. The marking probability is dependant on the observed bandwidth. Two classes of priority are present: priority and best-effort.

What happens when a customer does not hold to the traffic commitment?

This is a problem because the network could be flooded with traffic that should not be sent and the flows therefore not be guaranteed with its QoS parameters. The network components cannot estimate the future traffic which is the present situation in the internet. In future the network components need to learn how traffic can be estimated or shaped (classifying, marking, metering, dropping, shaping) so that traffic can be handled without congestion and therefore guaranteeing differentiated service flows with it's QoS parameters.
The policing should be accurate to remark the traffic condition violations. Traffic shaping then should do the rest with cell dropping, queuing and degrading the cells to a lower priority. Examples of such traffic condition violations are traffic greater than the Peak Cell Rate (PCR), traffic less than minimum cell rate or early cells.

Another general problem is that reverse traffic (traffic sent back from the receiver) can hardly be estimated in terms of amount and burst.

> Dropping is one rather bad solution, the alternative is an aggressive flow detection. Same problem when a World Wide Web (WWW) application gets traffic back from steadily changing locations. The QoS parameters nevertheless need to be supported on each flow.

Will best-effort users suffer or profit from the implementation of diffserv?

> They will profit with better queuing implementation in the routers that is not only done for the better services but also for Best-Effort Service to protect the other services (RED with in and out (RIO) [22]). However, users will suffer somehow that more resources can be reserved by better services and diminish the space to be used by Best-Effort Service. In unused times, Best-Effort Service can still profit from large bandwidth.
> An overall desire from network suppliers is uttered that applications need to control their load for the network and not to put all the burst on the network whenever they want, otherwise it is not imaginable that all applications will have this functionality implemented leaving the task to handle these bursts to routers.

3.3.3 Mapping of Differentiated Services on ATM Service Categories

Following every combination of the former described six Differentiated Services with the five ATM Service Categories is investigated, whether it’s a good mapping or not. In a separate column the resource wastage of such combinations is shown.

3.3.3.1 *Premium Service on ATM*

	Description:	suited / not suited for:	mapping:	resource wastage:
CBR	The constant bitrate specified for Premium Service and CBR fit well together. A virtual leased line can be formed with this combination (from ISPs to customers, like private persons and companies, or a telephone company leasing the virtual lines to connect company subsidiaries together with Premium Service, while the ATM backbone of the telephone company is set to CBR. The maximum bandwidth parameter on Premium Service is therefore set to PCR. When PCR is specified accordingly even telephone and banking transactions can go over this construction of combined services with a certain rate of unused bandwidth.	**suited:** Audio, Video, TV, Video Conferencing, Weather Satellite Pictures **possible:** Telephone, Fax, Banking Transactions, Airline Reservation, ftp, http, telnet, email, news **impossible:** -	good	low
rt-VBR	Premium Service allows no bursts but rt-VBR does. If there is bursty traffic like telephone calls the rt-VBR can process the data well and the excessive data can be used for other customers while Premium Service has unused capacity when a person listens to his partner on the other side of the line. This combination is thinkable bringing Premium Service to the customer assuring him a virtual leased line. It is supposed that customers still like to have a separate line for traditional reasons, like backwards compatibility and psychological ones (with switching systems it is not anymore a real physical leased line), while the ATM backbone processes the effectively incoming data only, saving capacity on Service Provider side. This mapping leads to a CBR like service with certain resource wastage on ATM side..	**suited:** - **possible:** Audio, Video, TV, Video Conferencing, Telephone, Fax, Banking Transactions, Airline Reservation, Weather Satellite Pictures, ftp, http, telnet, email, news **impossible:** -	good - medium	medium-high
nrt-VBR	nrt-VBR is well suited for bursty response time critical transaction processing, while Premium Service is not supporting bursts on the other side; Premium Service is supporting real-time applications but nrt-VBR does not. To fit somehow together the parameters of the Premium Service would need to be set to very high values but would therefore be very costly also.	**suited:** - **possible:** Fax, Weather Satellite Pictures, ftp, http, telnet, email, news **impossible:** Audio, Video, TV, Video Conferencing, Telephone, Banking Transactions, Airline Reservation	medium	high

	Description:	suited / not suited for:	mapping:	resource wastage:
UBR	Premium Service cannot be mapped on UBR because UBR supports no QoS. In congestion situations UBR is transmitting almost no data, while Premium Service is running idle and processing the incoming data in real-time. On the other hand if UBR has a large bandwidth available, Premium Service is limited to a maximum bitrate. This mapping would be a wastage of high quality Premium Service.	**suited:** - **possible:** ftp, http, telnet, email, news **impossible:** Audio, Video, TV, Video Conferencing, Telephone, Banking Transactions, Airline Reservation, Fax, Weather Satellite Pictures	bad	high
ABR	Like UBR, but ABR can vary its bitrate in congestion situations, so discarded cell rate can be kept low. On one hand ABR is not supporting minimum cell rate and on the other hand Premium Service is not supporting bursts, therefore this mapping would be no good idea either.	**suited:** - **possible:** ftp, http, telnet, email, news, Fax, Weather Satellite Pictures **impossible:** Audio, Video, TV, Video Conferencing, Telephone, Banking Transactions, Airline Reservation	bad	high

Table 5: Premium Service on ATM

3.3.3.2 Assured Service on ATM

	Description:	suited / not suited for:	mapping:	resource wastage:
CBR	Assured Service brings a better data transmitting rate than Best-Effort Service, especially in congestion situations. Bursts are supported by excessive Assured Service but CBR cannot absorb bursts therefore the cells are discarded. If the line is in little use the CBR is running idle using costly service. There is no need to bring high quality services together like CBR and Assured Service that don't fit together. With a big buffer some traffic shaping can be performed with setting the CBR parameter to a high value.	**suited:** - **possible:** Telephone, Banking Transactions, Airline Reservation, ftp, http, telnet, email, news, Fax, Weather Satellite Pictures **impossible:** Audio, Video, TV, Video Conferencing	medium	high
rt-VBR	Assured Service brings no real-time support but rt-VBR does. It makes no sense to support real-time on one part but not on the other. Anyhow in Assured Service the packets are treated with priority so that some better than Best-Effort Service can be brought to ATM even when rt-VBR is the far better service. This mapping might work in most of the times but if Assured Service fails to bring the service expected, there is no use to buy expensive ATM links when the weaker link part makes suffer the whole link.	**suited:** - **possible:** Telephone, Banking Transactions, Airline Reservation, ftp, http, telnet, email, news, Fax, Weather Satellite Pictures **impossible:** Audio, Video, TV, Video Conferencing	good - bad	medium
nrt-VBR	In Assured Service the packets are marked with a higher priority and forwarded over the nrt-VBR link. Bursts are supported by both Assured Service and nrt-VBR. In congestion situations the packets are treated with higher priority on both differentiated services internet and ATM network. The mapping is good because both parts support the same quality aspects suited for response time critical connections. To support high quality on ATM, with Assured Service, the traffic and QoS parameters need to be set to a very high value (internally big buffers are allocated and a big amount of resources is required).	**suited:** Banking Transactions, Airline Reservation, Remote Process Monitoring, Weather Satellite Pictures, ftp with high priority **possible:** fax, email, news, ftp, real-time Applications that can adjust their frame rate **impossible:** Real-time applications with stringent real-time requirements, Audio, Video, TV, Video Conferencing	good	medium - high

	Description:	suited / not suited for:	mapping:	resource wastage:
UBR	Assured Service offers a better service than Best-Effort Service but UBR doesn't. There is no need to feed packets over priority based transport mechanism when the other part of the link makes no promises to the transportation of his part. The link as a whole habits like Best-Effort Service even one part is supplied with a higher quality and possibly better paid also.	**suited:** - **possible:** ftp, http, telnet, email, news **impossible:** Telephone, Banking Transactions, Airline Reservation, Weather Satellite Pictures, Fax, Audio, Video, TV, Video Conferencing	bad - medium	low
ABR	Assured Services offers priority based transport mechanisms, ABR doesn't. With flow control applications can adjust their bitrate so that the Cell Loss Rate (CLR) is kept at a minimum, - ABR is supporting a minimum Cell Rate - while congestion situations are not as disastrous as for normal Best-Effort Service. Anyhow these services do not well fit together. Assured Service stays a better service than ABR even in congested times.	**suited:** - **possible:** ftp, http, telnet, email, news **impossible:** Telephone, Banking Transactions, Airline Reservation, Weather Satellite Pictures, Fax, Audio, Video, TV, Video Conferencing	medium	low

Table 6: Assured Service on ATM

3.3.3.3 USD Service on ATM

	Description:	suited / not suited for:	mapping:	resource wastage:
CBR	These services do not fit together because if there is no congestion, bursts are not supported on CBR, otherwise, if congestion is present, CBR will not be fully utilised. It is hard to imagine that there is a setting of parameters to receive a good service.	**suited:** - **possible:** ftp, http, telnet, email, news, Telephone, Banking Transactions, Airline Reservation, Weather Satellite Pictures, Fax **impossible:** Audio, Video, TV, Video Conferencing	bad	high
rt-VBR	Bursts are supported on rt-VBR but not on USD. In congestion situations the service is limited through USD. It makes no sense supporting real-time transmission on ATM side, while not guaranteeing any real-time support on USD side.	**suited:** - **possible:** ftp, http, telnet, email, news, Telephone, Banking Transactions, Airline Reservation, Weather Satellite Pictures, Fax **impossible:** Audio, Video, TV, Video Conferencing	bad	high
nrt-VBR	USD has a guaranteed minimal bandwidth and nrt-VBR treats cells with priority. This guarantees an effective treating of the traffic. Bursty traffic can make advantage of such links. USD is in any case (besides badly chosen parameters on ATM side) the weaker service. Anyhow, these services fit together quite good, but it is not the optimum of a mapping.	**suited:** ftp, http, telnet, email, news, Banking Transactions, Airline Reservation, Weather Satellite Pictures, Fax **possible:** Telephone **impossible:** Audio, Video, TV, Video Conferencing	medium - good	medium

	Description:	suited / not suited for:	mapping:	resource wastage:
UBR	UBR is Best-Effort Service and suffers from every occurring congestion, but USD has improved attributes like minimal bandwidth and fair share of excessive resources. Anyhow besides an absolute congestion on UBR the service combination is quite a good one.	**suited:** - **possible:** ftp, http, telnet, email, news **impossible:** Banking Transactions, Airline Reservation, Weather Satellite Pictures, Fax, Telephone, Audio, Video, TV, Video Conferencing	medium	low - medium
ABR	ABR and USD have a minimum bandwidth available (if specified on ABR) and both use the full bandwidth in not congested times, which is limited through ABRs Peak Cell Rate (PCR). Though the capacities are limited over the capacity of a subnet with the minimal bandwidth properly set, response time critical applications can be transported.	**suited:** ftp, http, telnet, email, news, Banking Transactions, Airline Reservation, Weather Satellite Pictures, Fax **possible:** Telephone **impossible:** Audio, Video, TV, Video Conferencing	good - medium	low

Table 7: USD Service on ATM

Olympic Service on ATM

	Description:	suited / not suited for:	mapping:	resource wastage:
CBR	CBR has a constant bitrate and supports no bursts, on the other side Olympic Service does not have a minimal bitrate to be specified. Therefore the link combination suffering from congestion on Olympic Service while the CBR link running idle. These two services do not fit together.	**suited:** - **possible:** ftp, http, telnet, email, news **impossible:** Telephone, Banking Transactions, Airline Reservation, Weather Satellite Pictures, Fax, Audio, Video, TV, Video Conferencing	bad	high
rt-VBR	Like CBR, but bursts are supported on rt-VBR. In congestion situations the service is limited through Olympic Service, because Olympic Service has no real-time support neither a minimal bitrate specified. Therefore it's no use paying a high quality service like rt-VBR while Olympic Service supports not an appropriate service.	**suited:** - **possible:** ftp, http, telnet, email, news **impossible:** Telephone, Banking Transactions, Airline Reservation, Weather Satellite Pictures, Fax, Audio, Video, TV, Video Conferencing	bad	high
nrt-VBR	No minimal bandwidth is guaranteed through Olympic Service. nrt-VBR is always the better service, if well configured. This combination could be thought to be implemented in campus areas or as access to ISPs where the ISP traffic or the internetworking traffic travels over ATM. Gold-Service could be mapped on nrt-VBR.	**suited:** 1st class service on campus areas, ISP access of most services **possible:** ftp, http, telnet, email, news, Banking Transactions, Airline Reservation, Weather Satellite Pictures, Fax, Telephone **impossible:** Audio, Video, TV, Video Conferencing	medium - good	medium

	Description:	suited / not suited for:	mapping:	resource wastage:
UBR	Any subnet (Olympic or ATM) is limited through the capacities available in these subnets. In congestion situations the user (especially the Bronze user will be squeezed to almost zero) will suffer from lack of bandwidth, but on lightweight traffic times the user can benefit from the maximum burst rate. Silver-Service could be mapped on UBR.	**suited:** 3rd class service on campus areas or access to ISPs, http and ftp (at night), mirror server updates **possible:** ftp, http, telnet, email, news **impossible:** Banking Transactions, Airline Reservation, Weather Satellite Pictures, Fax, Telephone, Audio, Video, TV, Video Conferencing	medium - good	low - medium
ABR	ABR has a minimal bandwidth parameter to be set, but Olympic Service does not. This combination forms a slightly better service than Best- Effort-Service on both sides. Applications that can adjust their bitrate can benefit from a small packet/Cell Loss Ratio (CLR). Both subnets are limited through the capacity of the other subnet, therefore no critical response time applications can be used over this link. Bronze-Service could be mapped on ABR.	**suited:** 2nd class service on campus areas or access to ISPs, http and ftp (at night), mirror server updates **possible:** ftp, http, telnet, email, news **impossible:** Banking Transactions, Airline Reservation, Weather Satellite Pictures, Fax, Telephone, Audio, Video, TV, Video Conferencing	medium - good	low - medium

Table 8: Olympic Service on ATM

3.3.3.5 SRP Service on ATM

	Description:	suited / not suited for:	mapping:	resource wastage:
CBR	SRP can easily support real-time applications because the estimators can adjust well to the traffic parameters required. CBR supports this functionality. The connection that is established will first be tested on delay. If the delay fits to the real-time data, the packets are transmitted. The bandwidth requested over SRP will be taken to CBR as Peak Cell Rate (PCR). Problems will occur when the datarate breaks down, then the resources will not be supported for the rest of the linked time. This combination is especially interesting for telephone companies that want to support the normal telephone lines and sell the rest of the capacity for computer data transmission. The aggregated telephone lines will scale up and down over time, needing a different service contract (bitrate) on ATM from time to time.	**suited:** Audio, Video, TV, Video Conferencing, Weather Satellite Pictures, Fax, Aggregated Telephone Lines **possible:** ftp, http, telnet, email, news **impossible:** Single - not aggregated - Telephone Lines, Banking Transactions, Airline Reservation, Weather Satellite Pictures, Fax	medium - good	low
rt-VBR	Bursts are badly supported through the slowly reacting estimators of SRP. rt-VBR is the better service and SRP must probably hold the line through a minimal packet rate inserted not to loose the connection. Anyhow, good aggregation of telephone lines is possible. The average capacity of many aggregated telephone lines can easily be estimated.	**suited:** Aggregated flows (Telephone Lines), real-time applications with bitrate not collapsing to zero **possible:** ftp, email, news, Weather Satellite Pictures, Fax, Audio, Video, TV, Video Conferencing **impossible:** Banking Transactions, Airline Reservation, telnet, http, Single - not aggregated - Telephone Lines	medium - good	medium
nrt-VBR	nrt-VBR is made for bursty traffic but SRP can't support bursty traffic because of the slowly reacting estimators. Starvation of links is possible when the link falls under a certain bitrate. Therefore this is no good combination of services.	**suited:** - **possible:** ftp, http, telnet, email, news **impossible:** Banking Transactions, Airline Reservation, Telephone, Weather Satellite Pictures, Fax, Audio, Video, TV, Video Conferencing	bad	high

	Description:	suited / not suited for:	mapping:	resource wastage:
UBR	UBR promises a fair share and can support bursts in not congested times. SRP has a priority transmission, is Best-Effort Service in the case that no reservation can be made and bursts are not supported. With this different attributes these services can't fit together.	**suited:** - **possible:** ftp, http, telnet, email, news **impossible:** Banking Transactions, Airline Reservation, Telephone, Weather Satellite Pictures, Fax, Audio, Video, TV, Video Conferencing	bad	medium
ABR	ABR can select the Minimum Cell Rate (MCR) to equal or greater than zero, it supports bursts and scales well. On the other hand SRP looses the link reservation when the bitrate falls under a certain packet rate, does not support bursts and scales down but not in any case up. Therefore the capacity is dependent on the capacity of each subnet. Anyhow the services ABR and SRP scale up and down well in normal conditions.	**suited:** - **possible:** ftp, http, telnet, email, news **impossible:** Banking Transactions, Airline Reservation, Telephone, Weather Satellite Pictures, Fax, Audio, Video, TV, Video Conferencing	medium	medium

Table 9: SRP Service on ATM

3.3.3.6 Best-Effort Service on ATM

	Description:	suited / not suited for:	mapping:	resource wastage:
CBR	CBR supports no peaks and Best-Effort Service supports no minimal bandwidth. That means, the weakness of each service type makes the service as a whole a bad service. No forwarding in congested times, no bursts supported, no real-time delivery.	**suited:** - **possible:** ftp, http, telnet, email, news **impossible:** Telephone, Banking Transactions, Airline Reservation, Weather Satellite Pictures, Fax, Audio, Video, TV, Video Conferencing	bad	high
rt-VBR	No forwarding is guaranteed over Best-Effort Service. If the Best-Effort Service part can send bursts over maximum speed, the rt-VBR can absorb this load even for real-time service when the parameters are well set, but it makes no sense to have a good but expensive service on one side and a cheap but lousy service on the other side that makes vanishing the advantages.	**suited:** - **possible:** ftp, http, telnet, email, news **impossible:** Telephone, Banking Transactions, Airline Reservation, Weather Satellite Pictures, Fax, Audio, Video, TV, Video Conferencing	bad	high
nrt-VBR	Like rt-VBR but without real-time support on ATM side. This mapping would be no good idea either because nrt-VBR is a high quality service, while Best-Effort Service makes no quality commitment at all.	**suited:** - **possible:** ftp, http, telnet, email, news **impossible:** Telephone, Banking Transactions, Airline Reservation, Weather Satellite Pictures, Fax, Audio, Video, TV, Video Conferencing	bad	high

	Description:	suited / not suited for:	mapping:	resource wastage:
UBR	UBR and Best-Effort Service make no commitment on transmission rate nor on delay. In times with low traffic both can transmit bursts at a high rate but in congestion situations every bit might get stuck in the net. The transport protocol will recognise this loss of data. Therefore sending the packets again and again, thus leading to an even more congested situation. This could be a cheap service connection which is well known today from the Internet extended over ATM.	**suited:** ftp, http, telnet, email, news **possible:** - **impossible:** Telephone, Banking Transactions, Airline Reservation, Weather Satellite Pictures, Fax, Audio, Video, TV, Video Conferencing	medium - good	low
ABR	With flow control it is possible to get a low Cell Loss Ratio (CLR). If all applications can do so, less congestions are occurring than without flow control. This is expected to be an inexpensive service; it can profit from the high bandwidth, but with a slightly better behaviour than a simple Best-Effort Service when facing congestion.	**suited:** ftp, http, telnet, email, news **possible:** as today's known telephone over the Internet **impossible:** Telephone, Banking Transactions, Airline Reservation, Weather Satellite Pictures, Fax, Audio, Video, TV, Video Conferencing	good	low

Table 10: Best-Effort Service on ATM

3.3.4 Recommended Mappings

The results from the pages above are now presented in a table:

	CBR	rt-VBR	nrt-VBR	UBR	ABR
Premium	++ [1]	+ + / -	+ / -	- -	- -
Assured	+ / -	+ / -	+ + [2]	+ / - -	+ / -
USD	- -	- -	+ + / -	+ / -	+ + / -
Olympic	- -	- -	+ + / - [3]	+ + / - [4]	+ + / - [5]
SRP	+ + / -	+ + / -	- -	- -	+ / -
Best-Effort	- -	- -	- -	+ + / - [6]	+ +

Table 11: Mapping of Differentiated Services on ATM Service Categories

Notes:	+ +	(very) good mapping
	+ + / -	medium mapping (has more advantages than drawbacks)
	+ / -	medium mapping (about the same amount advantages and drawbacks)
	+ / - -	medium mapping (has more drawbacks than advantages)
	- -	(very) bad mapping

[1] Virtual Leased Line
[2] Response Time Critical Link
[3] Gold-Service
[4] Bronze-Service
[5] Silver-Service
[6] Best-Effort Service over ATM. No QoS supported

Here the results from the pages above concerning the wastage of resources in the specific combination:

	CBR	rt-VBR	nrt-VBR	UBR	ABR
Premium	- -	+ + / -	+ +	+ +	+ +
Assured	+ +	+ / -	+ + / -	- -	- -
USD	+ +	+ +	+ / -	+ / - -	- -
Olympic	+ +	+ +	+ / -	+ / - -	+ / - -
SRP	- -	+ / -	+ +	+ / -	+ / -
Best-Effort	+ +	+ +	+ +	- -	- -

Table 12: Wastage of resources on service category mappings

Notes:	+ +	high	wastage of resources
	+ + / -	medium to high	wastage of resources
	+ / -	medium	wastage of resources
	+ / - -	low to medium	wastage of resources
	- -	low	wastage of resources

With this information service structures can be built with the aspect of good mappings and the also important aspect of resource consumption, meaning high costs. Anyhow good services are sometimes costly also the customer is probably willing to pay for an good or extraordinary service. These services made possible on the internet will cost maybe less over the internet than over public lines.

The best mappings are described again in an overview:

Premium Service with CBR form a virtual leased line. It's well suited for constant flows like Audio and Video transmission. Even real-time applications with compressed data (like MPEG) but an overall average bitrate can be supported. Over this link, policing and shaping in the edge router with relatively big buffers will flat out the peaks to make the bitrate constant. This will be a commonly used service because it goes well with real-time applications with hard real-time requirements and is a good and save coupling of the two services. Many other services with soft real-time requirements will benefit from such a service combination like ftp, http, telnet, email and news.

Assured Service and nrt-VBR is a service combination that supports bursty non real-time traffic. This services would bring appropriate link behaviour to response time critical applications like Banking Transactions, Airline Reservation Systems, Virtual Reality, Remote Surgery, http and telnet.

USD and nrt-VBR as one service class is well prepared to deliver none real-time data. Applications like Weather Satellite Pictures and Fax can benefit from this combination to get a higher priority treating. This service is discriminated from the real best-effort traffic.

Olympic Service does not connect to ATM services optimally. It can be imagined that this service is established on campus areas to get a service discrimination while simultaneously get access to the Internet Service Provider (ISP). It is supposed that most of the traffic stays within the campus area or within an access provider's area.
A scenario like most of the traffic within a campus area travels over Olympic Service and the internetworking traffic maps to and travels over ATM service can be imagined. ISPs can then offer very special services - like full bandwidth to surfers for low price at night and so on - that are specially suited to customers needs.

SRP on CBR provides a constant bitrate, but here the flow rate can scale up and down to accomplish the requirements. The allocation of more bandwidth is not guaranteed and if the bandwidth falls under a certain level the link could be disconnected. Nevertheless this is a quite good model for aggregating flows to a single flow. With this aggregation the starvation of links is avoided. When more links go online the aggregated link could ask for more capacity and therefore make a new service contract. The link can then scale up and down well. This combination suits for telephone companies that want to sell the excessive capacity for computer data best-effort lines.

Best-Effort goes **with** both **UBR and ABR** together but ABR is always preferable as long the applications - or the underlying link levels - support flow control to avoid congestion. Ftp, http, telnet, email and news are well suited applications with this service combination.

From the application side no support for bursty real-time applications (rt-VBR on ATM side) like remote control of machines (remote surgery and virtual reality simulations) is guaranteed from any of the Differentiated Service models or combinations of Differentiated Service and ATM. With the emerge of virtual reality applications and their commercial use, these type of applications will demand for network support. There has to be further study to support these type of applications with Differentiated Services.

3.3.5 Deployment issues

In the following table some deployment issues are listed:

	Where must be additional services deployed?	Network information delivery	Granularity	Monetary cost
Premium	No changes to endsystems required when the border routers have additional forwarding primitives like classifiers, markers, policers and shapers as well as token buckets	traffic allocation through bandwidth broker	per-flow or aggregated flow	medium
Assured	No changes to endsystems required when the border routers have additional forwarding primitives like classifiers, markers, policers and shapers as well as token buckets	traffic allocation through bandwidth broker	per-flow or aggregated flow	medium
USD	No changes to endsystems, no admission control, deployment in bottlenecks only. Each router needs to know the share ratio and has a proportional bandwidth sharing scheduler implemented	not mentioned	per-user traffic isolation can be aggregated to any aggregation level	medium - low
Olympic	Rate based link share scheduler in every hop, classification at the edge of a network, no admission control, no shaping, no policing	traffic allocation through bandwidth broker	per-flow or aggregated flow	medium
SRP	Active role of end systems needs additional software. Routers survey the datastreams through estimators that try to estimate the future flows	reservation protocol and feedback protocol	aggregated in every node	high
Best-Effort	none	none	per-flow or aggregated flow	none

Table 13: Deployment issues

3.3.6 Socket

A socket needs to make an abstraction of the underlying layers while supporting all necessary traffic parameters. The socket hides the implementation specific and network topology specific details from the application. The application can then choose from a variety of links (e.g., a video link with DVD quality, a telephone line or a line for e-mail download). These should be selected by the application bringing appropriate service to the user. Anyhow the user can select some of these abstract links when the user needs a special link (for playing a video or to store this video with DVD quality need different QoS parameters for the line). Only a few possibility of the wide range of the possible parameters / links are given to the user.

At the socket level all traffic and QoS parameters are available for a wide variety of links. Some selected parameters are propagated to the next higher layer which is the application layer. Here the application can select a smaller amount of parameters which preferably are clustered to links, so that the application can select links instead of parameters. In most of the cases the user selects an application he wants to run (e.g., the user wants to hear news from a certain radio station); the application selects the appropriate link and gives these information to the socket which chooses the appropriate parameters to meet these requirements.
In certain cases it is necessary that the user itself chooses a link with a certain quality. The user will therefore have to answer simple questions only like: ‘Do you want a line with CD quality or with telephone quality?’ So the user does not need to know about the underlying QoS parameters.

When, lets say a internet surfer, clicks on a real-audio button, a new line is required and established to support these requirements for that audio transmission, meaning having opened several links with different QoS parameters simultaneously.

Some of the QoS parameters may be:

- Traffic parameters (like bandwidth and parameters which have influence on delay and dependability). These parameters make a traffic commitment based on bandwidth and are responsible for the stringent delay of the packets.

- Other QoS parameters (like packet loss and packet error). These parameters are responsible to check the accuracy of the QoS flow, whether the flow commits to the QoS parameters set. The lost and errored packets as well as packets that travel to a wrong destination (misinserted packets) are compared to the traffic commitment. Counter measurements are taken when these measured parameters are not accurate.

- Short-cut routing / hop-by-hop-routing. The application can choose whether it sends a short time duration flow by selecting hop-by-hop-routing or a long time duration flow by selecting the short-cut routing option.

It can be imagined that these parameters are standardised for every application, that the application can tell the socket which type of line to be established (e.g., a line with CD quality, 44.1 kHz, 16 bit, Stereo has its exact standardised parameters), or the user can select the required line.

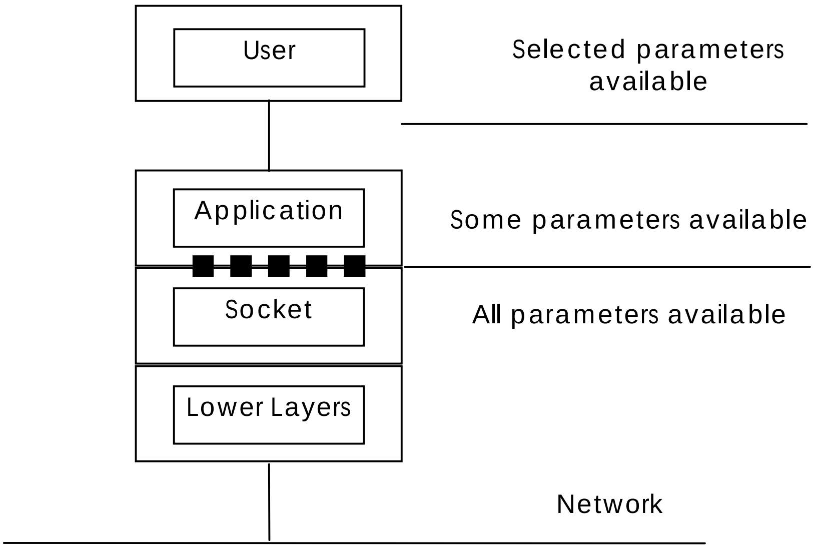


Figure 22: Socket

Anyhow, it is not imaginable that every application will support all of the features. It is more likely that the traffic estimation is done in routers which make the reservations needed. The user will - in most of the cases - not select any of the parameters; it is furthermore desirable do make an automation as far as possible.

According to the mapping from application to service categories the link is established with the appropriate QoS parameters. While the application transmits its data, the lower layers (e.g., TCP) are responsible for the connection, the policed and shaped transmission of the data and so on.

3.3.7 Mapping of applications to service categories

Sockets need to map the application type to a service type. Each application belongs to a type that is communicated to the socket when link reservations are made. Then the socket (and the underlying layers) is responsible for establishing the link with the appropriate Quality of Service. The following table gives an example how to support most of the services.

Service Type	Application Type	Application
Premium Service / CBR	Real-time applications with hard real-time requirements an with overall constant or maximum bitrate	Audio, Video on Demand, TV, Radio, Video Conferencing, Weather Satellite Pictures, ftp, http, telnet, email
Assured Service / nrt-VBR	Non real-time applications with a critical response time	Airline Reservation Systems, Banking Transactions, http, telnet
Best-Effort Service / ABR / UBR	Non real-time applications that can or cannot adjust the bitrate with feedback control. Traditional Best-Effort Service	ftp, http, telnet, email
Not supported services	Bursty real-time applications with soft real-time requirements	Telephone, Fax, Virtual Reality, Remote control of machines (remote surgery)

Table 14: Mapping of applications to service categories

4. Example

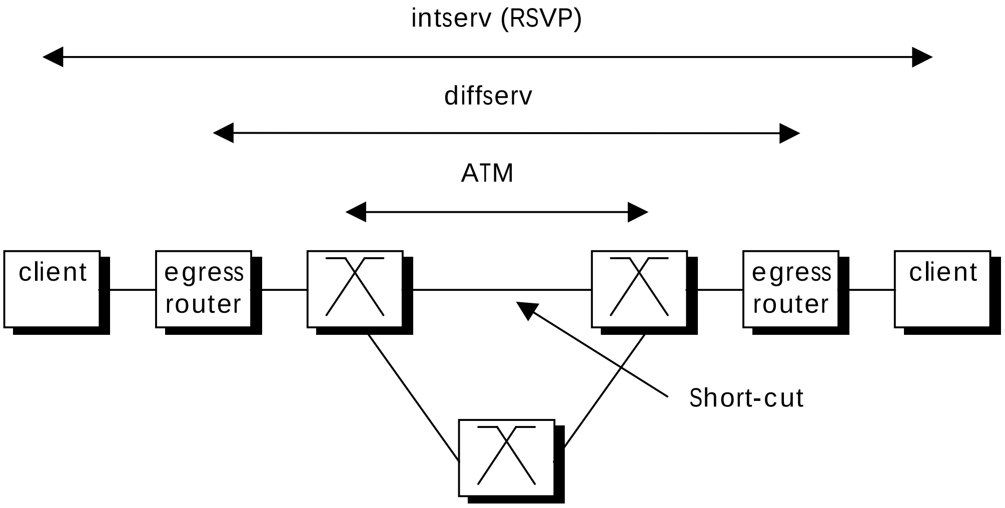


Figure 23: Quality of Service Network configuration

An IP packet is sent from a client in an integrated services (intserv) area with a packet classified to a certain QoS class (other than best-effort) and the Resource reSerVation Protocol (RSVP) tries to make a reservation (from end-to-end, or to the egress router of the diffserv area) for the desired traffic.
If the resources are granted and reserved, the packets arrive at the diffserv border area. There, the integrated service class needs to be mapped into differentiated service to go over the differentiated services area of the net. The reservation for this ATM cloud has to be made and the packets that are now chopped into cells are sent over the ATM network. However, here the ATM improvements take place.
First, the Next Hop Resolution Protocol (NHRP) comes into play. It sends a message to an ATMARP server (ATM Address Resolution Protocol) that gives back the ATM address of the server towards the destination. The border router knows the address to which to send the cells. This is the shortest way (or the shortest way that has the recourses to fulfil the QoS requirements over the ATM net) towards the destination. The ATMARP server has learned the information over time with ATMARP investigation packets or has been taught by a system administrator.

Before now the border router can send the packets that are meanwhile arriving, stored in queues and sent in smaller cells over the ATM cloud, it checks whether it has already a flow or an aggregated flow (known as Virtual Channel VC or Virtual Path VP) travelling to the same destination. If this is the case, the router classifies the cells with the same number of VC or VP, so it has not anymore to look-up it's routing table. This information will help the following ATM switches to redirect the cells in the direction towards the destination router.

Now the first cell is sent to the next ATM switch. This ATM switch and the following ATM switches toward the destination router proceed in the same way. To make switching fast - ATM switches that can easily handle hundred of thousands of flows - new switching technique are implemented (IP Switching needs hardware support, while Tag Switching uses software only). The problem here is that every ATM switch on the way of a cell needs to make a routing decision for this cell. When the first cell of an IP flow arrives at an IP switch the switch looks-up his routing table whether it has information on that flow yet. This information is given to the ATM switch hardware that is capable to switch the following cells just by looking at the Virtual Channel / Virtual Path (VC/VP) block in the cell header.

At every router and switch the traffic agreements are verified. The packets are classified in the ToS field according to the class they belong to, marked with this class, policed (are the packets within the agreed policy) and finally shaped (packets are delayed, dropped or marked out-of-profile).

5. Conclusion

The internet is growing rapidly. To supply faster and better service, big efforts are made. Services like ftp, e-mail and surfing on the internet are OK to run on the current Internet architecture, but better services, like response time critical applications (airline reservation systems) and audio and video transmissions (video conferencing) need higher quality and service requirements. These requirements are different from application to application. One may need real-time support while the other needs an answer in a certain response time. These requirements are commonly known as Quality of Service, they differ from Best-Effort Service which is available on the internet today. Best-Effort Service means that packets are sent as fast as possible and as reliable as possible.

In some internet areas integrated services are supported with RSVP that reserves resources for the support of QoS. Integrated services work well over small areas but are not scaleable to wide area networks like the world wide internet. Differentiated services try to widen these areas where Quality of Service is required.

To support QoS, the Internet and its components need to be adjusted to the new proposals. Due to the different requirements a series of building blocks are constructed to support most of the services.

Some of the IP packets travel over ATM clouds. ATM need to be supported, especially because ATM networks support Quality of Service building blocks inherently.
The following proposals are discussed in this paper: Premium Service [15], Assured Service [16], User-Share Differentiation (USD) [19], Olympic Service [21] and Scalable Resource Reservation (SRP) [23].

The best mappings are presented in a small table:

Differentiated Services	ATM
Premium Service	CBR
Assured Service	nrt-VBR
USD Service	rt-VBR / ABR
Olympic Service	No extraordinary mapping
SRP Service	CBR
Best- Effort Service	UBR / ABR

Table 15: The best mappings between ATM and Differentiated Services

Notes: Bursty response time critical applications - like Banking Transactions, Remote Surgery - are not explicitly supported over a specific combination.

Olympic has no extraordinary mapping, but goes most preferably with nrt-VBR UBR and ABR.

These combinations can have certain drawbacks in wasting capacity of the networks resources.

It is a desire to make routers or switches faster because in internet hundreds of thousands and also millions of routes need to be routed within seconds. Quality can better - or in certain cases only - be supported, when greater speed is available.
Many attempts to increase speed are made including the ones that are discussed in this paper which are improvements within ATM networks:

- IP Switching to switch cells at hardware speed
- Tag Switching
- NHRP (Next Hop Resolution Protocol) for establishing the shortest routes possible
- Flow aggregation to cluster flows into one so that routers don't need to distinguish where the single flow has to go, the routing tables therefore getting shorter and the speed of routing greater

6. Glossary

6.1 General Glossary

ACK	Acknowledge packets from ATMARP servers.
AS	Autonomous Systems.
BA	Behaviour Aggregate classifiers which classify only on patterns in the DS Byte.
BAC	Behaviour Aggregate Classifier. A classifier that selects packets based only on the contents of the DS field. Such classifiers are used in DS interior nodes, and are typically used for policing at a DS ingress node.
Boundary	A link connecting the edge nodes of two domains.
CAC	Connection Admission Control is defined as the set of actions taken by the network during the call set-up phase in order to determine whether a connection request can be accepted or should be rejected (or whether a request for re-allocation can be accommodated).
CBQ	Class-Based Queuing.
Classical IP	Classical IP and ARP over ATM, see RFC 1577.
Classifier	A logical element of traffic conditioning that selects packets based on the content of packet headers according to defined rules.
CLP	Cell Loss Priority control: For some service categories the end system may generate traffic flows of cells with Cell Loss Priority (CLP) marking. The network may follow models which treat this marking as transparent or as significant. If treated as significant, the network may selectively discard cells marked with a low priority to protect, as far as possible, the QoS objectives of cells with high priority.
CLS	Controlled Load Service.
Codepoint	A specific value of the PHB field in the DS Byte.
COPS	Common Open Policy Service Protocol.
CoS	(Differentiated) Classes of Service.
Customer DS domain	A DS domain that has an SLA in place with another directly attached DS domain (the provider DS domain) governing the rules by which traffic from the customer DS domain will be serviced within the provider DS domain.
DACS	Differentiated Admission Control Service.
diffserv	Differentiated Services. The user commits a service profile with the ISP and the packets have a priority marking. The flows can be aggregated (all flows i.e. between subnets). It is scaleable for small and large networks. IETF working group with same name.

DS Byte	A small bit-pattern in each packet, in the IPv4 ToS octet or the IPv6 traffic class octet, is used to mark a packet to receive a particular forwarding treatment, or per-hop behaviour, at each network node.
DS behaviour aggregate	A stream of packets that have the same DS codepoint.
DS capable	Able to support Differentiated Services functions and behaviours.
DS codepoint	A specific bit-pattern of the DS field.
DS edge node	A DS node that connects one DS domain to a node either in another DS domain or in a domain that is not DS capable.
DS egress node	A DS edge node in its role in handling traffic as it leaves a DS domain.
DS destination host	A DS host that acts as a DS egress node.
DS domain	A contiguous set of nodes which operate with a common set of service provisioning policies and PHB definition; it consists of DS interior nodes and DS edge nodes.
DS edge node	A DS node that is not a DS interior node.
DS egress node	A DS edge node in its role in handling traffic as it exits a DS domain.
DS host	A host computer that can perform certain traffic conditioning functions and therefore acts as a special DS edge node.
DS ingress node	A DS edge node in its role in handling traffic as it enters a DS domain.
DS interior node	A DS node that is not a DS edge node.
DS node	A DS capable node.
DS source host	A DS host that acts as a DS ingress node.
dWFQ	Distributed Weighted Fair Queuing.
dWRED	Distributed Weighted Random Early Detection.
ERED	Enhanced Random Early Detection.
FAC	Flow Admission Control.
FIB	Forwarding Information Base.
GPRA	Generic Packet Rate Algorithm.
GCRA	Generic Cell Rate Algorithm.
GS	Guaranteed Service.
IETF	Internet Engineering Task Force.
intserv	Integrated Services. Suited for small networks but hardly scaleable. It is based on reserved resources witch are set-up with the RSVP protocol. IETF working group with same name.
IP	Internet Protocol.
issll	Integrated Services over Specific Link Layers. IETF working group with same name.
ITU	International Telecommunication Union.
LAN	Local Area Network.
LANE	LAN Emulation.
LIS	Logical Independent Subnets.
MAN	Metropolitan Area Network.

Marker
A logical element of traffic conditioning that sets the DS codepoint in the DS field based on defined rules.

MARS
Multicast Address Resolution Server.

MCS
MultiCast Servers.

Meter
A logical element of traffic conditioning that measures the properties (e.g., rate) of a packet stream selected by a classifier.

MFC
Multi-Field Classifiers which can classify on the DS Byte as well on any of a number of header fields.

MIB
Management Information Base.

MPLS
MultiProtocol Label Switching. Base technology that is expected to improve the price/performance of network layer routing, the scalability of the network layer and provide greater flexibility in the delivery of (new) routing services (by allowing new routing services to be added without a change to the forwarding paradigm).

MPOA
MultiProtocol Over ATM. IETF working group with same name.

NAK
Not-acknowledge packets from ATMARP servers.

NBMA
Non-Broadcast Multiple Access networks, such as ATM, SMDS and X.25.

NHC
Next Hop Resolution Client.

NHRP
Next Hop Resolution Protocol is defined for avoiding extra hops in the delivery of IP packets over Non-Broadcast Multiple Access (NBMA) networks with a single destination.

NHS
Next Hop Resolution Server.

PHB
A Per-Hop Behaviour (PHB) is a description of the forwarding behaviour of a DS node applied to a particular DS behaviour aggregate.

PHB group
A set of one or more PHBs that can only be specified and implemented simultaneously, due to a common constraint applying to all PHBs in the set such as a packet scheduling or discard policy.

PNNI
Private Network-Network Interface Specification supports QoS. PNNI includes two categories of protocols:

- A protocol is defined for distributing topology information between switches and clusters of switches. This information is used to compute paths through the network
- A second protocol is defined for signalling. Message flows are used to establish point-to-point and point-to-multipoint connections across the ATM network

Policing
The process of applying traffic conditioning functions such as marking or discarding to a traffic stream in accordance with the state of a

	corresponding meter. The policing action taken may be one of two possibilities only: 1. drop the over-rate packet and 2. hold the over-rate packet until it will be in compliance with the peak rate (shaping).
Provider DS domain	A DS domain that has an SLA in place with another directly attached DS domain (the customer DS domain) governing the rules by which traffic from the customer DS domain will be serviced within the provider DS domain. A single DS domain may be both a customer DS domain and a provider DS domain for different directions of traffic at the same time.
RED	Random Early Detection.
RIO	RED with In nd Out.
RFC	Request for Comments.
RSVP	Resource reSerVation Protocol [27,28,29].
SBM	Subnet Bandwidth Manager. A Proposal for Admission Control over Ethernet.
SCSP	Server Cache Synchronisation Protocol.
Service Provisioning Policy	A policy which defines how traffic conditioners are configured on DS edge nodes and how traffic streams are mapped to DS behaviour aggregates to achieve a range of service behaviours.
Shaper	A logical element of traffic conditioning that delays packets within a traffic stream to cause it to conform to some defined traffic properties.
SIMA	Simple Integrated Media Access.
SLA	Service Level Agreement. A service contract between a customer and a service provider that specifies the details of a TCA and the corresponding service behaviour a customer should receive. A customer may be a user, organisation or another DS domain.
SPF	Shortest Path First.
SVC	Switched Virtual Circuit.
TCA	Traffic Conditioning Agreement. An agreement specifying classifier rules and the corresponding traffic profiles and metering, marking, policing and/or shaping rules which are to apply to the traffic streams selected by the classifier.
TIB	Tag Information Base.
Traffic conditioner	An entity which performs traffic conditioning functions and which may contain header classifiers, meters, policers, shapers and markers. Traffic conditioners are typically deployed in boundary nodes only.
Traffic conditioning	Control functions that can be applied to a behaviour aggregate, application flow, or other operationally useful subset of traffic, e.g., routing updates. Traffic conditioning is used to enforce service level agreements between domains and

	to condition traffic to receive a differentiated service within a domain.
Traffic profile	A description of the expected properties of a traffic stream such as rate and burst size.
Traffic stream	An administratively significant set of one or more microflows which traverse a pathsegment. A traffic stream may consist of the set of active microflows which are selected by a particular classifier.
VPN	Virtual Private Network.
WAN	Wide Area Network.
WFQ	Weighted Fair Queuing.
WWW	World Wide Web.

6.2 ATM Glossary

AAL	ATM Adaptation Layer
ABR	Available Bit Rate
ATM	Asynchronous Transfer Mode
BE	Best-Effort
BT	Burst Tolerance
CBR	Constant Bit Rate
CDV	Cell Delay Variation
CDVT	Cell Delay Variation Tolerance
CLP	Cell Loss Priority (bit)
CLR	Cell Loss Ratio
CTD	Cell Transfer Delay
GCRA	Generic Cell Rate Algorithm
LLC	Logical Link Control
MBS	Maximum Burst Size
MCR	Minimum Cell Rate
MPL	Minimum Path Latency
MTU	Maximum Transfer Unit
nrt-VBR	Non-real-time VBR
PCR	Peak Cell Rate
PVC	Permanent Virtual Connection
QoS	Quality of Service
Rspec	Reservation Specification
rt-VBR	Real-time VBR
SCR	Sustainable Cell Rate
TCP	Transport Control Protocol
TM	Traffic Management
TSpec	Traffic Specification
UBR	Unspecified Bit Rate
UNI	User-Network Interface
UPC	Usage Parameter Control (ATM traffic policing function)
VBR	Variable Bit Rate
VC	(ATM) Virtual Connection
VCC	Virtual Channel Connection
VCI	Virtual Channel Identifier
VPC	Virtual Path Connection
VPI	Virtual Path Identifier

7. References

7.1 Important References

[1] af-tm-0056.000 ATM Forum, Traffic Management 4.0

[2] Computer Networks Andrew S. Tanenbaum, Prentice Hall, Third Edition, 1996. ISBN 0-13-394248-1

[3] newman0001.pdf IP Switching: ATM Under IP. Peter Newman, Greg Minshall, and Tom Lyon Ipsilon Networks

[4] draft-davie-tag-switching-atm-00.txt -05.txt
Use of Tag Switching with ATM

[5] draft-turner-diff-nhrp-00.txt
Differentiated Services over Symmetric NHRP Shortcuts

[6] draft-ietf-rolc-nhrp-00.txt -15.txt
NBMA Next Hop Resolution Protocol (NHRP)

[7] draft-ietf-ion-nhrp-appl-00.txt -02.txt
NHRP Protocol Applicability Statement

[8] RFC 2332 NBMA Next Hop Resolution Protocol (NHRP)

[9] draft-ietf-diffserv-arch-00.txt -01.txt
An Architecture for Differentiated Services

[10] Tag Switching http://www.cisco.com/warp/public/732/tag

[11] Cisco http://www.cisco.com

[12] Torsten Braun http://www.iam.unibe.ch/~braun/talks/mannheim/

[13] diffserv Differentiated Services Working Group
http://www.ietf.org

[14] RFC 1349 Type of Service in the Internet Protocol Suite (ToS)

[15] draft-nichols-diff-svc-arch-00.txt
A Two-bit Differentiated Services Architecture for the Internet

[16] draft-ietf-diffserv-precedence-00.txt
IP Precedence in Differentiated Services Using the Assured Service

[17] RFC 2211 Specification of the Controlled-Load Network Element Service

[18] draft-ietf-issll-atm-mapping-07.txt
Interoperation of Controlled-Load Service and Guaranteed Service with ATM

[19] draft-wang-diff-serv-usd-00.txt
User-Share Differentiation (USD). Scaleable bandwidth allocation for differentiated services

[20] CBQ http://ftp.ee.lbl.gov/floyd/cbq.html

[21] draft-nichols-dsopdef-00.txt
Differentiated Services Operational Model and Definitions

[22] draft-clark-diff-svc-alloc-00.txt deleted
An Approach to Service Allocation in the Internet

[23] draft-almesberger-srp-00.txt
Scaleable Resource Reservation for the Internet

[24] EPFL Lausanne A Survey of Differentiated Services Proposals for the Internet Constant Gbaguidi, Hans J. Einsiedler, Paul Hurley, Werner Almesberger, Jean-Pierre Hubaux sscwww.epfl.ch

[25] pmg.ps http://www.eecs.umich.edu/~wuchang/work/ Adaptive Packet Marking for Providing Differentiated Services in the Internet W. Feng, D. Kandlur, D. Saha, K. Shin, U. Michigan, October 1997.

[26] RED http://www-nrg.ee.lbl.gov/floyd/red.html Random Early Detection

[27] RFC 2209 Resource reSerVation Protocol (RSVP). Message Processing Rules

[28] RFC 2208 Resource reSerVation Protocol (RSVP). Applicability Statement Some Guidelines on Deployment

[29] RFC 2205 Resource reSerVation Protocol (RSVP). Functional Specification

[30] RFC 792 Internet Control Message Protocol (ICMP)

[31] af-lane-0021.000 LAN Emulation LAN Emulation over ATM 1.0

[32] af-lane-0038.000 LAN Emulation Client Management

[33] af-lane-0050.000 LANE 1.0 Addendum

[34] af-lane-0057.000 LANE Servers Management Spec 1.0

[35] RFC 1577 Classical IP and ARP over ATM

[36] dWRED http://www.cisco.com/univercd/cc/td/doc/product/software/ios111/cc111/wred.htm

[37] RFC2309 Recommendations on Queue Management and Congestion Avoidance in the Internet

[38] RED with ATM ftp://ftp.rennes.enst-bretagne.fr/pub/reseau/afifi/red-atm.ps

[39] RFC 2381 Interoperation of CLS and GS with ATM

7.2 Other References

Papers / Books:

IAM Bern Differentiated Services: Ein neuer Ansatz für Quality of Service im Internet. Braun / Habegger

Internet Drafts grouped by working groups (IETF) www.ietf.org

A mirror of internet drafts can be found under
ftp://ftp.gwd.de/pub/misc/standards/ftp.leo.org/internet-drafts/

diffserv Differentiated Services Working Group

draft-ietf-diffserv-framework-00.txt
A Framework for Differentiated Services

draft-ietf-diffserv-header-00.txt -01.txt
Definition of the Differentiated Services Field (DS Byte) in the IPv4 and IPv6 Headers

draft-ietf-diffserv-rsvp-00.txt
A Framework for Use of RSVP with Diff-serv Networks

draft-bernet-intdiff-00.txt
A Framework for End-to-End QoS Combining RSVP/Intserv and Differentiated Services

draft-ellesson-sla-schema-00.txt
Schema for Service Level Administration of Differentiated Services and Integrated Services in Networks

draft-ford-issll-diff-svc-00.txt
Integrated Services Over Differentiated Services

draft-blake-diffserv-marking-00.txt
Some Issues and Applications of Packet Marking for Differentiated Services

draft-guerin-aggreg-rsvp-00.txt
Aggregating RSVP-based QoS Requests

draft-ford-issll-diff-svc-00.txt
Integrated Services Over Differentiated Services

draft-ferguson-delay-drop-00.txt
Simple Differential Services: IP ToS and Precedence, Delay Indication, and Drop Preference

draft-worster-diffserv-gr-00.txt
Guaranteed Rate in Differentiated Services

draft-bernet-diffedge-00.txt
Requirements of Diff-serv Boundary Routers

draft-ietf-diffserv-phb-mgmt-00.txt
Management of PHBs

draft-ietf-diffserv-phb-ef-00.txt
An Expedited Forwarding PHB

issll *Integrated Services over Specific Link Layer Working Group*

draft-ietf-issll-is802-framework-05.txt
A Framework for Providing Integrated Services

draft-ietf-issll-is802-svc-mapping-01.txt
Integrated Service Mappings on IEEE 802 Networks

draft-ietf-issll-is802-sbm-06.txt
SBM (Subnet Bandwidth Manager): A Protocol for RSVP-based Admission Control over IEEE 802-style networks

draft-ietf-issll-diff-svc-00.txt
Integrated Services Over Differentiated Services

draft-ietf-issll-sbm-05
SBM Subnet Bandwidth Manager. A Proposal for Admission Control over Ethernet

intserv *Integrated Services Working Group*

draft-ietf-intserv-v2-mib-00.txt
Integrated Services Management Information Base

qosr *Quality of Service Routing Working Group*

draft-ietf-qosr-framework-06.txt
A Framework for QoS-based Routing in the Internet

mpls *Multiprotocol Label Switching Working Group*

draft-ietf-mpls-framework-02.txt
A Framework for Multiprotocol Label Switching

draft-davie-mpls-atm-00.txt
Use of Label Switching with ATM

rap *RSVP Admission Policy Working Group*

draft-ietf-rap-framework-00.txt
A Framework for Policy-based Admission Control

draft-ietf-rap-rsvp-ext-00.txt
RSVP Extensions for Policy Control

draft-ietf-rap-user-identity-00.txt
User Identity Representation for RSVP

draft-ietf-rap-rsvp-ext-00.txt
RSVP Extensions for Policy Control

idr *Interdomain Routing Working Group*

draft-ietf-idr-bgp4-08.txt
A Border Gateway Protocol 4 (BGP-4)

draft-ietf-idr-route-damp-03
BGP Route Flap Damping

draft-ietf-idr-bgp4-ipv6-01.txt
Use of BGP-4 Multiprotocol Extensions for IPv6 Inter-Domain Routing

draft-ietf-idr-bgp4-cap-neg-02.txt
Capabilities Negotiation with BGP-4

draft-ietf-idr-bgp4-mib-02.txt
Definitions of Managed Objects for the Fourth Version of Border Gateway Protocol (BGP-4)

draft-ietf-idr-bgp4-multiprotocol-v2-01.txt
Multiprotocol Extensions for BGP-4

draft-ietf-idr-aggregation-tutorial-01.txt
Route Aggregation Tutorial

draft-ietf-idr-aggregation-framework-03.txt
A Framework for Inter-Domain Route Aggregation

Other drafts

srri.ps
Scaleable Resource Reservation for the Internet

draft-ietf-ion-nhrp-mib-00.txt -04.txt
Definitions of Managed Objects for NBMA Next Hop Resolution Protocol (NHRP)

draft-carlson-nhrp-00.txt
Guidelines for Next Hop Client (NHC) developers

draft-halpern-ion-r2r-nhrp-00.txt -01.txt
NHRP for Destinations off the NBMA Subnetwork

draft-heinanen-diff-tos-octet-01.txt
Use of the IPv4 ToS Octet to Support Differential Services

draft-heinanen-nhrp-00.txt
NBMA Next Hop Resolution Protocol (NHRP) renamed to draft-ietf-rolc-nhrp-00.txt

draft-horikawa-mobile-nhrp-00.txt
Support for Mobile NHRP Devices in ATM Networks

draft-ietf-ion-nhrp-mobile-nhc-00.txt
NHRP with Mobile NHCs

draft-ietf-ion-discov-nhrp-00.txt -01.txt
ILMI-Based Server Discovery for NHRP

draft-ietf-ion-scsp-nhrp-00.txt -05.txt
A Distributed NHRP Service Using SCSP

draft-shivkuma-ecn-diffserv-01.txt
A One-bit Feedback Enhanced Differentiated Services Architecture

draft-doolan-tdp-spec-00.txt
Tag Distribution Protocol TDP

ATM Forum

www.atmforum.com/atmforum/specs/approved.html
ftp.atmforum.com/pub/approved-specs/

af-mpoa-0087.000	LAN Emulation/MPOA Multi-Protocol Over ATM Specification v1.0
af-pnni-0055.000	ATM Forum Private Network-Network Interface Specification, Version 1.0

Introduction Workshops on the Internet

Monish Rajpal	Department of Computer Science, Johns Hopkins University http://www.cnds.jhu.edu/courses/cs667/qos/index.htm
ATM Routing	http://www.crihan.fr/PEPSY/1997/coursATM20/3.0/3.0-ATM_Routing_-_ATMF_PNN.html
IP Switching	3Com Slide Show http://iworks.ecn.uiowa.edu/conference/IWorks97/sessions/sn180/paper.html
Cisco Tag Switching	http://www.cisco.com/warp/public/732/gallery/ra_tagswitch/

RFCs Lists of all RFCs under http://www.garlic.com/~lynn/rfcietf.html

RFC 2386	A Framework for QoS-based Routing in the Internet
RFC 2366	Definitions of Managed Objects for Multicast over UNI 3.0/3.1 based ATM Networks
RFC 2353	APPN/HPR in IP Networks. APPN Implementers' Workshop Closed Pages Document
RFC 2336	Classical IP to NHRP Transition
RFC 2333	NHRP Protocol Applicability Statement
RFC 2331	ATM Signalling Support for IP over ATM
RFC 2216	Network Element Service Specification Template
RFC2215	General Characterisation Parameters for Integrated Service Network Elements
RFC 2214	Integrated Services Management Information Base Guaranteed Service Extensions using SMIv2

RFC 2213	Integrated Services Management Information Base using SMIv2
RFC 2212	Specification of Guaranteed Quality of Service
RFC 2210	The Use of RSVP with IETF Integrated Services
RFC 1987	Ipsilon's General Switch Management Protocol Specification
RFC 1954	Transmission of Flow Labelled IPv4 on ATM Data Links Ipsilon
RFC 1953	Ipsilon Flow Management Protocol Specification for IPv4
RFC 1932	IP over ATM: A Framework Document
RFC 1735	Address resolution (NARP)

Mailing Lists

Diffserv archive	http://www-nrg.ee.lbl.gov/diff-serv-arch

WWW Links / Homepages

IAM Bern	http://www.iam.unibe.ch/~rvs/
IETF	http://www.ietf.org
Jon Crowcroft	http://www.cs.ucl.ac.uk/staff/J.Crowcroft/
Henning Schulzrinne	http://www.cs.columbia.edu/~hgs/
Wu Chang Feng	http://www.eecs.umich.edu/~wuchang/
IBM	http://www.networking.ibm.com