

# IP SWITCHING

Shishir Agrawal, [agrawal@cis.ohio-state.edu](mailto:agrawal@cis.ohio-state.edu)

---

The rapid growth of Internet and increase in real-time and multimedia applications have created a need to improve Internet routing technology in terms of bandwidth, performance, scalability and delivery of new functionalities. Several proposals involving application of layer 2 switching technology to layer 3 routing have been made to counter above challenges. Noticeable efforts in this directions are Ipsilon's IP Switching, Cisco's Tag Switching, IBM's ARIS, Toshiba's CSR and MPLS.

---

[Other Reports on Recent Advances in Networking](#)

[Back to Raj Jain's Home Page](#)

---

## Table of Contents

- [Introduction](#)
- [IP Over ATM](#)
  - [LAN Emulation \( LANE \)](#)
  - [Classical IP Over ATM \( IPOA \)](#)
  - [Next Hop Resolution Protocol \( NHRP \)](#)
  - [Multi-Protocol Over ATM \( MPOA \)](#)
- [IP Switching](#)
  - [IP Switch](#)
  - [Flow Classification](#)
  - [Flow Management Protocol \( IFMP \)](#)
  - [IP Switching Features](#)
  - [Simulation Results](#)
- [Tag Switching](#)
  - [Forwarding Component](#)
    - [Tags](#)
    - [Tag Information Base](#)
    - [Forwarding Procedure](#)
  - [Control Component](#)
    - [Destination Based Routing](#)
    - [Hierarchy of Routing Knowledge](#)

- [Multicast Forwarding](#)
  - [Flexible Routing](#)
  - [Quality of Service](#)
  - [Cell Switched Router](#)
    - [Architecture of CSR based Internetwork](#)
    - [Features of CSR](#)
    - [Flow Classification](#)
  - [Aggregate Route Based IP Switching \(ARIS\)](#)
    - [ARIS mechanism](#)
    - [ARIS Messages](#)
    - [Egress Identifiers](#)
    - [ISR Information Base](#)
    - [ARIS Features](#)
  - [Multi-Protocol Label Switching \( MPLS \)](#)
    - [Definitions](#)
    - [Core MPLS Components](#)
      - [Basic Routing Approach](#)
      - [Labels](#)
      - [Encapsulation](#)
    - [Observation, Issues and Assumptions](#)
    - [Technical Approaches](#)
  - [Summary](#)
  - [References](#)
- 

## Introduction

The last couple of years have seen several major changes in the internetwork traffic pattern. First, there has been an explosive growth of Internet in terms of its size and volume. Second, there has been an increase in number of multimedia and real-time applications in the network. These changes have put significant pressure on the network to support higher bandwidth and provide quality of service guarantees. Till now IP has fared well in terms of scalability because of its connectionless nature. However the hop-by-hop packet forwarding paradigm of IP is turning out to be insufficient in supporting the newer networking demands and the routers are becoming a bottleneck. As such, there is need for improvement in router/routing technology in terms of

packet forwarding performance, adapting to newer routing functionalities and providing sufficient network guarantees to support desired quality of service. Another major change has been the introduction of ATM (Asynchronous Transfer Mode) technology. Because of its fast label switching capabilities, it is being viewed by the industry and academia as a promising technology for faster networking. However, most of the current applications are based on IP and hence the success of ATM depends a lot on its capabilities to support IP and other higher layer protocols.

To counter one or more of above problems, several independent approaches have been made which can be divided into two broad categories. One is to run IP over ATM and other is to use the label switching technology of ATM for packet forwarding. This paper does a survey of these approaches.

The paper is organized as follows. Section 2 discusses initial approaches of IP over ATM, section 3 discusses Ipsilon's IP switching, section 4 discusses Cisco's Tag Switching, Section 5 discusses IBM's ARIS, section 6 discusses MPLS proposed by IETF and section 7 summarizes the paper. This is followed a list of references.

[Back to Table of Contents](#)

---

## **IP Over ATM**

This section gives a very brief overview of the efforts made by ATM Forum and IETF for running IP over ATM.

### **LAN Emulation (LANE)**

This architecture has been proposed by ATM Forum and can be considered as one of the earliest efforts for running IP over ATM. The idea is to make an ATM LAN appear as a set of logical shared medium LANs interconnected via routers. A shared LAN is emulated by setting up an ATM multicast group between all of the nodes that belong to the same logical LAN. For data transfer between nodes, an address resolution server is used to translate MAC address to an ATM address and then a point-to-point virtual channel is established between the nodes. Main disadvantages of this solution are the use of routers for transferring data within same physical ATM LAN and servers being single point of failures.

### **Classical IP Over ATM ( IPOA )**

This architecture has been proposed IETF's IPOA WG. Here a group of ATM stations are divided into Logical IP Subnets ( LIS ), interconnected via routers. Each LIS has an ATMARP server for resolution between IP address and ATM address. There is no broadcast service within a LIS. In this architecture also nodes in different LIS have to communicate via routers even if they are directly connected.

### **Next Hop Resolution Protocol ( NHRP )**

This protocol has been proposed by IETF's Routing Over Large Clouds ( ROLC ) WG to counter the problem of going through the routers between logical subnets. The goal here is to locate an exit point in the ATM cloud closest to the destination and to obtain its ATM address. Here we have NHRP servers which provide the next hop towards the destination. These NHRP servers interact with each other to get resolve the exit point closest to the destination.

## Multi-Protocol Over ATM ( MPOA )

This has been proposed by ATM Forum. It is an extension of LANE. MPOA uses NHRP to resolve the ATM address of an exit point closest to destination and provide direct layer 3 connectivity across an ATM fabric. MPOA operates at both layer 2 and layer 3. It also includes protocols to replicate servers and distribute the database for reasons of capacity and availability.

[Back to Table of Contents](#)

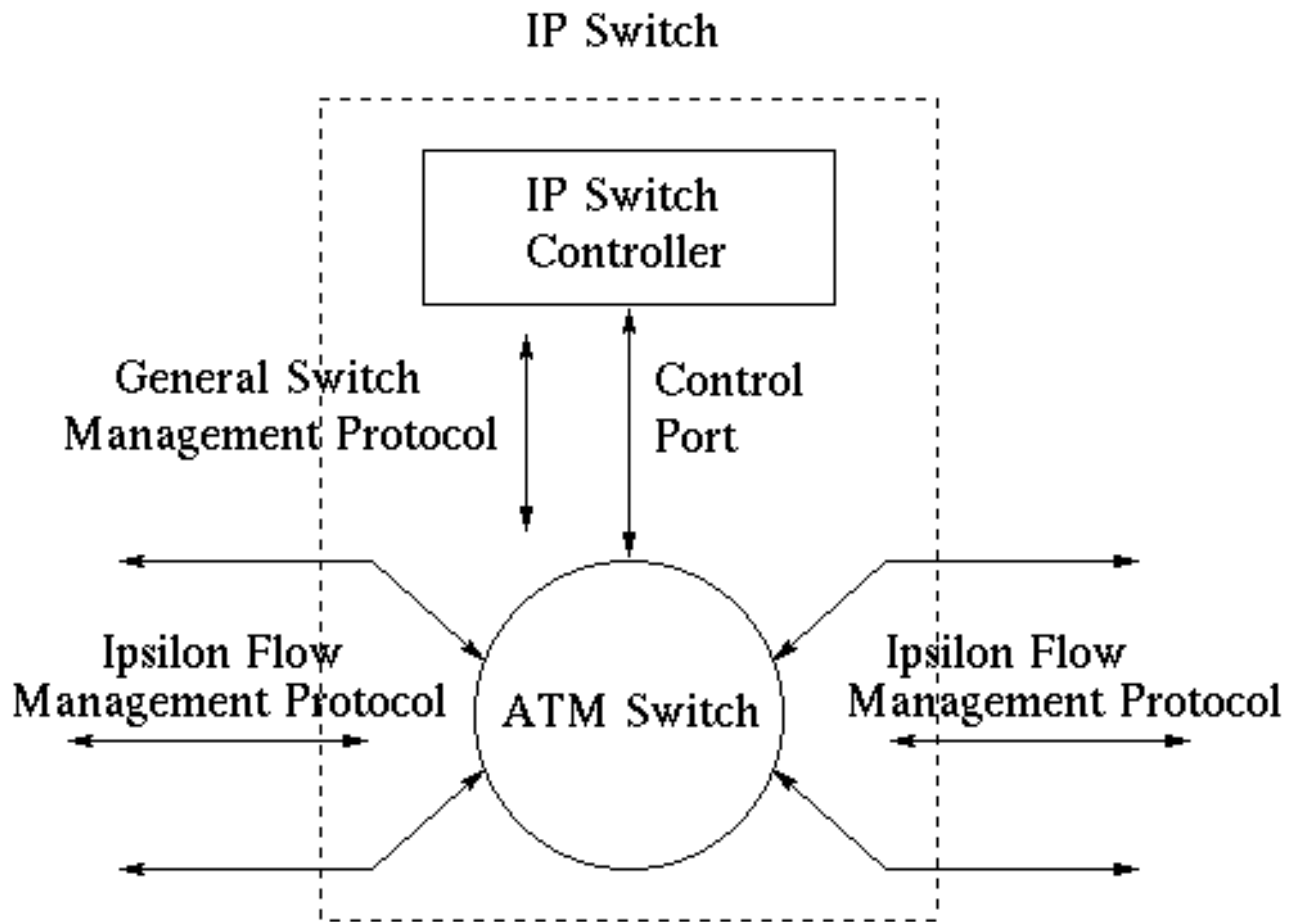
---

## IP Switching

This technology [4,Flow Labeled IP] has been developed by Ipsilon Networks Inc. The goal of Ipsilon is to make IP faster and offer the quality of service support. The "IP over ATM" approach tries to hide the underlying network topology from IP layer by treating the datalink layer as large, opaque network cloud. However, this leads to inefficiency, complexity and duplication of functionality in the resulting network[5,ATM under IP]. Ipsilon's approach is to discard the connection-oriented ATM software and implement the connectionless IP routing directly on the top of ATM hardware. This approach takes the advantage of robustness and scalability of connectionless IP and speed, capacity and scalability of ATM switches.

## IP Switch

IP switch ( refer 2 Figure 1, taken from [5, ATM under IP] ) is basically an IP router with attached switching hardware that has the ability to cache routing decisions in switching hardware. To construct an IP switch, a standard ATM switch is taken, the hardware is left untouched, but all the control software above AAL-5 is removed. It is replaced by standard IP routing software, a flow classifier to decide whether to switch a flow or not and a driver to control the switch hardware. At system startup a default virtual channel is established between the control software of the IP switch and its neighbors, which is then used for default hop-by-hop forwarding of IP datagrams. To gain the benefits of switching, a mechanism has been defined to associate IP flows with the ATM labels.



**Figure 1: Structure of An IP Switch**

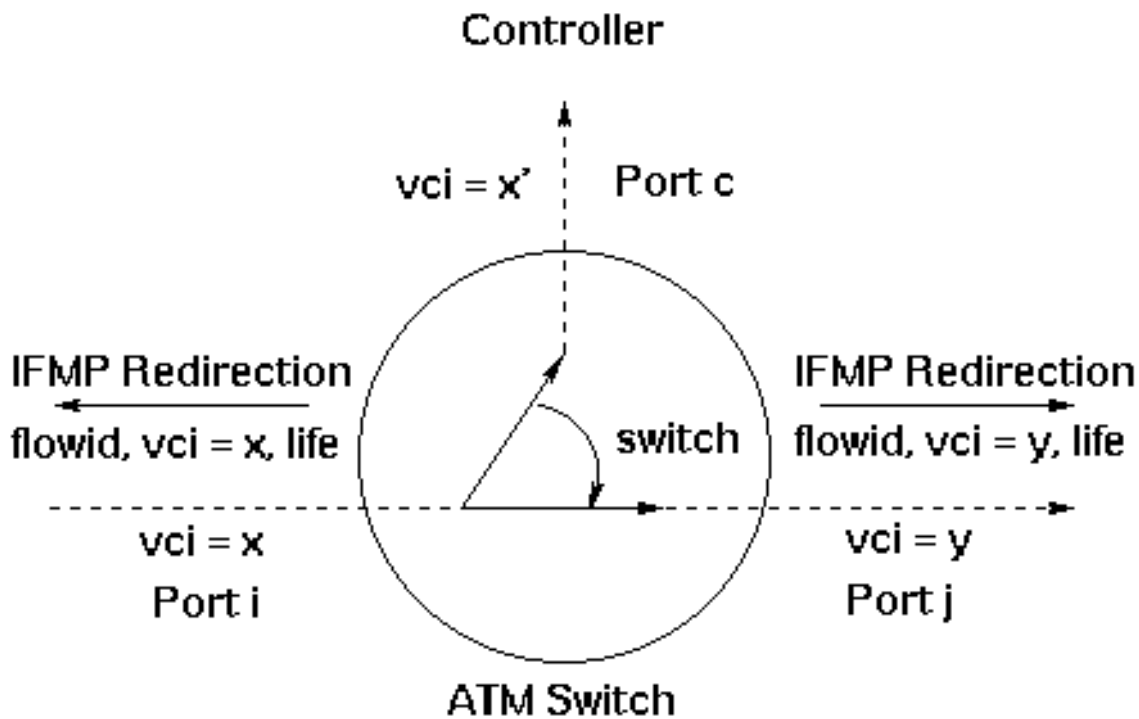
## Flow Classification

A flow is a sequence of packets from a source to a destination which get the same forwarding treatment at a router. An IP flow is characterized by the fields in IP/TCP/UDP header such as source address, destination address, port number, etc. Currently two types of flows have been defined by Ipsilon. Host pair flow type is defined for traffic between same source and destination IP address. A Port pair flow type is defined for traffic between same source and destination port between same source and destination IP address.

When a packet is assembled and submitted to controller in IP switch, apart from forwarding it to the next hop it also does flow classification. Flow classification decision is a policy decision local to a IP switch. Depending on the classification, the switch decides to forward or switch the next packets of that flow. Usually the controller would decide to forward short term flows like database queries, DNS messages and switch long term flows like ftp or telnet data.

## Flow Management Protocol (IFMP) [ RFC 1953]

When an IP switch controller decides to switch a flow, it performs the following steps to establish a flow. Please refer to Figure 2 ( taken from [5, ATM under IP]).



**Figure 2: Establishing a Switched Flow**

1. It selects a free label  $x$  on the input port  $i$  on which the packet was received, selects a free label  $x'$  on its control port  $c$ , and instructs the switch driver to map label  $x$  on input port to label  $x'$  on control port  $c$ .
2. It then sends an IFMP message ( containing flow identifier, label and lifetime ) upstream to the previous hop from which the packet came, requesting to send all further packets of that flow on the ATM virtual channel specified by the label in the message. From this point onwards the packets arrive at switch controller with label  $x'$ . Though the packets are still reassembled, the process is speeded up because the previous routing decision of the flow has been cached.
3. The switch may receive similar redirection request on port  $j$  from router at the next hop, to redirect the flow on label  $y$ . The decision to entertain such request is local and can be ignored.
4. On accepting the redirection request, the switch maps label  $x$  on input port  $i$  to label  $y$  on output port  $j$ . From this point onwards all the packets of that flow are switched in hardware and no reassembly is done.

Each IP switch maintains a refresh timer. When the timer goes off, the state of each flow is examined and either a flow is refreshed by resending the IFMP Request message or the label is reclaimed for reuse by sending an IFMP Reclaim message upstream. Also, when an IP switch redirects packets on a specific virtual channel, it removes all the header fields that characterize the flow from the header of each packet. The removed fields are stored by the router that issues the redirection message. This allows IP switch to act as a security firewall without having to inspect the contents of each packet.

## IP Switching Features

Following are some important features of IP Switching :

- **Point-to-Point** IP Switching advocates point-to-point network model for ATM rather than a logical shared medium model as proposed by some competing approaches.

- **Multicast** In case of multicast, a flow coming to an IP Switch will branch out to multiple destinations. Each of these branches can be individually redirected. If the whole flow is labeled, then multicast capabilities of ATM can be used.
- **Quality of Service** An IP Switch can make QoS decision according to its local policy. The QoS information can be included in the flow classification process. Individual QoS requests for each flow can be supported using resource reservation protocols like RSVP. An IP Switch can be configured to utilize traffic management capabilities of ATM to guarantee the desired QoS.
- **Latency** The latency in setting up of a virtual channel from source to destination can be low for connection oriented protocols like TCP. Due to 3-way handshake in TCP setup procedure, a virtual channel can be established even before the first data packet is sent across. Also the tolerance in case of failures is good because the IP Switches can always go back to connectionless packet forwarding via a different route if any link fails.

## Simulation Results

According to the simulations performed by Ipsilon[7, Iterop96], 80 - 90 percent of data in the Internet traffic is suitable for switching. The number of connection setups per second is also well within the capabilities of ATM.

[Back to Table of Contents](#)

---

## Tag Switching

This technology has been developed by Cisco Systems Inc. to enhance routing in terms of bandwidth, scalability, support to newer routing functionalities, multicast, hierarchy of routing knowledge and flexible routing control. It uses the label ( called tag here ) switching technology for layer 3 packet forwarding. The tag switching technology consists of forwarding and control components[8, RFC 2105] which are described below:

### Forwarding Component

The forwarding component uses the tags carried by the packets and the tag forwarding information stored in the switch to perform packet forwarding. It consists of following:

#### Tags:

Packets in tag switching environment can carry tags in one of the following ways:

- as a small "shin" tag header inserted between layer2 and layer3 header.
- as a part of layer 2header.
- as a part of layer 3 header.

### Tag Information Base

Routers/Switches which support tag switching are called tag-switches. A tag-switch carries tag forwarding information in a database called Tag Information Base (TIB). Each entry in the TIB is in the form of an

incoming tag with one or more subentries of the form (outgoing tag , outgoing interface, outgoing link level information)

## Forwarding Procedure

When a tag-switch receives a packet with tag, it uses the tag as an index to lookup information in TIB. On finding an entry with incoming tag equal to the tag of the packet, for every subentry of the entry, the switch replaces the tag and the link level information in the packet by the outgoing tag and link information in the subentry and sends the packet to the outgoing interface.

Since the forwarding procedure involves fixed length matching of tag entries, it can be implemented in hardware and thereby increasing the packet forwarding performance. Secondly, the forwarding procedure is independent of routing functionality ( unicast/multicast, routing protocol used, etc). As such the forwarding component is independent of the Network Layer.

## Control Component

In a tag switching there is a binding between a tag and network layer routing. A tag could be bound to an individual application flow, a single route, group of routes or a multicast tree. The primary function of control component is to create tag bindings and distribute the tag binding information among the interconnected tag-switches. The control component consists of a set of modules, one handling each routing function, thus supporting scalability of routers in terms of routing functionalities. Some of the routing functions which tag switching supports are described below.

### Destination Based Routing

Tag Switching supports destination based routing in the following manner. A tag switch, like a router, would participate in network layer routing protocols and construct its Forwarding Information Base (FIB) using the information obtained from these protocols. Then it would use one of the following 3 methods for tag allocation, creation of TIB and distribution of tag bindings.

1. **Downstream Tag Allocation** Here the downstream switch does the job of allocating a tag and binding it to a particular route or set of routes (address prefix in FIB).
2. **Downstream Tag Allocation on Demand** Here the downstream switch does the job of allocating tag and binding with address prefix, but only when it is requested to do so by an upstream switch.
3. **Upstream Tag Allocation** Here the upstream switch does the job of allocating tag and binding it with address prefix.

Once the tag allocation is done and binding created, the switch which allocated the tags, distributes it to the other interconnected tag switches by either piggybacking the binding on existing routing protocols or using a separate Tag Distribution Protocol[9, TDP] . When a previously untagged packet enters a network of tag switches, a tag is added to its header by the first tag switch in its route, then the packet is switched through the network and the tag is removed by the last tag switch in its route. Since the existence of entry in FIB causes tag allocation, hence tag switching is topology driven.

### Hierarchy of Routing Knowledge

IP routing views network as a set of routing domains. Though it separates intra-domain routing from inter-domain routing, all the routers in a domain which route transit traffic, maintain same amount of

information which degrades network performance and increases routing convergence time. To recognize the domain boundaries, tag switching allows a packet to carry a set of tags arranged in the stack form. Thus, in the IP environment a packet will have 2 tags, for inter-domain and intra-domain switching. This helps to improve overall network performance and decrease routing convergence time.

## Multicast Forwarding

To support multicast forwarding function, tag switching associates a tag with multicast tree. When a tag-switch creates a multicast forwarding entry in its TIB, it creates a local outgoing tag and a subentry for each outgoing interface.

## Flexible routing

Destination based routing do not support routing on basis of any parameter other than the destination address. This limits the capacity of network to do load balancing. Tag switching overcomes this shortcoming by allowing binding of tags to specific routes, which can be different from the routes decided by the destination address.

## Quality of Service

To support QoS selection a router should be able to classify packets as belonging to a particular class and forward them such that the QoS selection is satisfied. Tag switching helps QoS by assigning a tag to a class of packets once the classification has been done. This eliminates the need to reclassify packet at each router.

[Back to Table of Contents](#)

---

## Cell Switched Router

Cell Switched Router ( CSR ) technology has been introduced by Toshiba Corporation, Japan. "IP over ATM" approach discussed in section 1 doesn't take into account operation with newer IP protocols like RSVP. On the other hand, just the use of resource reservations and flows in IP has basic throughput limitations as compared to the cell switching mechanism of ATM. Toshiba has proposed CSR based Internetworking Architecture, which tries to merge the two approaches by extending the current routers to handle resource reservation and IP flows using ATM's cell switching capabilities.

## Architecture of CSR based ATM Internetwork

Cell switched router is a router which has cell switching functionality of ATM in addition to conventional IP datagram forwarding capabilities. As such, for datagrams which cross IP subnet boundaries, both options of IP header based routing or VPI/VCI based cell switching are available. A CSR can bypass datagram assembly/disassembly by concatenating the incoming and outgoing ATM VCs. By doing such bypass at consecutive CSRs, ATM level connectivity (called "ATM Bypass-pipe") can be provided. There are 2 different kinds of VCs defined between adjacent CSRs or CSR and end host/router.

1. **Default VC:** For datagrams which are assembled/disassembled at the CSRs.
2. **Dedicated VCs:** These carry IP flows and can concatenate with other dedicated VCs to constitute a Bypass pipe.

It is important to note that there can't be any non-CSRs in a Bypass pipe. The route for datagram forwarding is decided on the basis of a routing table entry at each CSR. The same route is also used for creating the Bypass pipe. The route for VCs between two adjacent CSRs is determined using ATM routing protocols like PNNI. The ingress router/host does the job of flow classification. It sees the header fields of each datagram (src address, destination address, port number, etc.) and decides whether the datagram should be sent on default or dedicated VC. The intermediate CSR look at the VPI/VCI of the incoming cells. If there is an entry in the ATM routing table for the VPI/VCI of the incoming cells then they are cell switched, if not then they are assembled and forwarding is done on the basis of entry in IP routing table.

## Features of CSR

The main functionality of CSR is similar to that of NHRP based IP over ATM architecture, however following are some of the features that bring out the differences between the two.

- In NHRP model, first IP level routing is used to resolve the egress point of ATM cloud and then ATM level routing is used to set up the connection between the entry and egress node. In CSR, IP level routing is used for inter-subnet routing and ATM level routing is used for intra-subnet routing. Though CSR based approach is more structured, it may not find the optimal VCC between the entry and egress nodes.
- CSR based networks have the flexibility to dynamically change routes for Bypass pipes in case of change in IP level routing information or errors in the link.
- CSR provides a natural support to multicast. A intra-subnet multicast is achieved by using point to VCs, while inter-subnet multicast is achieved by concatenation of VCs at the CSRs.

## Flow Classification

The flow classification is done by the end hosts/routers and depending of the flow they trigger Bypass pipe setup procedures. If the flow is a short lived flow, then the datagrams are forwarded on the default VCs. If the flows are long-term flows (like FTP, NNTP, etc) which don't require any bandwidth or QoS guarantees, then the end-hosts/routers may decide to forward datagrams on dedicated VCs. In this kind of flow the Bypass pipe setup procedure is sender initiated. If the flows are such that they demand some bandwidth or QoS selection (eg, RSVP), then again the datagrams are forwarded on the dedicated VCs. However the Bypass pipe setup is receiver initiated in this case. Depending on the requirements of the flow, the CSRs can use corresponding service class (ABR/UBR/CBR/VBR) provided by the ATM.

[Back to Table of Contents](#)

---

## Aggregate Route Based IP Switching

This concept has been introduced by IBM. The goal of ARIS[15, Internet Draft] is to improve aggregate throughput of IP and other network layer protocols by switching datagrams at the media speed.

## ARIS Mechanism

An Integrated Switched Router (ISR), is a switch augmented with the standard network layer routing support. A switched path is a path in the network along which datagrams can be "switched" at media speed. In a network which uses destination based hop-by-hop datagram forwarding, ARIS uses ISRs instead of switches/routers and pre-establishes switched paths to well known egress nodes. These nodes are recognized from the information provided by routing protocols like OSPF and BGP. The switched paths are established by the ISRs by exchanging ARIS messages described later. They are in the form of tree routed at the egress points. On receiving a datagram, the ingress ISR performs a conventional longest prefix match lookup in its forwarding information base to get the associated switched path label for the destination and then transmits the datagram on that switched path. If no entry is found in the FIB then the datagram is forwarded to the next hop using the default hop-by-hop forwarding path.

## ARIS Messages

ARIS is protocol independent. In case of IP, ARIS messages are transmitted with IP protocol number 104. Following messages are used by ARIS to setup and manage the switched paths.

- **Init** This is the first message sent by an ISR to its neighbors to notify its existence. It is periodically retransmitted till a positive acknowledgment is received.
- **KeepAlive** This message is sent by an ISR to its neighbors to inform them of its continued existence. This message is sent only if no other ARIS messages have been sent to the neighbors in a given time period.
- **Establish** This message is sent by an ISR periodically or in response to a Trigger message, to each of its upstream neighbors to setup or refresh a switched path. An ISR on receiving an Establish message for an egress identifier, verifies whether the path is correct and loop-free. If not then it demolishes the old path and sets up a fresh one using Establish messages.
- **Trigger** This message is sent by an ISR to its new downstream neighbor requesting an Establish message, after it tears-down its old switched path to its egress identifier.
- **Teardown** This message is sent when an ISR loses its connectivity to the egress identifier.
- **Acknowledgment** This message is sent in response to ARIS messages. The acknowledgment can be positive or negative.

## Egress Identifiers

Egress identifiers define a routed path through the network. They are classified into different types according to the routing protocol information. The various types of egress identifiers in ARIS are:

1. **IP destination prefix** Here each IP destination prefix has its own switched path tree. This type is used for routing protocols like RIF.
2. **Egress IP Address** Used in BGP and OSPF.
3. **OSPF router ID**
4. **Multicast (source, group) pair** Used by multicast protocols like DVMRP, MOSPF, etc.
5. **Multicast (ingress-of-source, group) pair** Used for multicast protocols like MOSPF, PIM-SM.

## ISR Information Base

An ISR maintains following three information bases for management of switched paths and routing datagrams.

### Routing Information Base (RIB)

This is same as RIB in a standard router. In an ISR, apart from computing best-effort routes, RIB is also used to identify egress points and egress identifiers used in other two information bases.

### Forwarding Information Base (FIB)

This contains entries of the form destination prefix, outgoing interface, next hop IP address, egress identifier and the label for the associated switched path. An ISR looks into its FIB for packet forwarding. While the association between next hop IP address and egress identifier is created by IP routing protocol, the association between egress identifier and switched path label is created by ARIS protocol.

### VC Information Base (VCIB)

This information base maintains the mapping between egress identifiers and upstream/ downstream labels. This mapping is created using ARIS protocol.

## ARIS Features

Following are some of the important features of ARIS.

- **TTL Decrement** ARIS can support TTL Decrement in IP datagrams by maintaining hop count per egress identifier. Then, the ingress ISR would decrement the TTL field in the datagram before forwarding it to the switched path.
- **Loop Prevention** ARIS prevents the creating of switched path loops by sending "ISR ID" list with the establish messages while creating the switched paths. Secondly, even if transient IP loops are created by route changes, ARIS prevents creating of corresponding switched path loops by modifying the switched paths with the route changes.
- **Explicit Routes** ARIS supports building of switched paths through explicit routes. This is to support source based routing, setting up of bidirectional switched paths and paths for multicast.
- **Multicast** To support multicast, ARIS establishes point-to-multipoint switched path tree routed at the ingress node. To support source specific multicast tree, ARIS uses (source, Group) pair for the egress identifier and to support shared tree ARIS uses (\*, Group) or (RP, Group) for the egress identifier.
- **Label Conservation** ARIS does label conservation by mapping several IP destinations to same egress switch and enabling merging of switched paths.
- **Flexibility** Unlike other competing technologies, ARIS useful even in networks in which only some of the routers are ISRs or arbitrary network routing topologies exist.
- **L2 Tunneling** In L2 tunneling a L2 pdu is encapsulated in another L2 pdu. ARIS supports L2 tunneling by allowing inner shell labels to be advertised to the upstream ISRs by using the ARIS establish messages.
- **Quality of Service** ARIS can be extended to support QoS. However the current version doesn't address this issue.

## Multi-Protocol Label Switching (MPLS )

As discussed in previous sections, several approaches were made in last couple of years to apply layer 2 switching technology (based on label swapping) to layer 3 routing. This led IETF to form MPLS group. The goal of this group is to develop a standard [18, Internet Draft] for integration of layer 2 switching with layer 3 routing in order to improve price/ performance of network layer routing, scalability of network layer and provide greater flexibility in providing new routing services. Though MPLS focuses on IP, it is extendible to other network layer protocols and no assumption is made about layer 2 protocols, except that they are capable of forwarding network layer packets.

### Definitions

A few definitions that would be used extensively in this section are given below. They have been taken directly from the Internet Draft [18, Internet Draft]:

- **Label** A short fixed length physically contiguous locally significant identifier which is used to identify a stream
- **Label Information Base ( LIB )** The database of information containing label bindings
- **Forwarding Equivalence Class** A group of layer 3 packets which are equivalent in the way forwarding is applied to them and they can be assigned same label.
- **Label Stack** It is an ordered set of labels.
- **Label Switched Router ( LSR )** It is a router with MPLS capabilities.
- **MPLS Domain** A contiguous set of nodes in one routing or administrative domain which are capable of doing MPLS routing.

### Core MPLS Components

There a set of core components defined by MPLS which can be applied in various ways to achieve the objectives of MPLS.

#### Basic Routing Approach

The approach taken by MPLS is to use standard layer 3 routing protocols like OSPF and BGP without any change. The information maintained by these protocols can be used for generation and distribution of labels. There should be a single routing protocol for routers and switches. Also, the MPLS-capable and MPLS-ignorant routers should be able to co-exist in the same routing domain.

#### Labels

Labels facilitate packet forwarding. The key issues involved with labels are:

##### Label Semantics

MPLS applications should be clear about the semantics they associate with labels. At the simplest level,

labels can be a short fixed length representation of packet header, used for indexing the forwarding decision at the routers. However, its possible to associate more semantics with the labels. For example, a label can represent a loop free path or can be the identity of a higher layer protocol type. In any case, labels always have local significance.

### Label Granularity

The essential point about labels is that all the packets with the same label will receive same forwarding treatment ( ie, forwarded on the same port, with same next hop label (if any) and using same encapsulation ). The granularity of element associated with labels can vary widely. On one hand a label can represent IP flows, explicit routes or host routes. On the other hand, it could represent an egress identifier or egress router (in case of IP unicast ) and a multicast tree ( in case of IP multicast ).

### Label Assignment

The key point in label switching is the binding between a label and network layer routing. Label assignment involves allocating a label, binding it to routing and distributing it to MPLS switches. Label assignment can be control-traffic driven or data-traffic driven. Though control-traffic driven label assignment has many advantages, data-traffic driven label assignment cannot be done away with. In both cases, label withdrawal is an essential associated functionality which should be taken care of.

### Label Stack and Forwarding Operations

The basic forwarding operation involves label swapping as follows. When a packet is received by a label switch, its label is used for index lookup in LIB and information like encapsulation, outgoing interface and outgoing label is retrieved. When a packet first enters an MPLS domain, a label is inserted along-with normal packet forwarding, the process being called "Label Push". When a packet leaves an MPLS domain, the label is removed from the packet, the process being called "Label Pop". In general, a packet can carry a stack of labels, with the label on the top of stack being relevant at any instance.

### Encapsulation

Label swapping uses several pieces of information like label, TTL field, etc. This information can be encoded in a layer 2 header, then called "MPLS Encapsulation" or an explicit "MPLS Header" could be added to packets. The size, contents and format of MPLS Encapsulation/ Header are still being discussed.

## Observations, Issues and Assumptions

Several observations, issues and assumptions are being discussed by the MPLS group.

1. **Layer 2 Vs Layer 3 Forwarding** The relative performance of layer2 and layer 3 forwarding can vary from node to node. Also layer 3 forwarding cannot be done away with.
2. **Scaling Issues** MPLS can achieve "O(n) scalability" by using the facts that routing follows an inverted tree routed at the destination and the number of destinations can be reduced using routing aggregation. However the definition of "n" in O(n) scalability is imprecise.
3. **Types of Stream** There are four types of stream recognized by MPLS. They are point-to-point, point-to-multipoint, multipoint-to-point, multipoint-to-multipoint.
4. **Loops Handling** Packet looping, at the least causes incorrect delivery and network congestion. MPLS will be designed with mechanisms to prevent loop and / or limit the amount of resources that will be

eaten up due to presence of loops.

- 5. Operations and Management** MPLS must support O & M at least as extensively as supported by IP currently. Though in most cases it is straightforward to do so, its not clear how to implement Traceroute over ATM or Frame Relays (with the concept of TTL absent in them).

## Technical Approaches

This section mentions some of the technical approaches for MPLS framework. For details please refer to [18, Internet Draft].

- **Label Distribution** Label Distribution Protocol (LDP) being developed by MPLS WG is going to adopt elements from several approaches. Two approaches in consideration are topology driven (as in Tag Distribution Protocol, ARIS) and traffic driven (as in FANP). In this both downstream and upstream label allocation method are being considered. Apart from above "Explicit Label Distribution", some implicit distribution schemes like Piggybacking on other control messages is also being considered.
- **Stream Merging** MPLS uses stream merging to achieve scalability. While Frame merging is straightforward, it is not so easy to perform stream merging in ATM environment. VC Merge requires buffering at the routers and introduces latency while VP merging requires distinct VCIs within a VP to distinguish source.
- **Loop Handling** Loop Handling can be done in three ways: Loop Survival, Loop Detection and Loop Prevention. Methods for loop survival are use of TTL or Dynamic Routing and fair queuing. The schemes that are being considered for loop detection are transmitting of "Loop Detection Control Packets ( LDCP )" along the path towards a destination whenever the route to the destination changes and counting the hops to each egress node based on the routes currently available. Any loop prevention method would involve some kind of explicit signaling.
- **Interoperation with NHRP** When MPLS is used over ATM, then there is a possibility of LSR also being the NHC ( Next Hop Client ). In order to switch cells between two technologies without reassembly, encapsulation must be consistent between MPLS and NHRP, NHRP should understand the granularity of stream and MPLS should understand the granularity of VC. Another problem that arises is that VC is bi-directional while MPLS is uni-directional.
- **Multipath** There are three approaches that are being considered for handling multipaths in MPLS. One maintains a separate switched path from each ingress node via multipath nodes to the destination, other establishes only one switched path from any ingress node to destination and the third approach allows splitting of a single incoming stream into multiple outgoing streams at a multipath node.

[Back to Table of Contents](#)

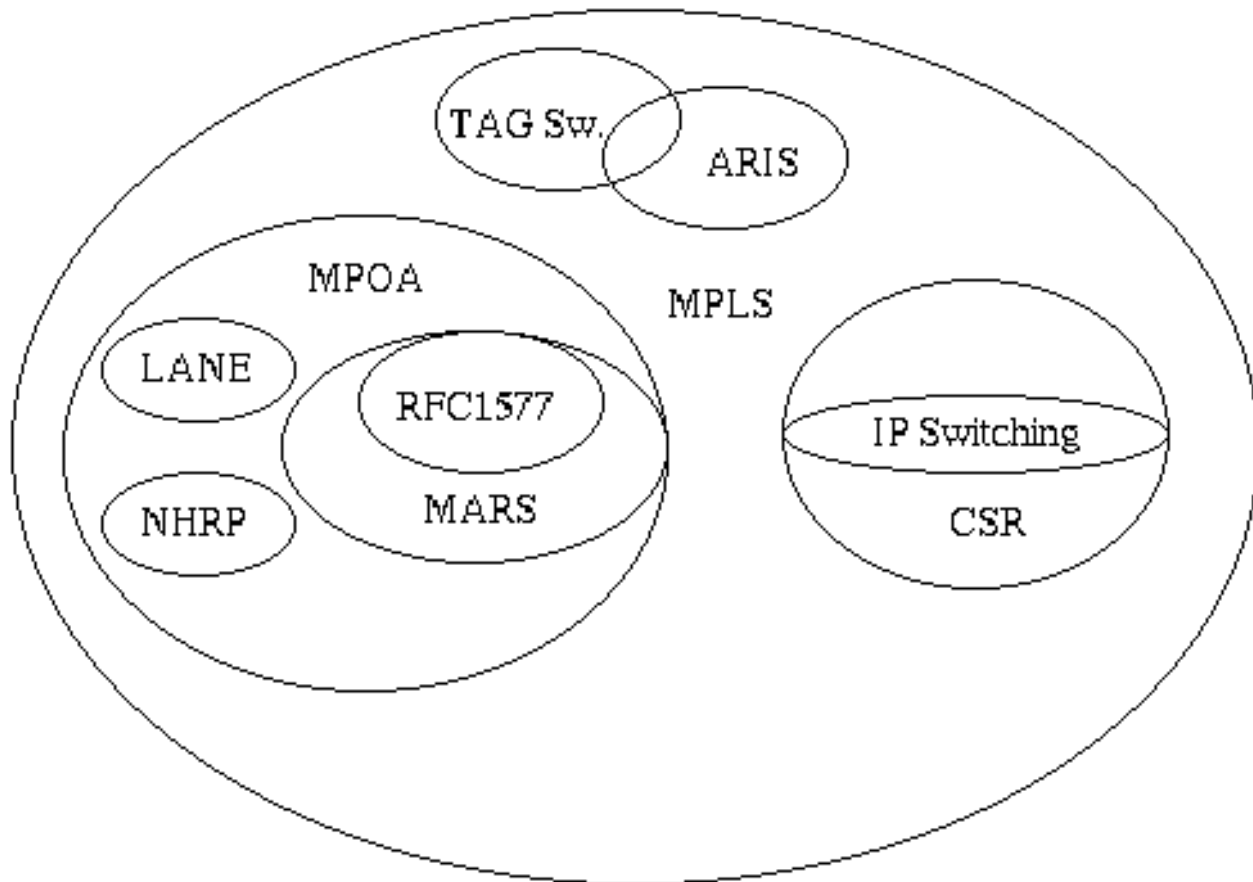
---

## Summary

Rapid growth in size of Internet, increase in number of real-time and multimedia applications, and some other changes in network traffic pattern have motivated several companies and organizations like ATM Forum and IETF to use label swapping technology for network layer forwarding. In this paper we studied different approaches to achieve the same.

Section 2 presented a brief overview of approaches made by IETF and ATM Forum for running IP over

ATM. In section 3 we discussed "IP Switching" technology proposed by Ipsilon. Here ATM switches are modified to do flow classification and switch IP flows at layer 2. In section 4 we discussed "Tag Switching" proposed by Cisco Systems. Here a small tag is carried in the packet header ( implicitly or explicitly ) and the tag is used to switch packets at the routers. In section 5 we studied "CSR based Internetworking Architecture" proposed by Toshiba Corporation, Japan. Here ATM Bypass pipes are used to switch IP packets at layer 2. In section 6 we studied " Aggregate Route based IP Switching" proposed by IBM. This approach establishes a set of switched paths in the network and uses binding of labels to the switched paths for switching packets. Finally, in section 7 we studied "Multiprotocol Label Switching" that is being developed by MPLS WG of IETF. This aims a providing a set of "core" MPLS mechanisms for standardising integrating of layer 2 switching with layer 3 routing. All these approaches can be seen together in a venn diagram ( taken from [20, Raj Jain] ) as shown below:



**Figure 3: Approaches in Layer 3 Switching**

[Back to Table of Contents](#)

## References

1. RFC 1953, " Ipsilon Flow Management Protocol Specification for IPv4, Version 1.0," May 1996,
2. RFC 1954, " Transmission of Flow Labelled IPv4 on ATM Data Links Ipsilon Version 1.0", 05/22/1996,

3. RFC 1987, "Ipsilon's General Switch Management Protocol Specification Version 1.1", 08/16/1996,
4. P. Newman, T. Lyon, and G. Minshall, "Flow Labelled IP : A Connectionless Approach to ATM", IEEE Infocom, March 1996. *A paper describing Ipsilon's IP switching technology.*
5. P. Newman, G. Minshall, and T. Lyon, " IP Switching: ATM Under IP ", *An extended version of Infocom96 paper with implementation details and simulation results.*
6. Peter Newman, Greg Minshall, Tom Lyon, and Larry Huston, "IP Switching and Gigabit Routers", preprint of paper in IEEE Communications Magazine, Jan. 1997, *Compare and contrasts IP Switching with fast routers.*
7. Peter Newman, Tom Lyon, and Greg Minshall, "Flow Labelled IP: Connectionless ATM under IP", Networld+Interop, April 1996, *A paper describing Ipsilon's IP Switching technology.*
8. RFC 2105, "Cisco Systems' Tag Switching Architecture Overview", 02/06/1997, 13 pp.,
9. "Tag Distribution Protocol", 05/23/1997, "Use of Tag Switching With ATM", 01/17/1997,
10. RFC 2098, "Toshiba's Router Architecture Extensions for ATM: Overview", 02/04/1997, 18pp.,
11. RFC 2129, " Toshiba's Flow Attribute Notification Protocol ( FANP ) Specification", April 1997,
12. Internet Draft, "ARIS: Aggregate Route-Based IP Switching", 03/26/1997,
13. Internet Draft, "ARIS Specification", 03/26/1997,
14. Internet Draft, "ARIS Support for LAN Media Switching", 03/27/1997,
15. Internet Draft, "A Framework for Multiprotocol Label Switching", 05/12/1997.
16. Internet Draft, "A Proposed Architecture for MPLS", July 1997,
17. Raj Jain, " Recent Advances in Networking ( Summer'97 Class Handouts )", <http://www.cis.ohio-state.edu/~jain/cis788-97/index.html>

[Back to Table of Contents](#)

---

Last Modified August 16, 1997