

## QoS Switching

H. T. Kung  
Division of Engineering  
and Applied Sciences  
Harvard University

November 4, 1998

1 of 40

## Acknowledgments

Many slides are from Professor Raj Jain of Ohio State.

But opinions/biases and any mistakes are mine.

3 of 40

## Two Related Areas to Cover

- (1) Switched IP Forwarding
- (2) 802.1Q (Virtual LANs) and 802.1p (GARP/Priorities)

Both efforts had origins in ATM:

- ATM switches-->IP switching.
- ATM Forum tried to define VLANs in 1994, and decided to let IEEE do it.

Both efforts intend to realize certain advantages that ATM has been trying to achieve, e.g., L2 overlay and virtual connections for LANs

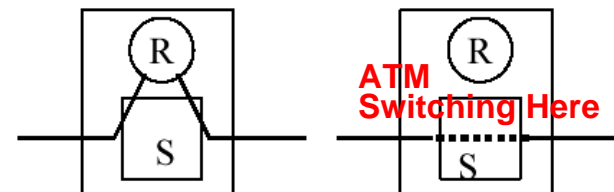
2 of 40

## IP Switching

Developed by Ipsilon (1995-96), trying to sell ATM switches.

Routing software in ATM switches

Initially, packets are reassembled by the routing software and forwarded to the next hop. Then long term flows are transferred to separate VCs. No cell reassembly. A similar scheme is called "Cell Switched Router (CSR)".



4 of 40

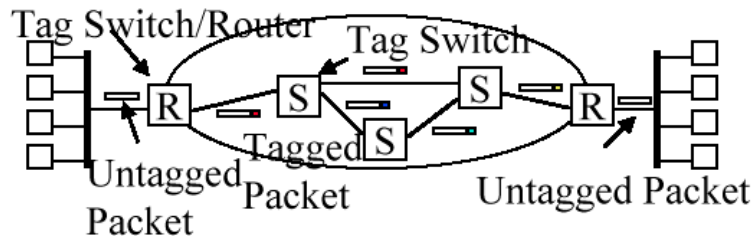
## Tag Switching /1

Proposed by Cisco

Entry router/switch attaches a label (tag) to the packet based on the route

Switches switch packets based on labels ==> Fast

Exit router/switch strips off the tag



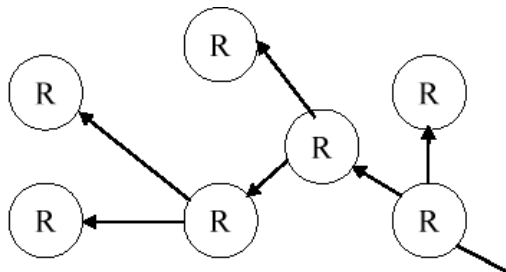
5 of 40

## ARIS (Aggregate Route-Based IP Switch)

Proposed by IBM

Topology based; one VC per egress router.

Egress router initiates the setup of switched path



7 of 40

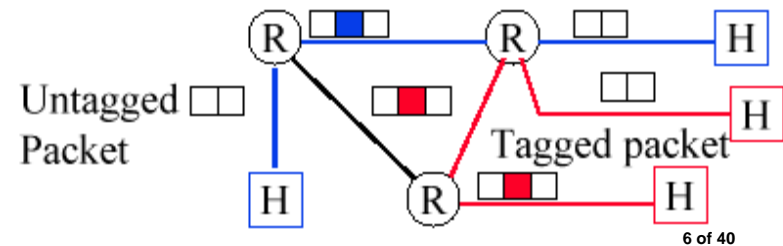
## Tag Switching /2

Similar to VLAN tags (discuss later)

Tags can be explicit or implicit L2 header



Ingress router/host puts a tag



6 of 40

## Switched IP Forwarding Comparison

Issue	IP Switch	CSR	Tag	ARIS
Datalink	ATM	ATM, FR	ATM, FR, Ethernet	ATM, FR
Network Layer	IP	IP	IP, XNS, ...	IP
Initiator	Downstream	Both	Both	Egress
VC Setup Protocol	IFMP	FANP	TDP	ARIS
Mapping	Traffic	Traffic	Topology	Topology
# of VCs	# of L4 flows	# of L3 flows	# of routes	# of Egress routers

8 of 40

# MPLS (Multiprotocol Label Switching)

IETF working group to develop switched IP forwarding

Combining various features of IP switching, CSR, tag switching, ARIS

Technology extendible to other L3 protocols

Not specific to ATM. Not specific to a routing protocol (OSPF, RIP,...)

9 of 40

# MPLS for Traffic Engineering: A Useful Way of Looking at MPLS

“Requirements for Traffic Engineering Over MPLS” by authors from UUNET-Worldcom, Internet Draft, Oct 1998.

- Conventional IP routing is driven by topology (i.e., #hops), and does not take into account of bandwidth availability and traffic characteristics
- MPLS can support traffic engineering via L2 trunking:
  - Classify packets (only once!)
  - Map packet classes onto trunks
  - Map trunks onto physical routes

10 of 40

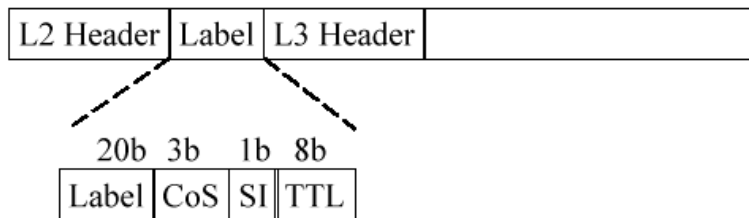
## Label Format in MPLS

Labels = Explicit or implicit L2 header

TTL = Time to live

CoS = Class of service

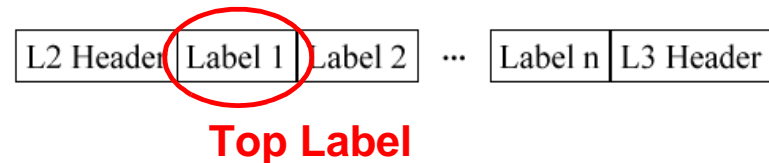
SI = Stack indicator



11 of 40

## Label Stack

Labels are pushed/popped as they enter/leave MPLS domain



12 of 40

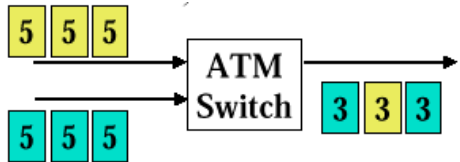
## Stream Merging

Required for egress based labels.  
Helpful for mpt-to-pt streams

In ATM/AAL5, cells of frames on the same VC cannot be intermingled  
⇒ VCs cannot be merged.

VC-merge: Store all cells of a frame and forward together  
⇒ Buffering, Delay.

VP Merge: VPI = Labels, VCI = source



13 of 40

## MPLS Issues

For ATM, multiple tags per destination may be used to avoid frame merging

Loop prevention, detection, survival

Multicast: Multiple entries in label information base

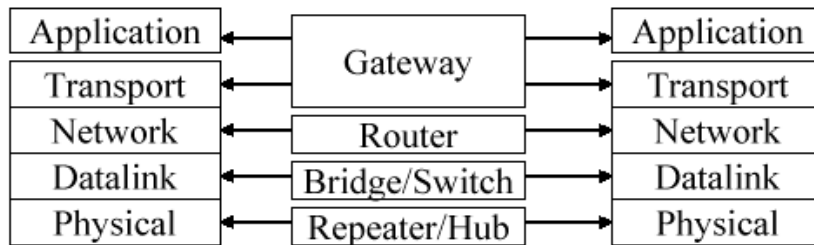
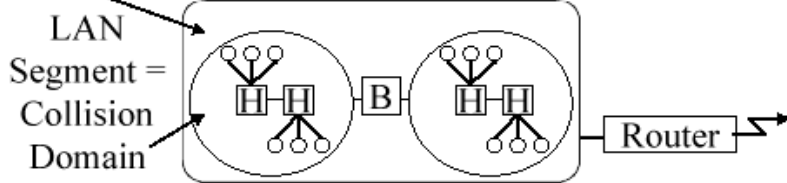
Multipath: Streams going to the same destination but different sources/port # may be assigned separate labels.

Host involvement: Label-enabled hosts will avoid first hop reassembly

14 of 40

## Review of Interconnection Devices

LAN = Broadcast domain



15 of 40

## Interconnection Devices

Repeater: PHY device that restores data and collision signals

Hub: Multiport repeater + fault detection and recovery

Bridge: Datalink layer device connecting two or more collision domains. MAC multicasts are propagated throughout the “extended LAN.”

Router: Network layer device. IP, IPX, AppleTalk. Does not propagate MAC multicasts.

Switch: Multiport bridge with parallel paths

16 of 40

## Traffic Classes in LAN

IEEE 802.1p working group

Goal: To support time-critical  
(continuous media) traffic

Method:

1. Prioritizing traffic
2. Efficient support of multicasting

Bridge filtering database for each port  
indicates whether any members of the  
group exist on the port  
⇒ Need group registration protocol

17 of 40

## IEEE 802.1p: Features

Allows up to 8 traffic classes (priorities)

Priority ⇒ Both queueing and access  
(Allows queueing priority on LANs that  
have no access priorities, e.g., Ethernet)

Different number of priorities on different  
ports

Allows dynamic multicast filtering

Applies to all 802 MAC protocols + FDDI

802 MAC Protocols: 802.3 (Ethernet),  
802.4 (Token Bus), 802.5 (Token Ring),  
802.6 (DQDB), 802.9 (Integrated  
Services), 802.12 (Demand Priority)

19 of 40

## What's in a Name?

The “p” in 802.1p is lower case.

Uppercase letter ⇒ Base standard

Lowercase letter ⇒ supplement

802.1p is a supplement to 802.1D bridge  
standard

802.1Q is a base VLAN standard

802.3z is a 1000 Mbps supplement to  
Ethernet Standard

18 of 40

## Recommended Four Priorities

Time and safety critical

Time critical

Non-time critical, loss sensitive

Non-time critical, loss insensitive

20 of 40

## How is Priority Set?

Priority may be set by user, destination address, input port, output port, access priority, or by VLAN

A priority may be assigned for a port  
==> For a source station connected to a switch

In some LANs, priority can be encoded in frames. In some other LANs, use 802.1Q VLAN tags.

21 of 40

## Scheduling

Separate queue for each priority

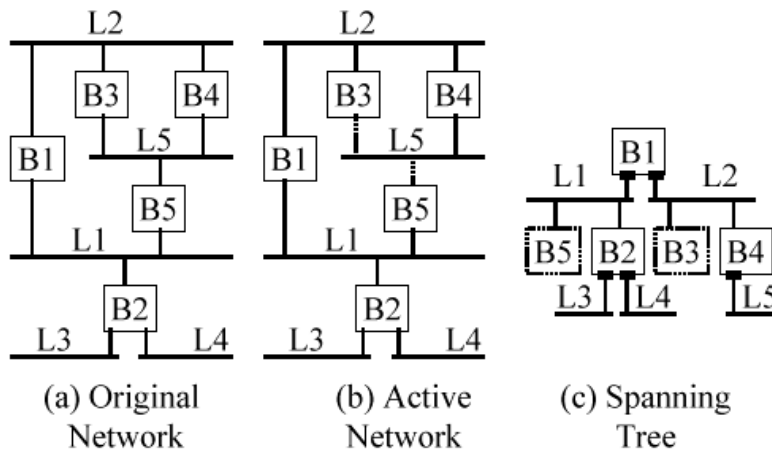
Higher priority first (strict priority)  
==> This is dangerous. Why?

No reordering of frames for a given priority and a given source and destination address pair

22 of 40

## Multicast: Today

Bridges forward multicast on all active ports. A spanning tree is formed to avoid loops



23 of 40

## Dynamic Multicast Filtering

Send multicast frames only on LANs where receivers exist.

Filtering database may be used.

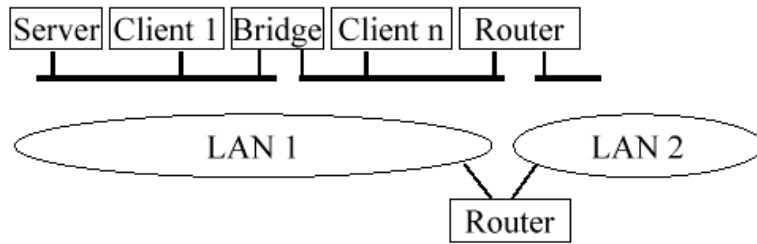
A directed subtree of the spanning tree is formed for each group. Multicast frames are forwarded along the directed graph

Source Pruning: Sources can stop transmission if there are no members

GARP (Generic Attribute Registration Protocol) for general purpose registration/distribution

24 of 40

## Traditional LAN



LAN = Single broadcast domain  
= Subnet

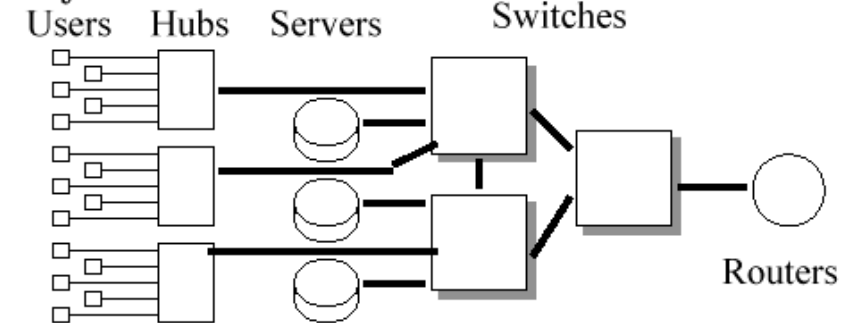
No routing between members of a LAN

Routing required between LANs

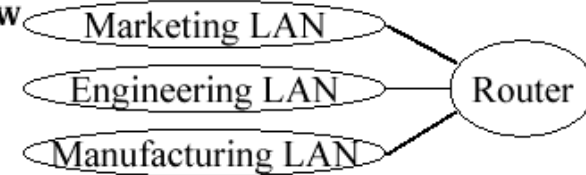
25 of 40

## VLAN (Virtual LAN)

### Physical View

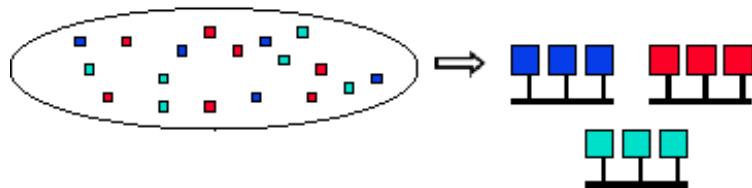


### Logical View



26 of 40

## VLAN



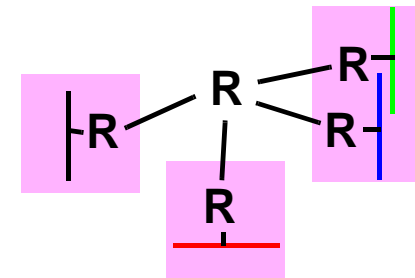
Virtual LAN = Broadcasts and multicast goes only to the nodes in the virtual LAN

LAN membership defined by the network manager ⇒ Virtual

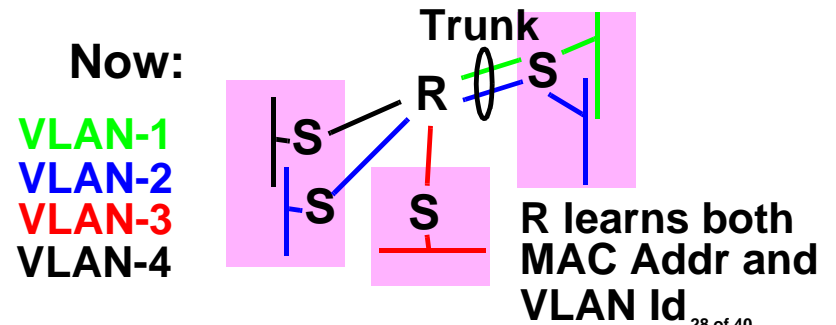
27 of 40

## A Typical Usage of VLANs

Before:



Now:



28 of 40

# 1. VLAN Eases Management

Location-independent

⇒ Marketing LAN can be all over the campus

Users can move but not change LAN

Traffic between LANs is routed

⇒ Better to keep all traffic on one LAN

Switch when you can, route when you must

⇒ Do not VLAN over expensive WAN links

## Types of VLANs

Layer-1 VLAN = Group of Physical ports (Harvard FAS uses this.)

Layer-2 VLAN = Group of MAC addr

Layer-3 VLAN = IP subnet

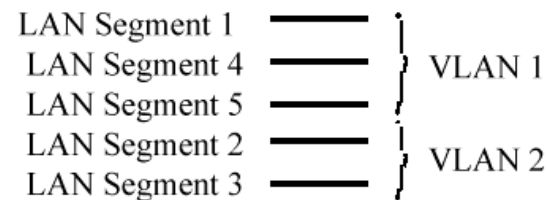
Switch Port	VLAN		VLAN1		VLAN2	
	1	2	MAC	IP	MAC	IP
A1	√		A1B234565600	23.45.6	21B234565600	
A2		√	D34578923434		634578923434	
A3	√		1345678903333		8345678903333	
B1		√	3438473450555		9438473450555	
B1	√		4387434304343		5387434304343	
			4780357056135		6780357056135	
			4153953470641		9153953470641	
			3473436374133		0473436374133	
			3403847333412		8403847333412	
			3483434343143		8483434343143	
			4343134134234		0343134134234	

# 2. VLAN Saves Money

Many of today's campus environments are **fiber-scarce**, **bandwidth-rich** (100s Mbps backbone) and **packet-processing-easy** (only 10,000s packets per second needed). Thus use of VLANs makes great sense.

- For Harvard FAS, a centralized VLAN routing module on a switch takes care of \*all\* 60+ VLAN subnets of 15,000 users.
- A single fiber is typically shared by many VLANs

## Layer-1 VLANs



Also known as port switching  
 Can be used to provide security and isolation

Does not allow user mobility  
 Moved user has a new subnet  
 ⇒ new IP address

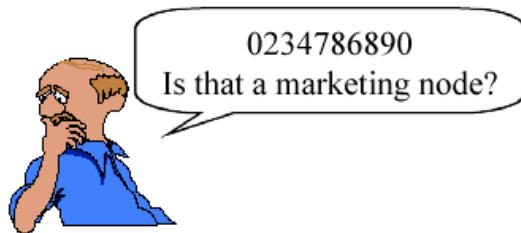
## Layer-2 VLANs

LANs defined by a list of MAC addresses

Provides full user movement

Clients and server always on the same LAN regardless of location

Problem: Too many addresses need to be entered and managed



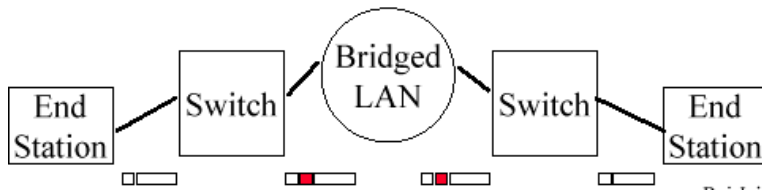
33 of 40

## VLAN Tagging

Dest. Addr	Src. Addr	VLAN Tag	Prot. Type
------------	-----------	----------	------------

First switch adds VLAN id to all incoming packets, and last switch removes the tag

Intermediate switches do not recompute the VLAN id, unlike VC Id or labels



35 of 40

## Layer-3 VLANs

Dest. Addr	Src. Addr	Protocol Type
------------	-----------	---------------

IP Dest. Addr	IP Source Addr
---------------	----------------

Also known as virtual subnet

VLAN membership implied by MAC-layer protocol type field and subnet field

VLAN configuration is learned by the switches

Stations do not belong to VLANs, packets do. Multiprotocol stations are put into multiple VLANs

34 of 40

## IEEE 802.1Q Features

Allows up to 4095 VLANs

Allows port based, MAC address based, and higher-layer VLANs

Allows mixing legacy bridges and VLAN-aware bridges

Extends 802.1p priority mechanism to priority based on VLAN membership

Allows priority associated with each VLAN

36 of 40

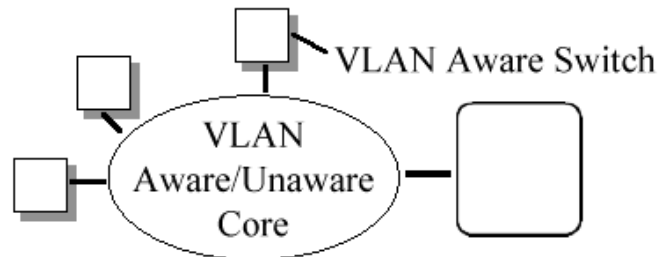
## Communication Between VLANs

Need routers

Can use 1-armed VLAN-aware router

VLAN-aware switches can route between VLANs

Such switches can be placed in the core, in the edges, or everywhere



37 of 40

## Some Issues/Questions

### MPLS:

- Do we really want to take on yet another set of routing protocols, tag-base routing tables, and tag distribution protocols?
- Will “extended” Frame Relay or “IP Relay” do the same job?

### VLAN:

- Will one heavy VLAN user kill others?
- How to assign priorities to VLANs sharing the same link?
- Companies such as Cisco, Bay Networks, and 3Com have switches which support 802.1Q/p. Q is ok, but what about p?

38 of 40

## Concluding Remark: Be Careful and Be Real

This is a an (very) unforgiving field:

In 1993 ATM was in, but it was out of favor in 1997

In 1995 IP switching was in, but it was out of favor in 1997

In 1996 RSVP was in, but it is out of favor in 1998

In 1998 Diff-Serv and MPLS are in, but will they also be out of favor soon?

39 of 40

## References

1. Multiprotocol Label Switching (mpls) working group at IETF
2. RFC 2105, “Cisco Systems' Tag Switching Architecture Overview”, 02/06/1997, <http://ds.internic.net/rfc/rfc2105.txt>
3. “ARIS: Aggregate Route-Based IP Switching”, 03/26/1997, <http://www.internic.net/internet-drafts/draft-viswanathan-aris-overview-00.txt>
4. 802.1Q/p specs. See <http://grouper.ieee.org/groups/802/1/index.html>

40 of 40