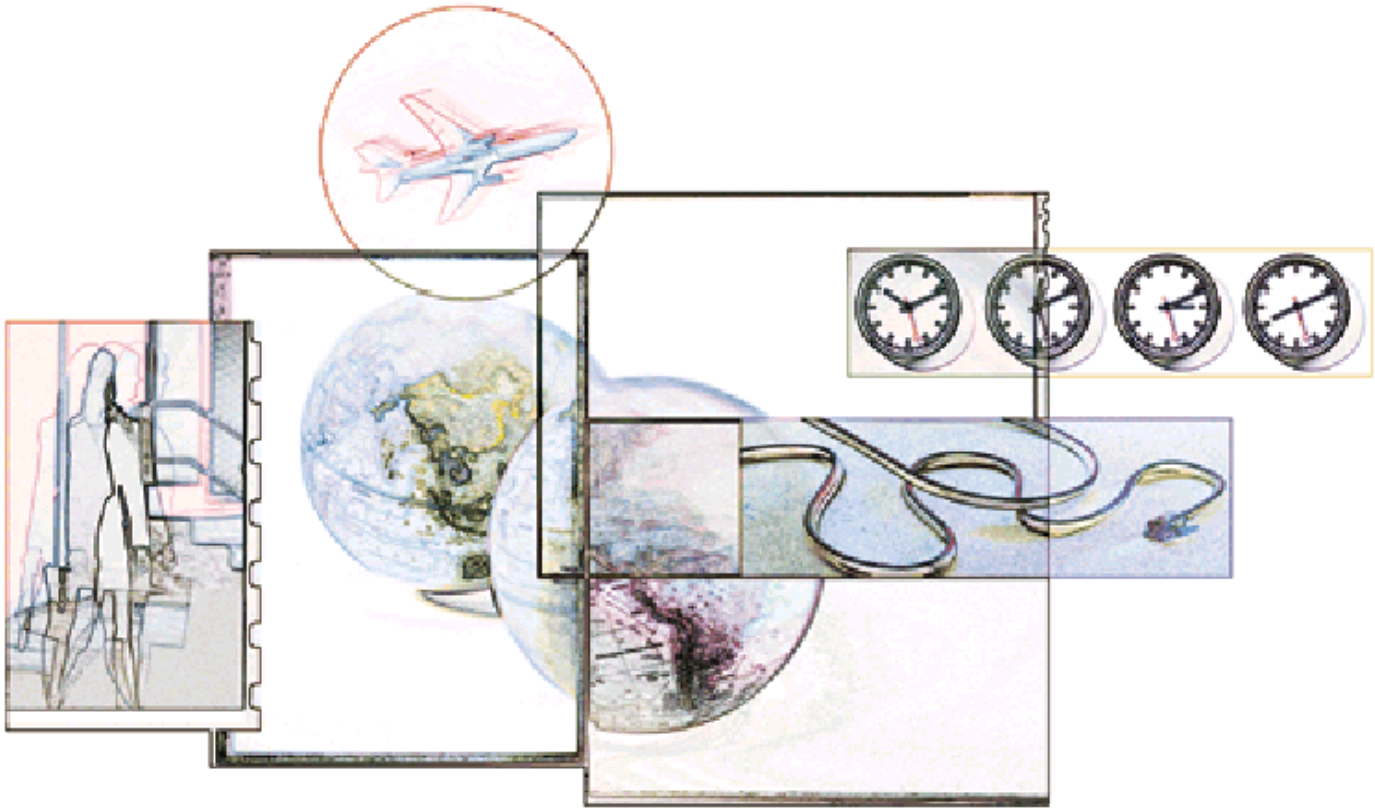
An Overview of IP Switching Technology

Chris Metz
IBM Corp.
cymetz@us.ibm.com



Contents

- Motivations
- Definitions
  - Peer vs Overlay
  - Label Swapping
  - VP/VC Merging
- IP Switching Types
  - Flow-based
  - Topology-based

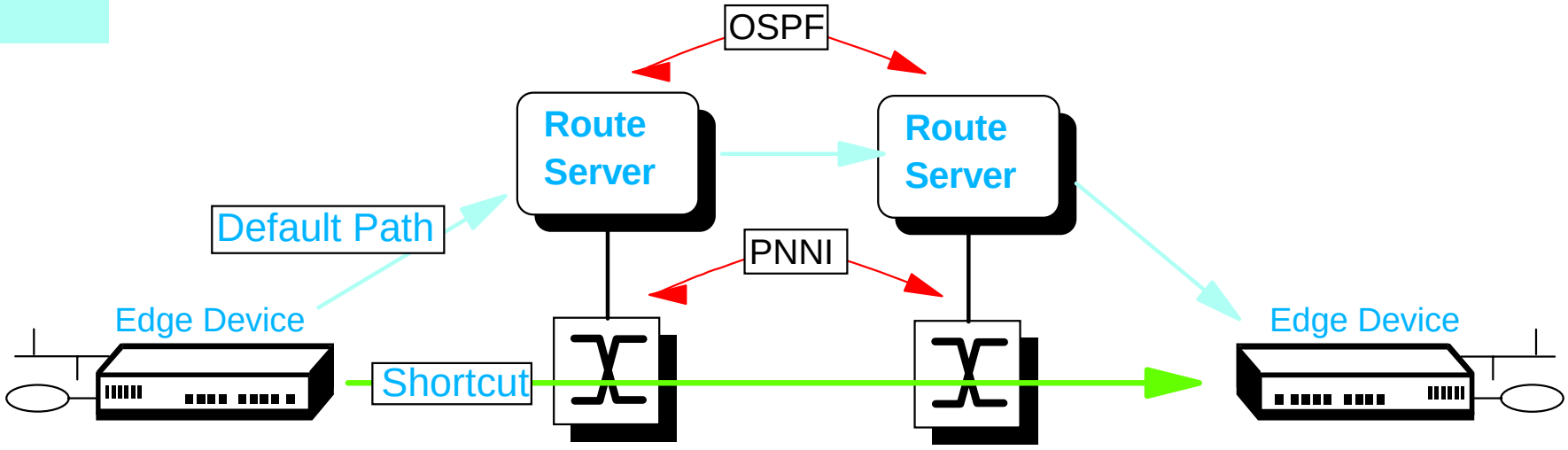
Contents

- Solutions
  - NHRP/MPOA
  - IFMP/GSMP (Ipsilon)
  - Tag Switching (Cisco)
  - ARIS (IBM)
- MPLS
- IBM Efforts
- Summary
- References

Motivations

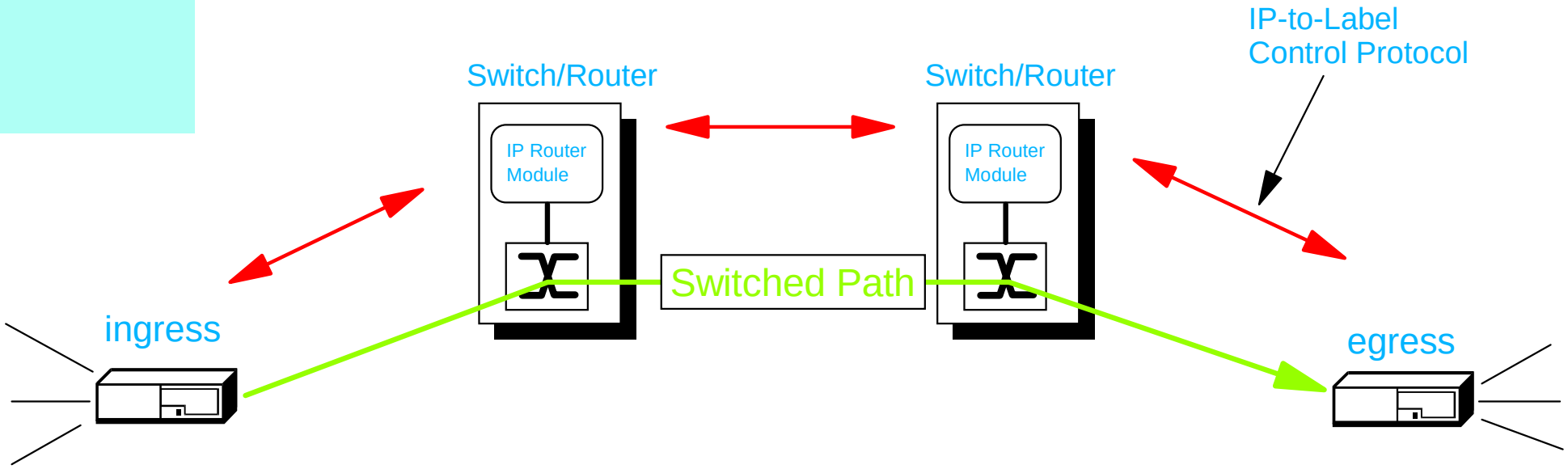
- Processor-based per-packet forwarding performed by today's routers is not fast enough
- Utilize hardware-based switching to improve price/performance of IP routing
- Maintain the dynamics, robustness and scalability of IP while exploiting hardware-based switching
- Simplify VC establishment through switch hardware
- Keep it simple and interoperable with today's router-based networks

Definition: Overlay Model



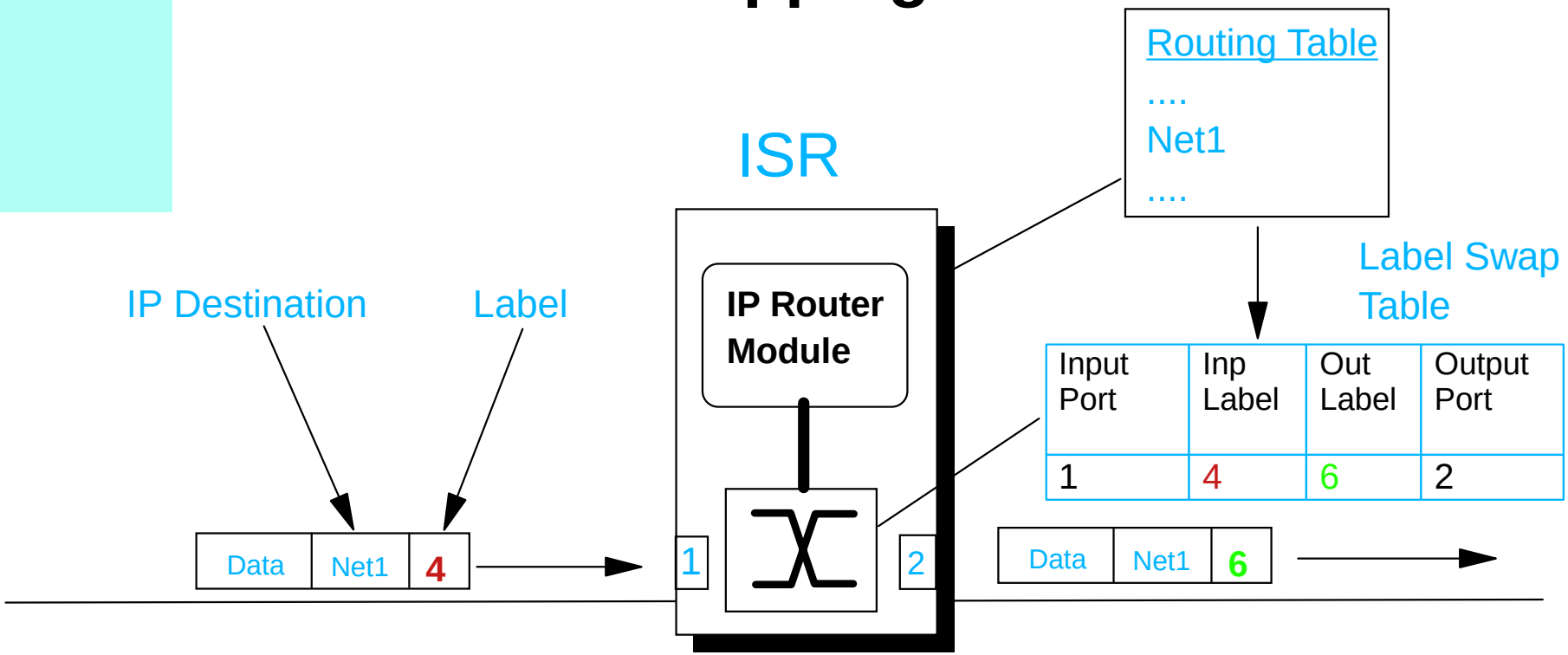
- Separate addressing and routing at layer-3 and layer-2
- Provides default routed path (thru route server) and shortcut path between edge devices/hosts
- ATM Forum MPOA is an example

Definition: Peer Model



- Switch/Router devices run single routing protocol, maintain a single address space and network topology
- Use specialized control protocol (e.g. IFMP, ARIS) to distribute routing/flow-to-label mappings among participating devices
- Packets/cells are switched from network ingress to egress

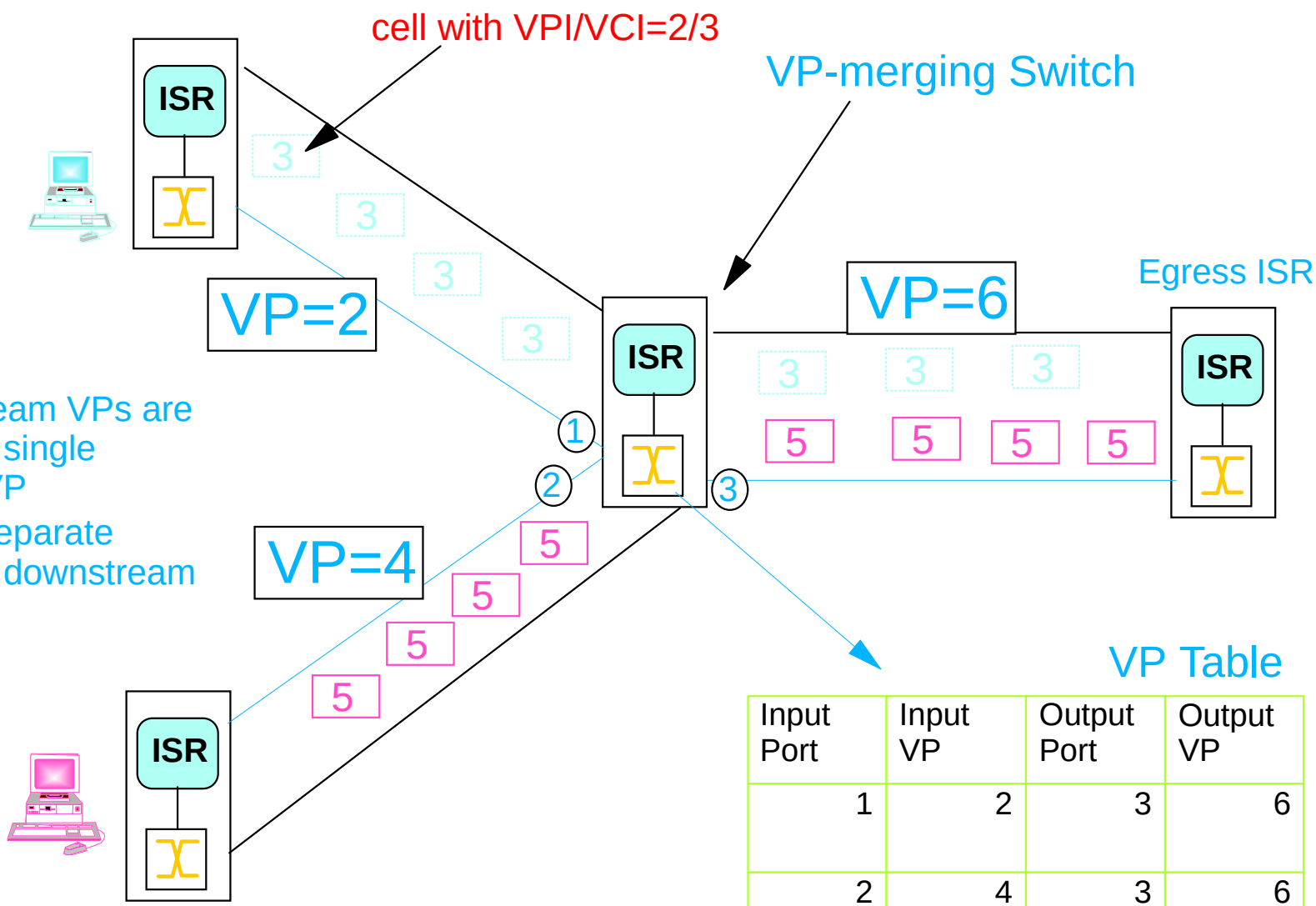
Definition: Label Swapping



- IP Routing (or flow) information is mapped to labels
  - switch's label-swap tables are updated with the new labels
- Packets/cells are forwarded using hardware-based switching (e.g. ATM). No per-packet processing at layer-3 is required!!
- Leverage switching maturity/price/performance
  - multigigabit throughput of switches exceeds that of routers

Definition: VP Merging

- **Observe that:**
  - multiple upstream VPs are merged into a single downstream VP
  - still requires separate upstream and downstream VCs

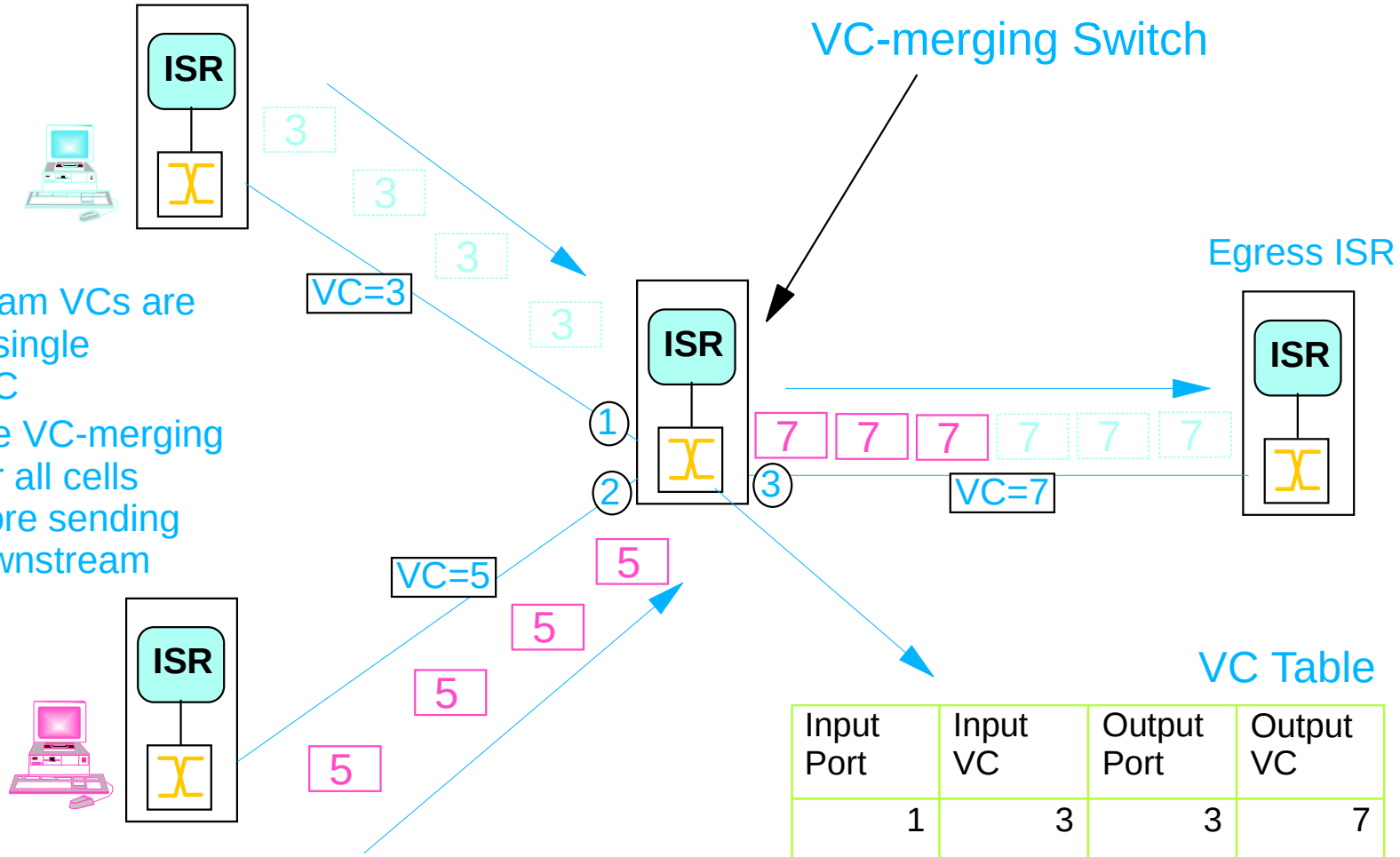


VP Table

Input Port	Input VP	Output Port	Output VP
1	2	3	6
2	4	3	6

Definition: VC Merging

- **Observe that:**
  - Multiple upstream VCs are merged into a single downstream VC
  - may require the VC-merging switch to buffer all cells of a frame before sending on the new downstream VC

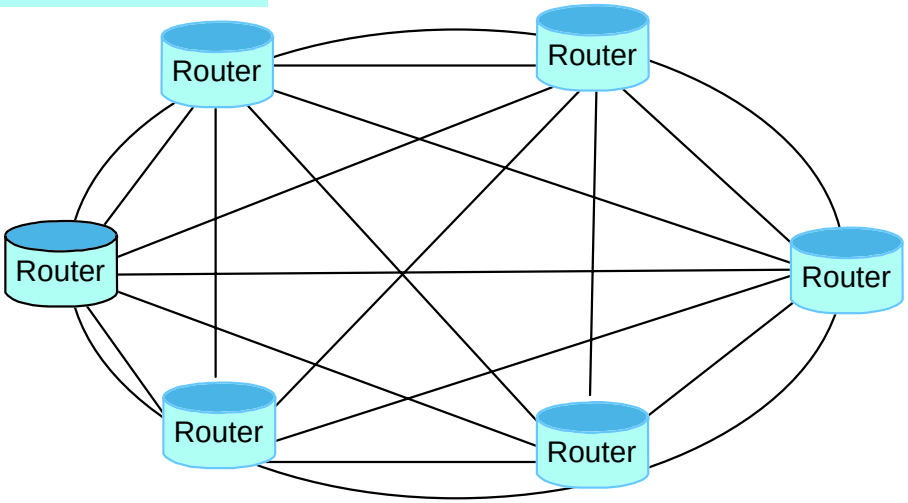


VC Table

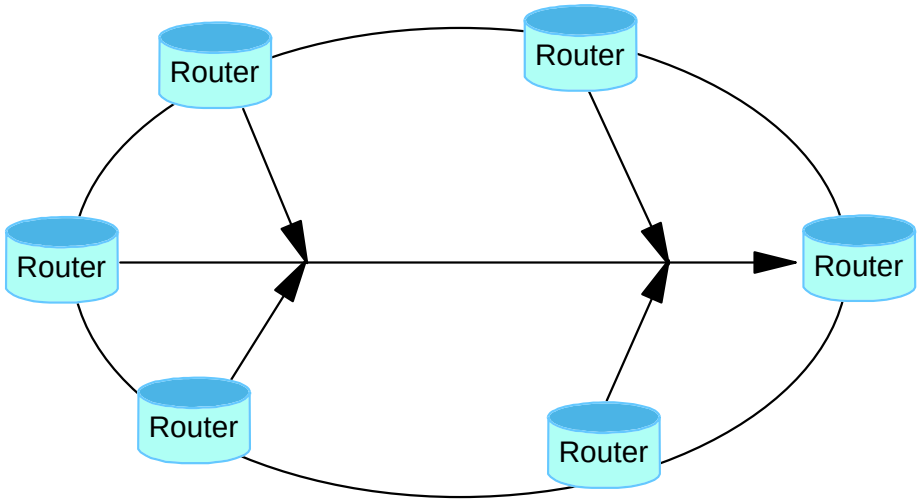
Input Port	Input VC	Output Port	Output VC
1	3	3	7
2	5	3	7

Why is Merging Important?

Full Mesh



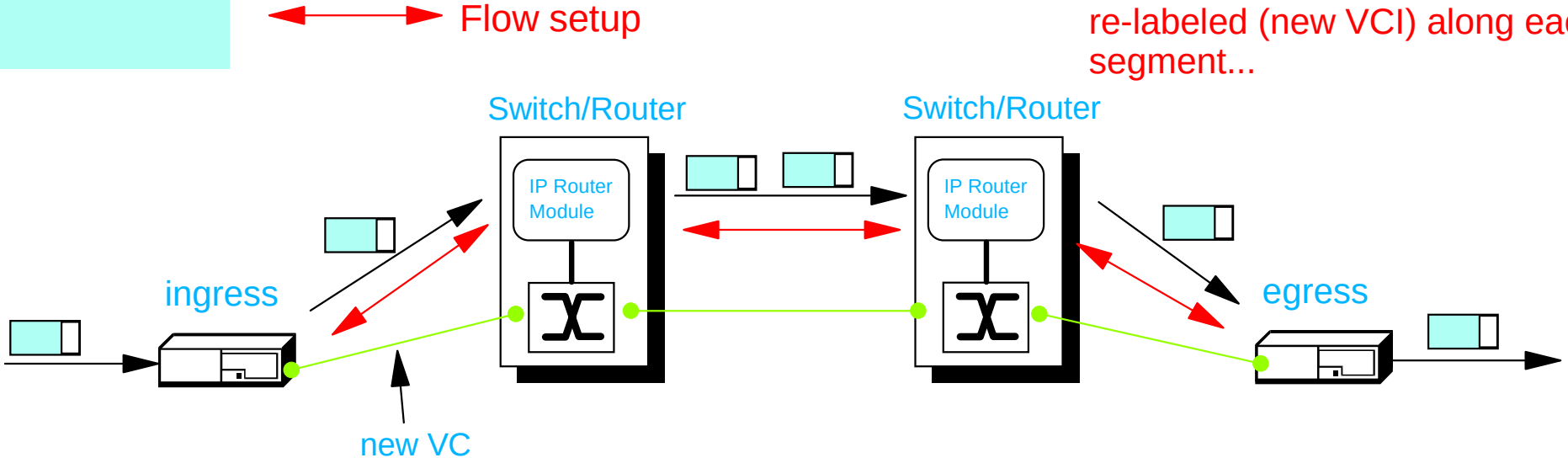
VP/VC Merging

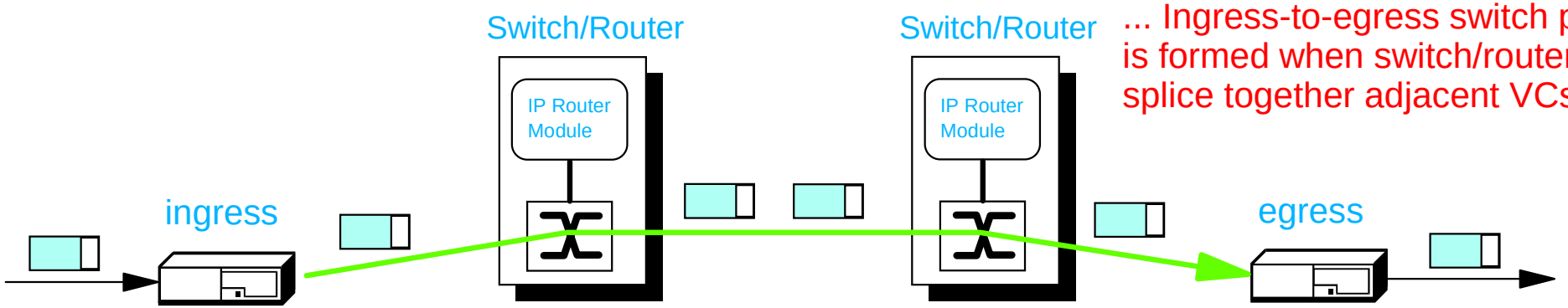


- Without Merging interconnecting routers in a full mesh requires O(N^2) connections where N=number of routers
  - may exceed network resource and operational capacity of the network
- With Merging, only need O(N) connections - N x number of multipoint-to-point VCs
- Important for large N
  - remember additional connections for special services (QoS) may be needed between any two nodes
- Scales like a router network

IP Switching: Flow-driven

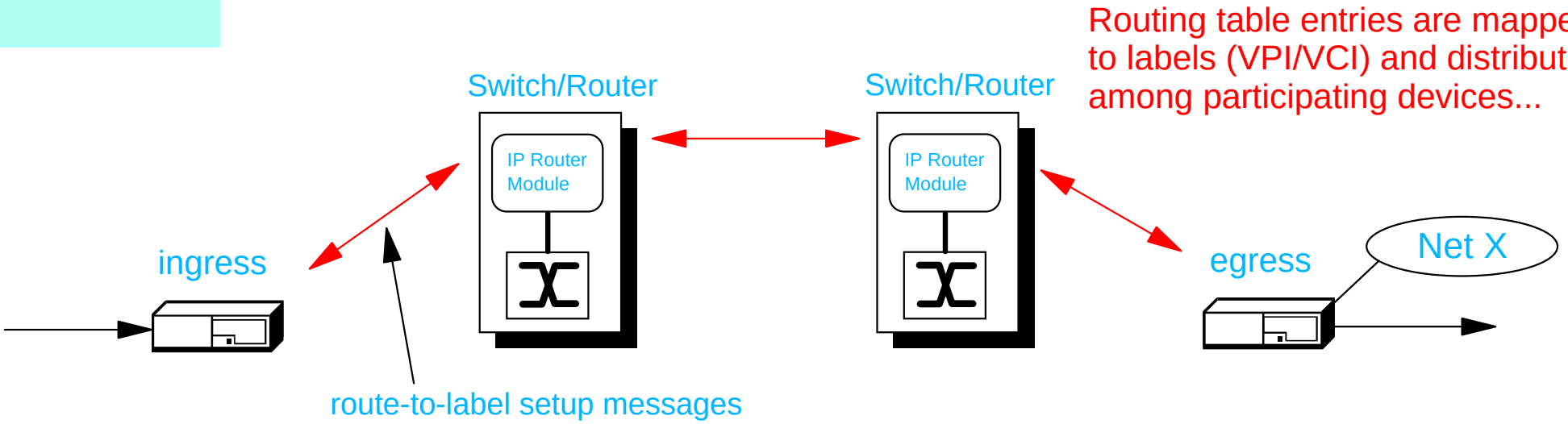
Arrival of flow triggers exchange of flow setup messages between adjacent devices. Cells of a flow are then re-labeled (new VCI) along each segment...

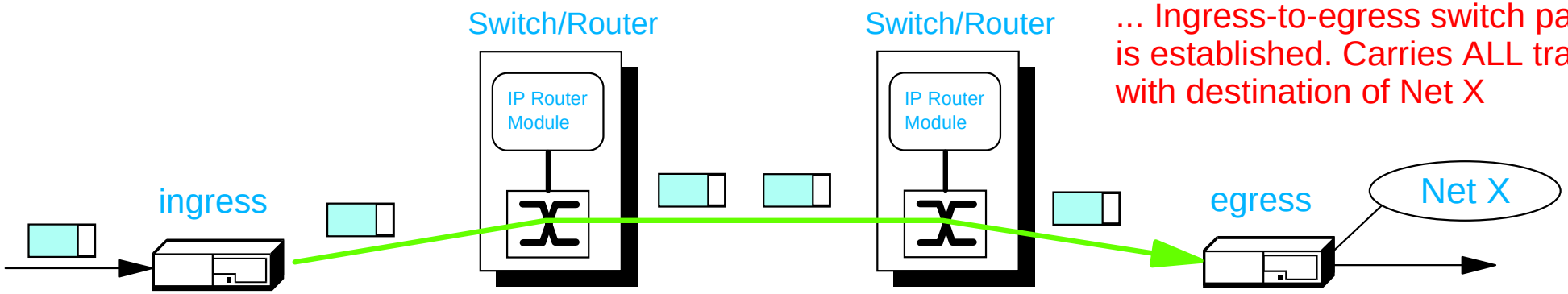




... Ingress-to-egress switch path is formed when switch/routers splice together adjacent VCs

IP Switching: Topology-driven



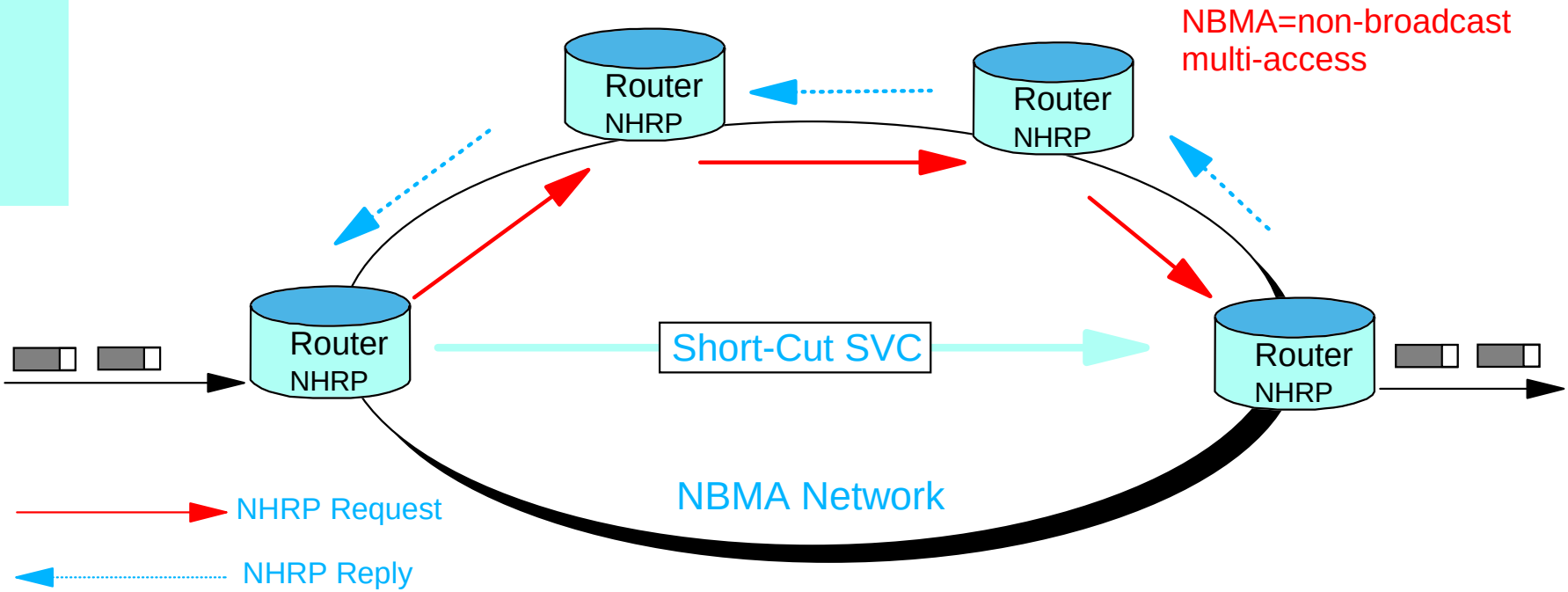


IP Switching: Flow-driven versus Topology-driven

- **Flow-driven**
  - Performed for each IP Flow
  - each host/host or appl/appl pair
- Advantages
  - provides both default and switched paths
  - on demand allocation of switched resources
- Disadvantages
  - path setup latency
  - does not scale for many flows
  - only "tail" of flow is switched - not the entire flow
- Examples: MPOA, IFMP

- **Topology-driven**
  - Performed only after topology change
- Advantages
  - scalability
  - low overhead
  - flow aggregation
  - all traffic is switched
- Disadvantages
  - possibility of loops
  - difficult to deliver QoS to aggregated flow
- Examples: Tag Switching, ARIS

Next Hop Resolution Protocol

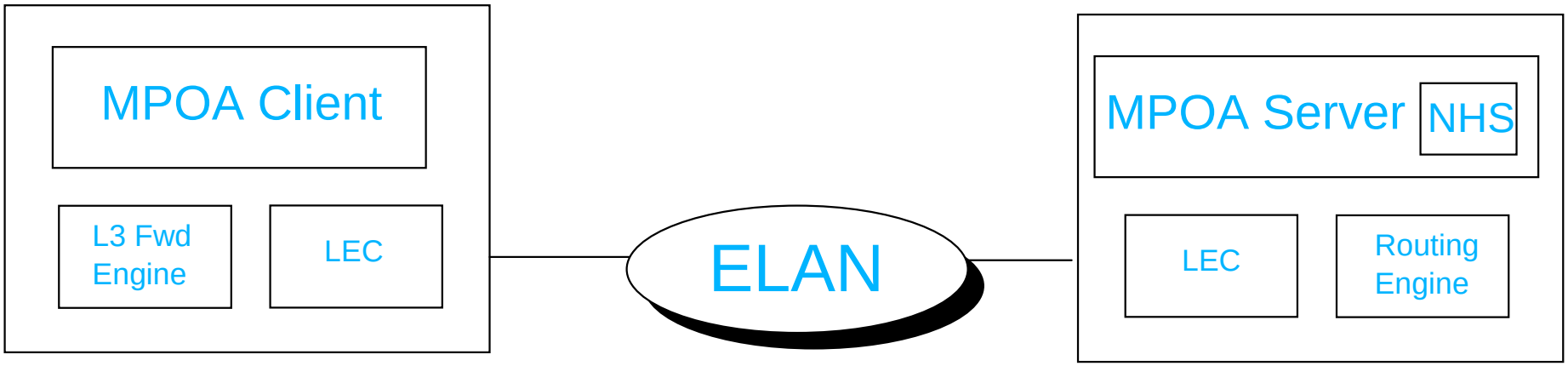


- NHRP is a query/response address resolution protocol
- Used to bypass router hops across NBMA network
- NHRP messages are routed hop-by-hop to the destination
- NHRP Router closest to destination returns NBMA address to requester who then sets up shortcut SVC
- Incorporated into MPOA

Solutions: MPOA

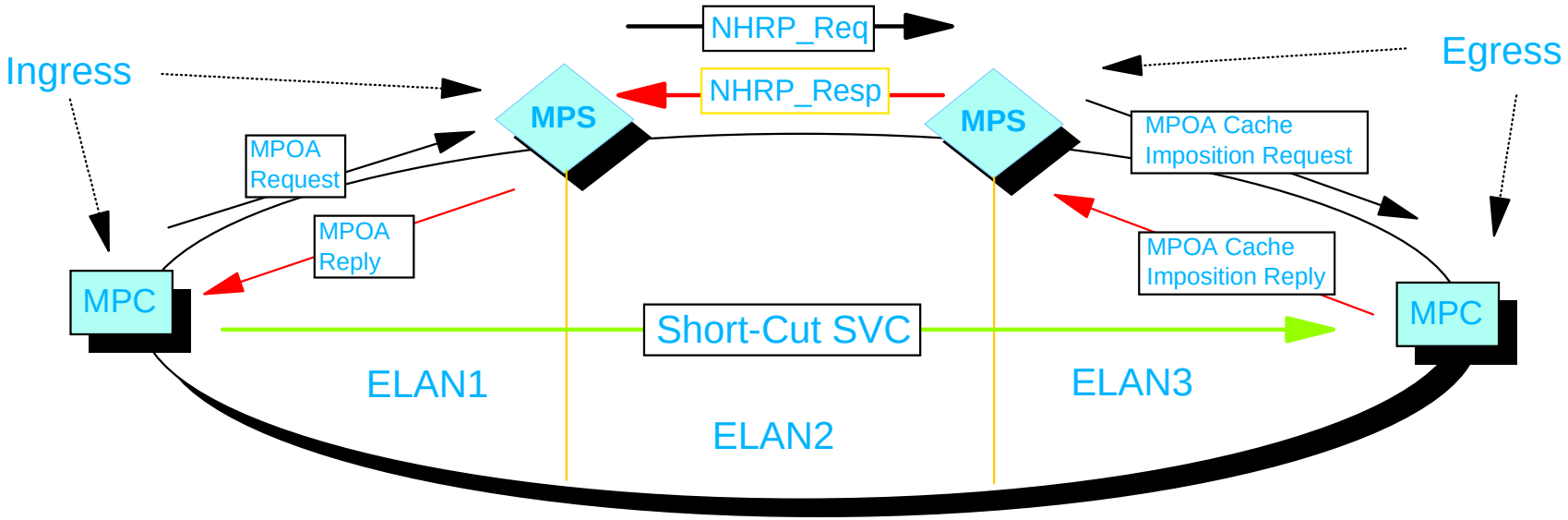
- Purpose is to provide "..efficient transfer of inter-subnet unicast data in a LANE environment.."
  - incorporates LANE and NHRP
- Separates the Route Calculation and Forwarding functions between Route Servers(MPOA Server) and Hosts/Edge Devices(MPOA Clients)
- Example of the Overlay Model
- Flow-driven
- Defines the protocols that enable MPCs and MPSs to communicate over an ATM network

MPOA Components



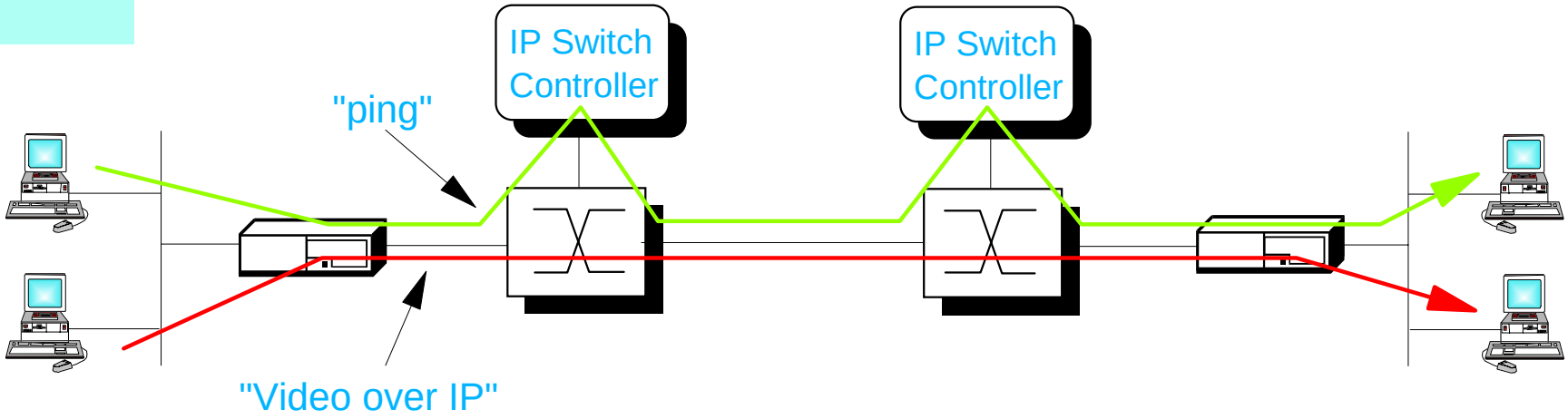
- MPC
  - MPOA client
  - source and sink of short-cut routes
  - performs layer-3 forwarding
  - issues MPOA queries to MPS
  - Examples: workstation, edge device
- MPS
  - MPOA Server
  - performs layer-3 route calculation
  - performs default forwarding
  - responds to MPOA queries
  - Examples: Route Server (MSS)

MPOA Flow



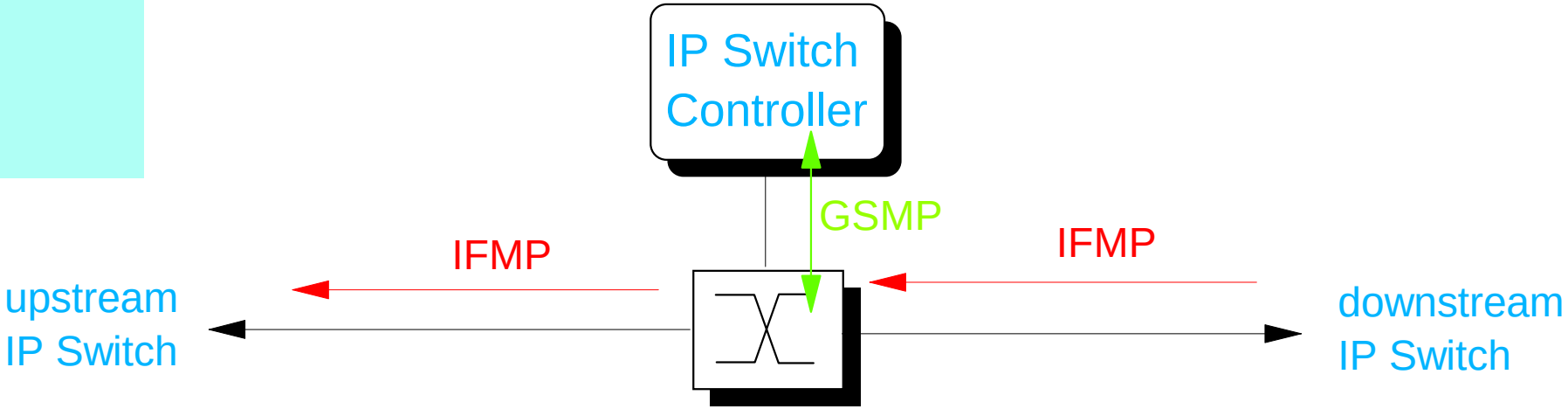
- MPC issues resolution request(src ATMaddr/IPaddr, dst IPaddr) to nearest MPS. This is forwarded over routed path until it reaches the egress MPS.
- MPOA Cache Imposition Reply/Request between egress MPS and MPC.
  - egress MPS collects information necessary to return NHRP reply to ingress MPS and MPC
- Egress MPS returns NHRP resolution reply(src ATMaddr/IPaddr, dst IPaddr/ATMaddr) which is then forwarded onto the ingress MPC.
- Ingress MPC establishes SVC to egress MPC

Solution: Ipsilon IP Switching



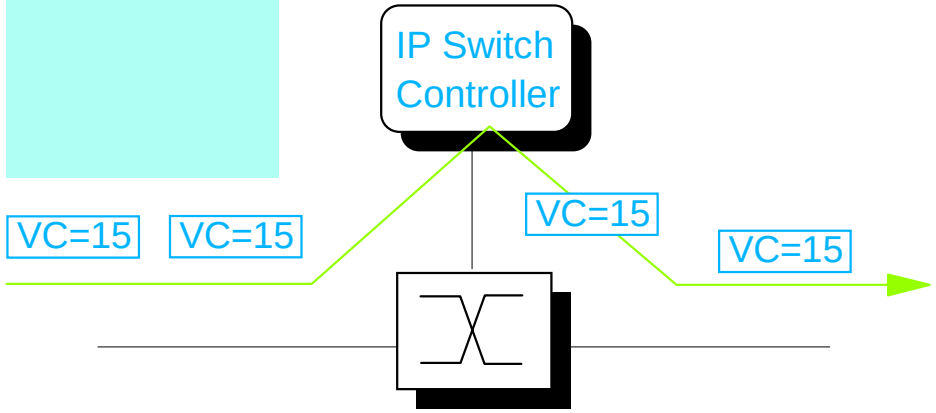
- Introduces an IP Switch (router on top of ATM hardware)
- IP Switch Controller
  - runs normal IP routing protocols and performs IP forwarding
  - identifies flows (based on policy, rate, etc.) that can be switched
  - exchanges flow-to-label bindings with neighbors and updates ATM switch label-swap tables
- Flow-driven

IFMP and GSMP

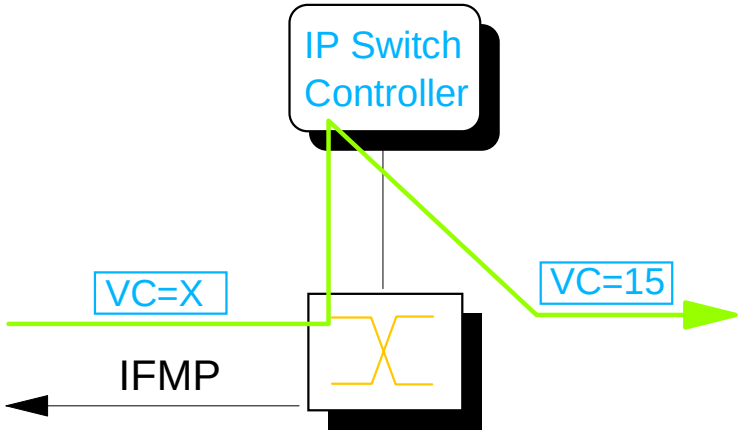


- IFMP - Ipsilon Flow Management Protocol, RFC1953/1954
  - instructs adjacent upstream node to re-label IP Flow with a new VPI/VCI for a specific period of time
- IFMP Redirect message contains:
  - Flow Type
  - Label - 32-bit field that contains VPI/VCI and is used to label an IP Flow
  - Flow Identifier - matches up with objects in IP header to identify an IP flow
  - Lifetime - period of time that redirection is maintained
- GSMP - General Switch Management Protocol, RFC1987
  - Used to manage switched resources
  - RFC1987

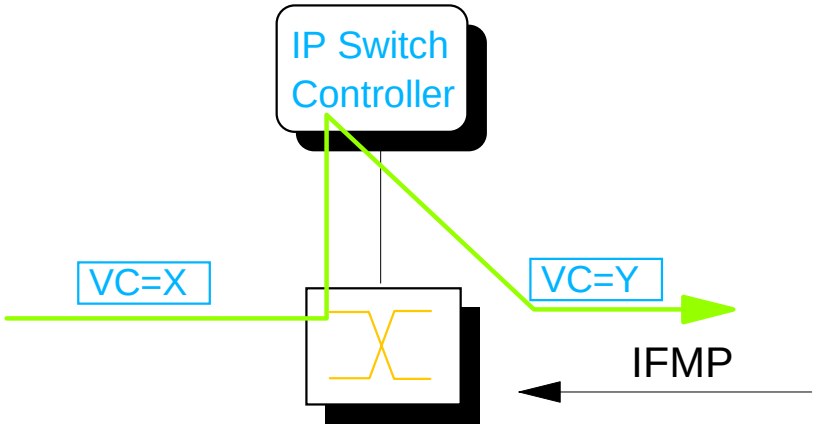
Ipsilon IP Switch Process



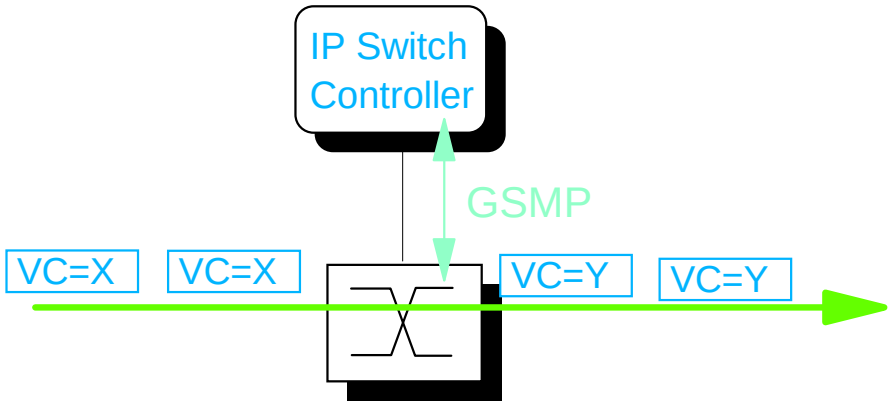
- Default Path thru IP Switch Controller
- IP Flow candidate for switching is identified



- IFMP Redirect (Flow ID, Label, Life) is sent upstream
- Cells switched from input port to output port upto router module (route information cached)



- Downstream IP Switch performs same function

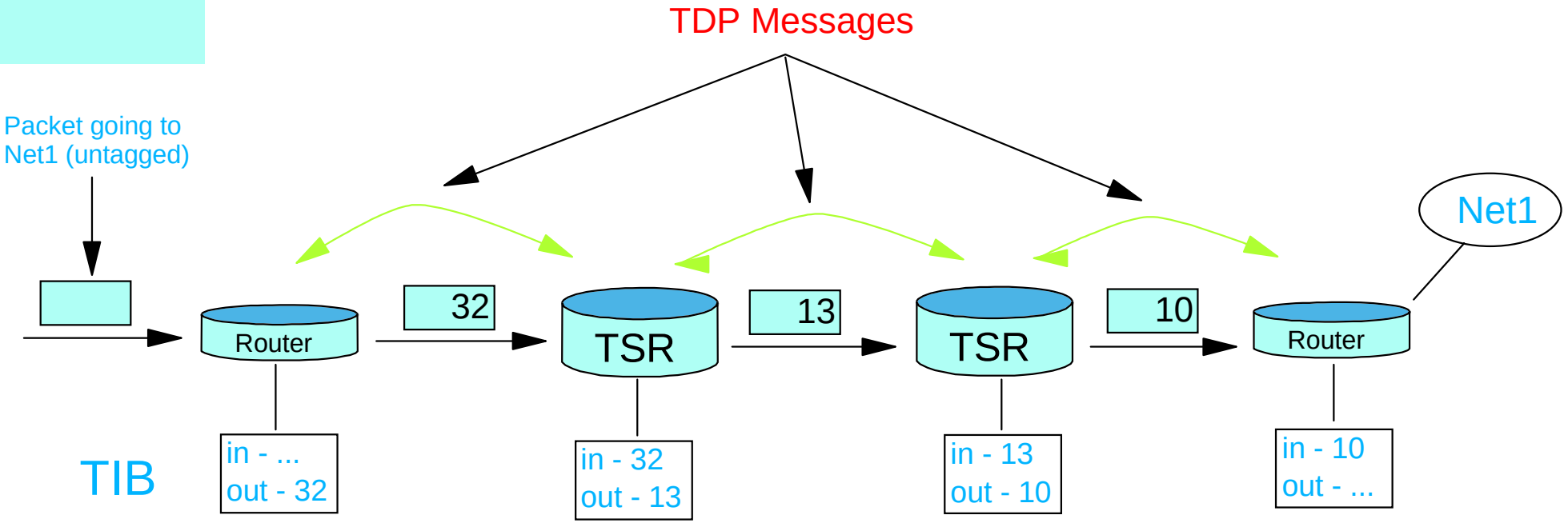


- GSMP maps input port to output port and IP Flow (cells) is now switched

Solution: Tag Switching (Cisco)

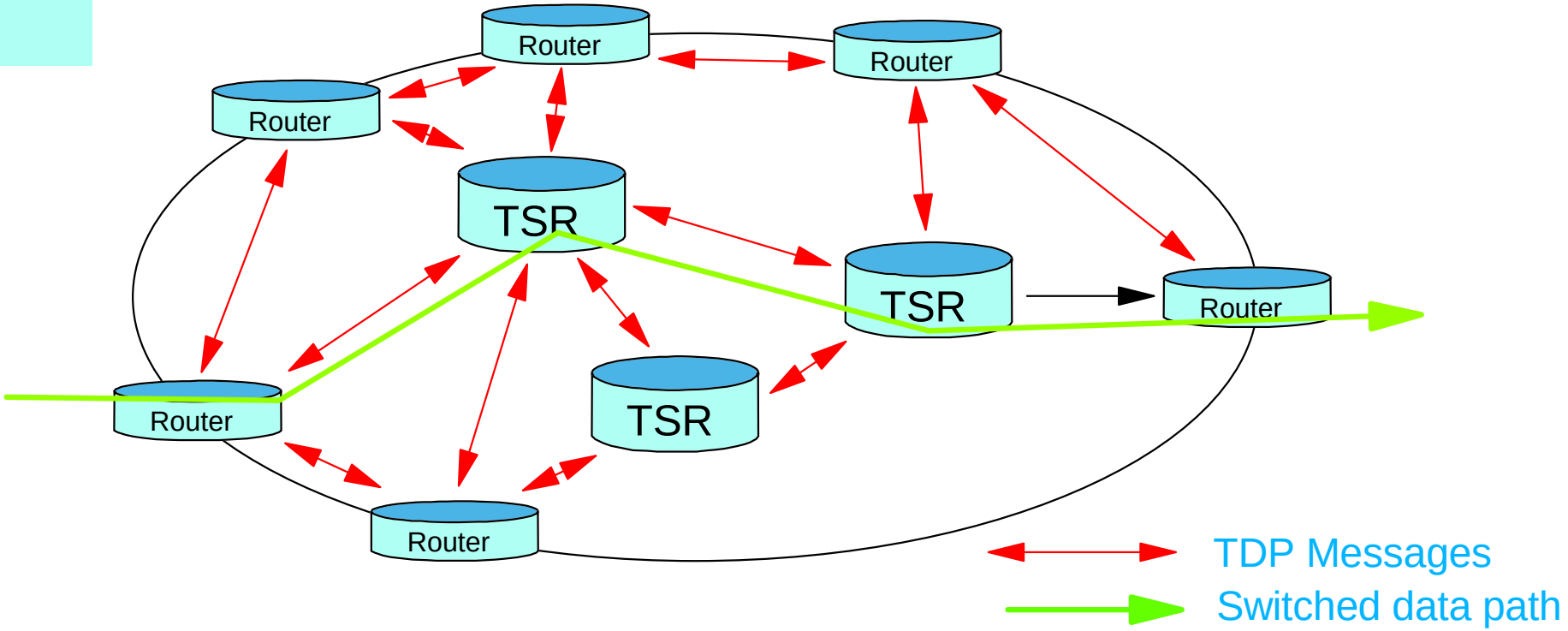
- Leverage layer-2 switching (label swapping) to improve IP packet forwarding performance
- Tag Switching Operation:
  - routes are mapped to tags. This is called a tag binding
  - Tag bindings are distributed among tag switch routers and stored in a tag information base (TIB) in the switch
  - packets/cells are marked with the appropriate tag and forwarded based on the contents of the tag
- Tag can be:
  - VPI/VCI field of an ATM cell, DLCI of a Frame Relay PDU
  - could be a "shim" layer between layer-3 and Mac addresses
- Topology-driven

Tag Switching



- Tag Distribution Protocol (TDP) communicates tag bindings (address prefix, Tag) between tag switch routers
  - tags bindings may also "piggyback" on other control protocols such as BGP, PIM, RSVP, etc.
- Packet is tagged at network ingress, switched to the egress, then tag is removed

Tag Switching Network

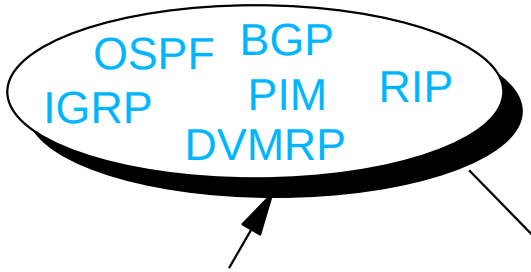


- Based on routing table entries, tag are generated and then advertised between adjacent TSRs via TDP
- Packets are tag switched from network ingress to egress

Solution: Aggregate Route-Based IP Switching (ARIS)

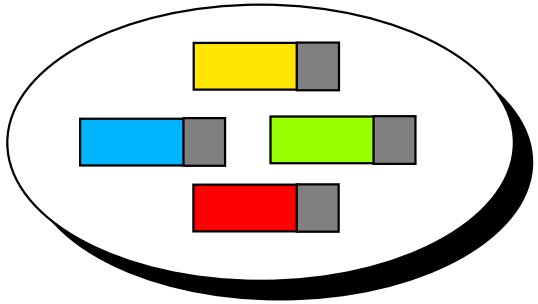
- Builds switched paths thru internetwork based on routing table information
- Switched path is VC "tree" rooted at the egress router and branching to all ingress routers
- Egress Identifier
  - defines a switched path to a one or more resources "behind" or associated with an Egress Node
  - enables many routing table entries to be aggregated into a single identifier and thus a single switched path
- Supports VP/VC merging
- Supports establishment of loop-free switched paths
- Topology-driven

Idea behind ARIS

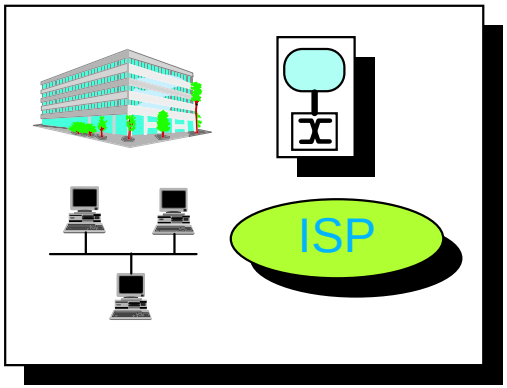




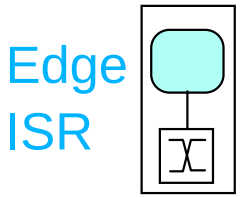
IP packets...



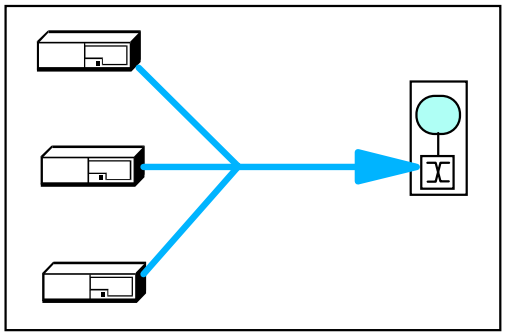
...going to a common destination...



..are associated with an identifier on an edge node...

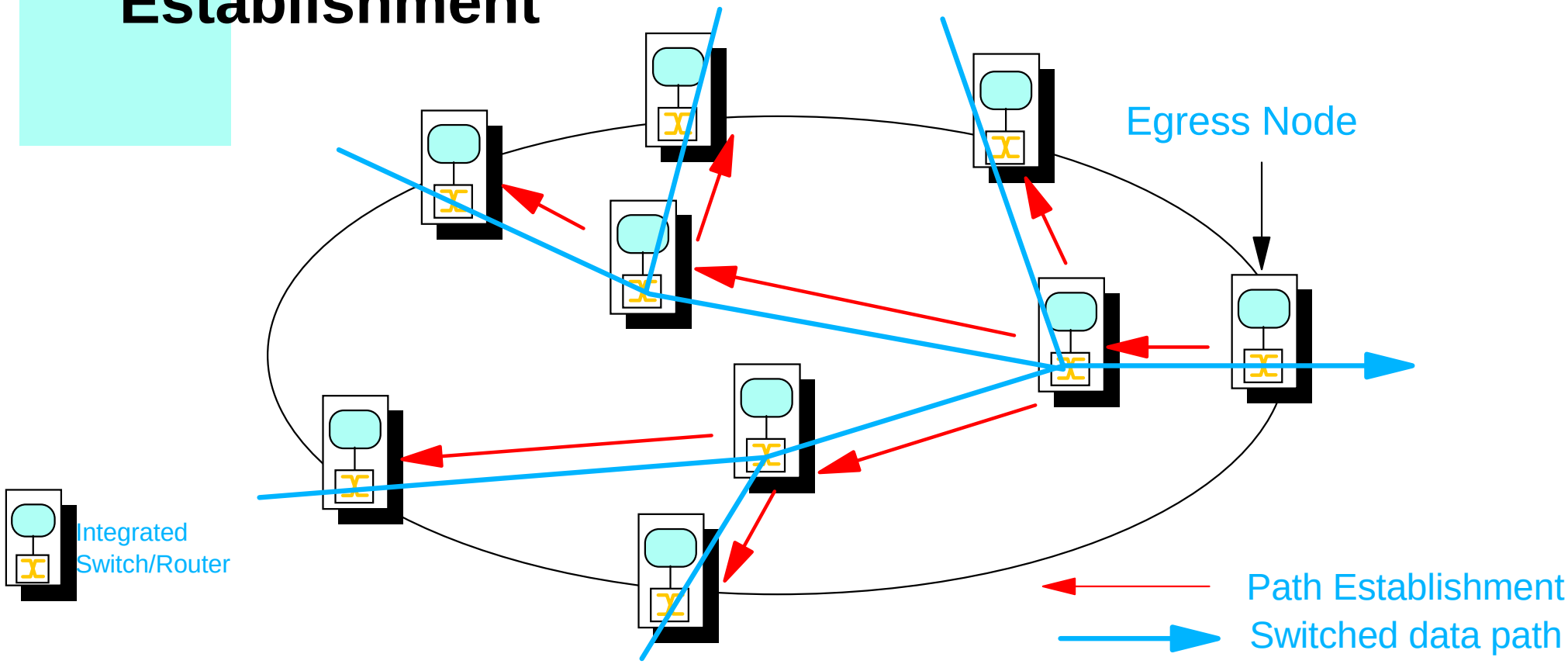


- **Observe that:**
  - ALL traffic going to a common destination flows on a single switched path
  - Edge identifier can be dynamically or manually configured on a single node



...which is then used to establish and identify a single switched path rooted at the edge.

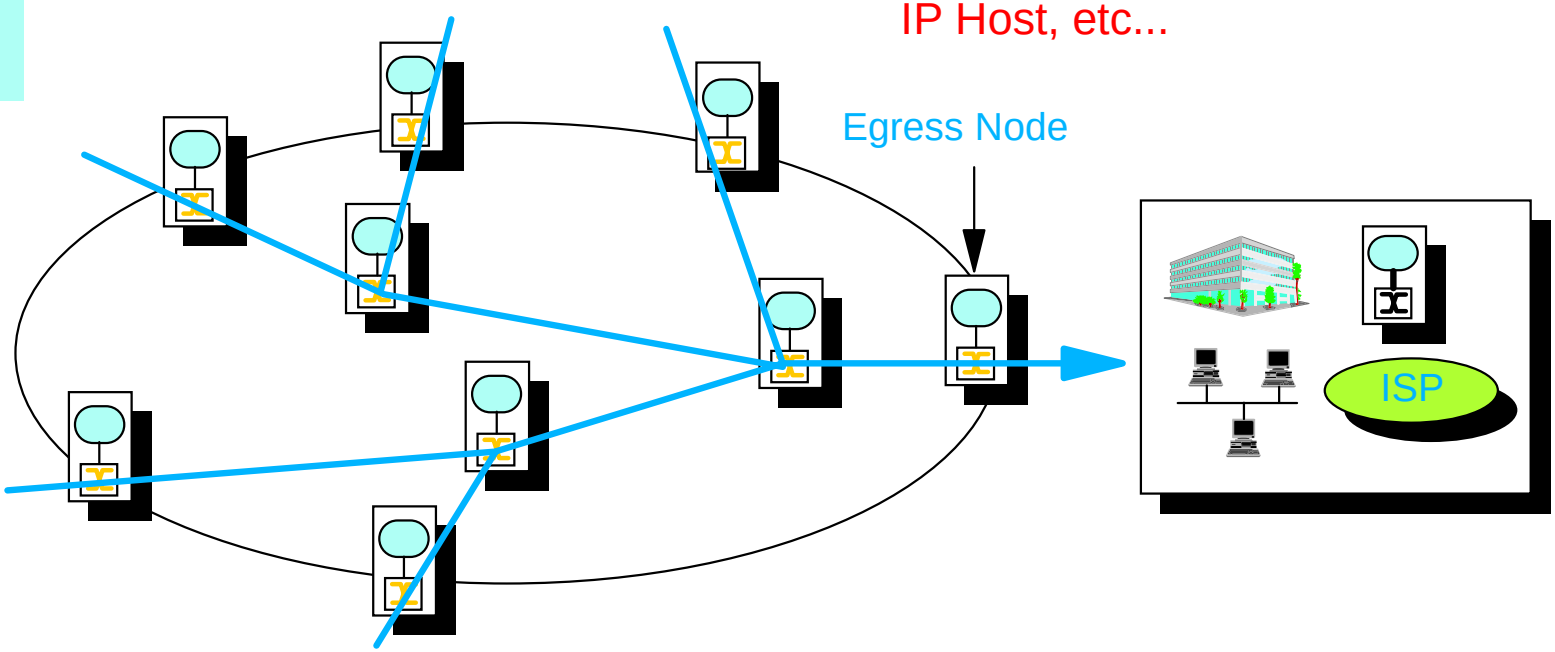
ARIS Switched Path Establishment



- ESTABLISH messages flow RPM-style from egress to all ingresses
- Routing (Egress ID)-Label mappings installed at each ISR along the way
- Switched paths follow the IP forwarding path
- Switched paths to an egress node form a tree rooted at the egress node

ARIS Egress Identifiers

Egress IDs can be associated with a CIDR prefix, IP subnet, Egress node, IP Host, etc...

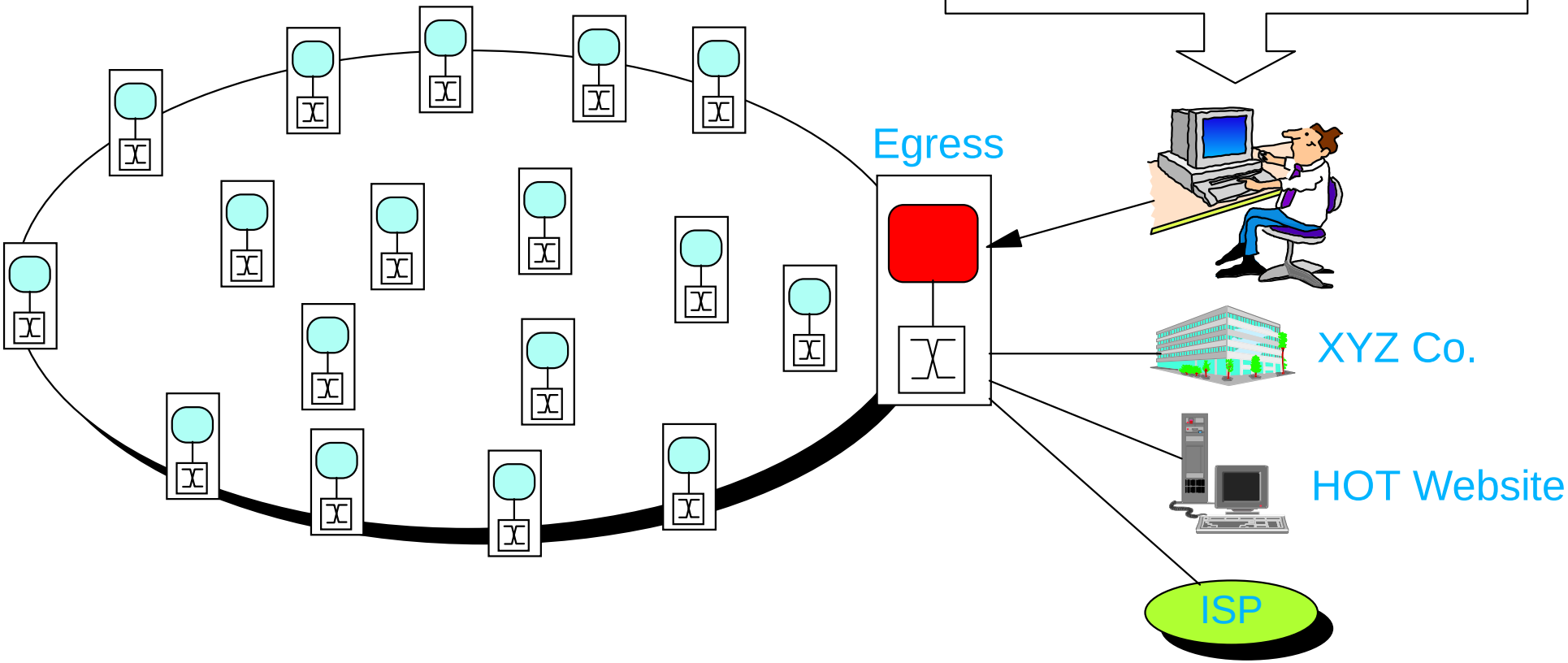


- Egress IDs are used to associate a group of packets that share a common set of routes with a single layer-2 switched path
  - the route can be a host route, a subnet, group of subnets, or something more abstract like an egress node itself, etc.
- Derived from routing protocols or manually configured (at the egress node)
- One switched path (mpt-pt tree) per egress ID; egress node can have multiple egress IDs

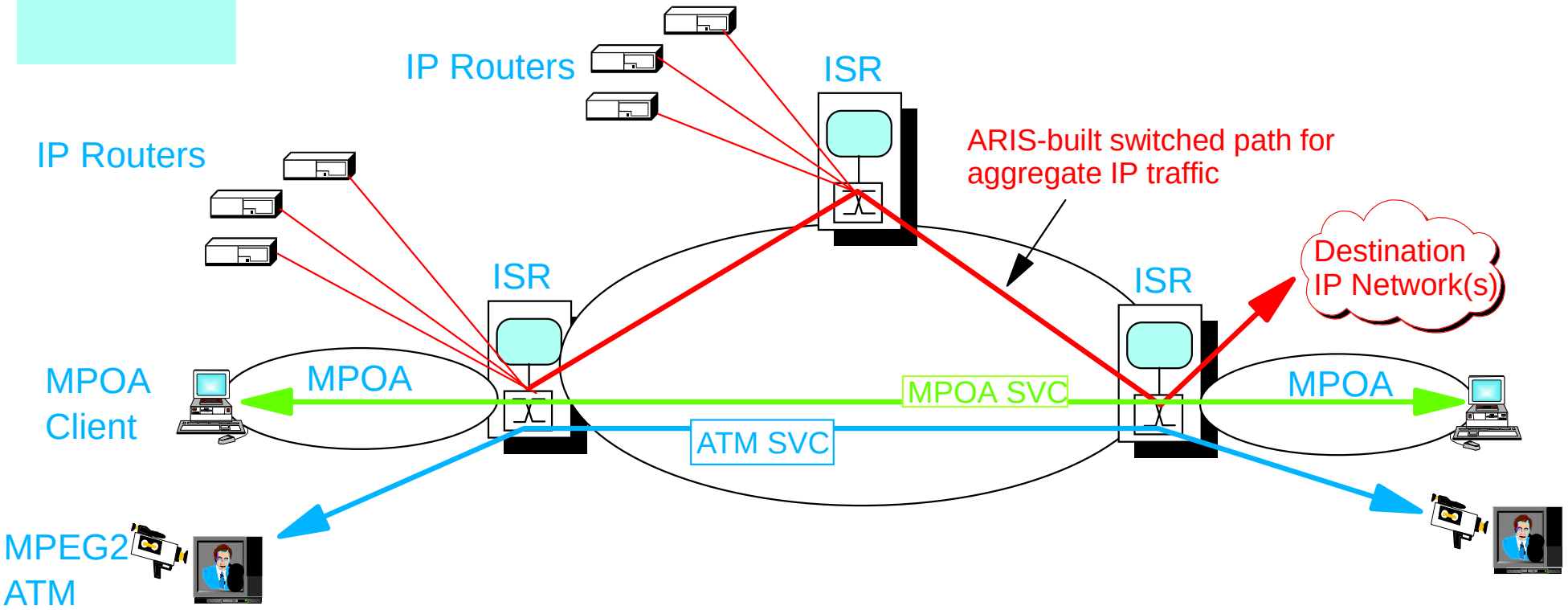
ARIS switched paths are administered from the Edge (egress) of the Network

network administrators manage provisioning of resources from the edge (egress) - not the ENTIRE network...

- No switched path for XYZ...
- Lotsa hits on the HOT website...
- Wanna be a good Internet Service Provider...



ARIS can coexist with native ATM services



- ISRs run Q.2931/PNNI in a "ships-in-the-night" configuration
- Enables Network Providers to offer IP and ATM services

ARIS

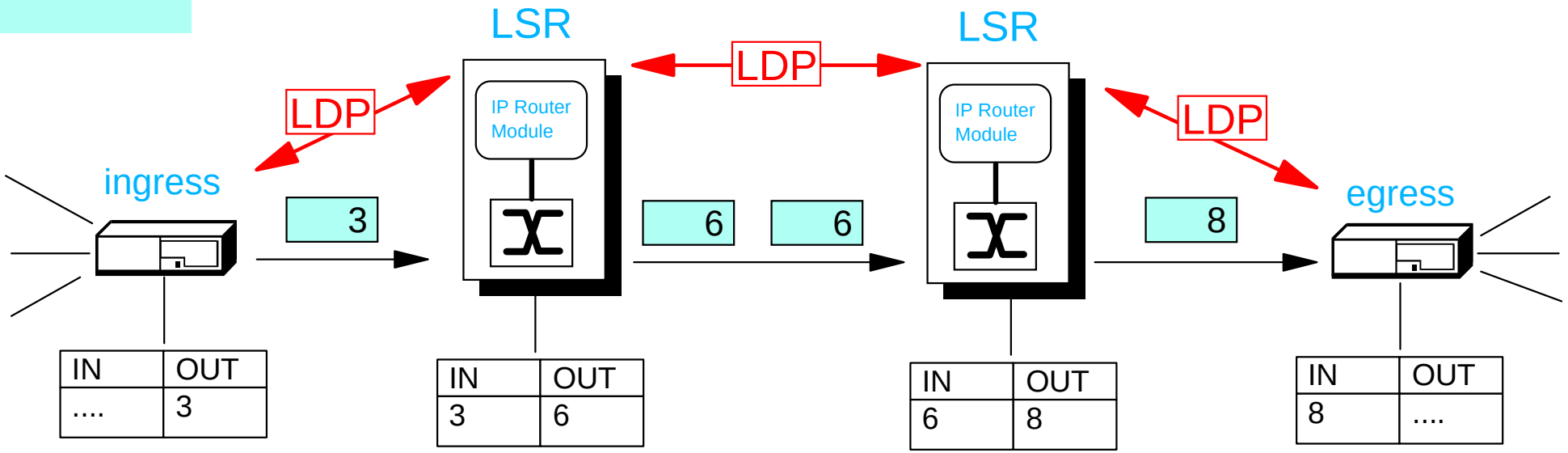
- Enables switched path establishment from network ingress(es) to egress based on:
  - destination network(s) or hosts
  - Multicast (DVMRP, PIM, etc.)
  - RSVP/QoS
  - Explicit Routes
  - Tunnels
  - etc.
- Promotes scalability
  - VP/VC merging requires O(N) switched paths
  - Egress aggregation - all traffic flowing to a common egress can share the same multipoint-to-point tree
- Edge (Egress) control simplifies network administration and policy setting
- Coexists with native ATM services

MPLS

- Multiprotocol Label Switching
- New Working Group in IETF chartered to develop solution that integrates network layer routing with layer-2 label swapping
- Objectives:
  - improve price, performance and scalability of IP routing
  - Enable and simplify integration of new IP-layer services without changing forwarding paradigm
- IBM Participation:
  - co-chair the MPLS Working Group
  - submitted ARIS as a solution for MPLS

MPLS Operation

LSR=Label Switch Router

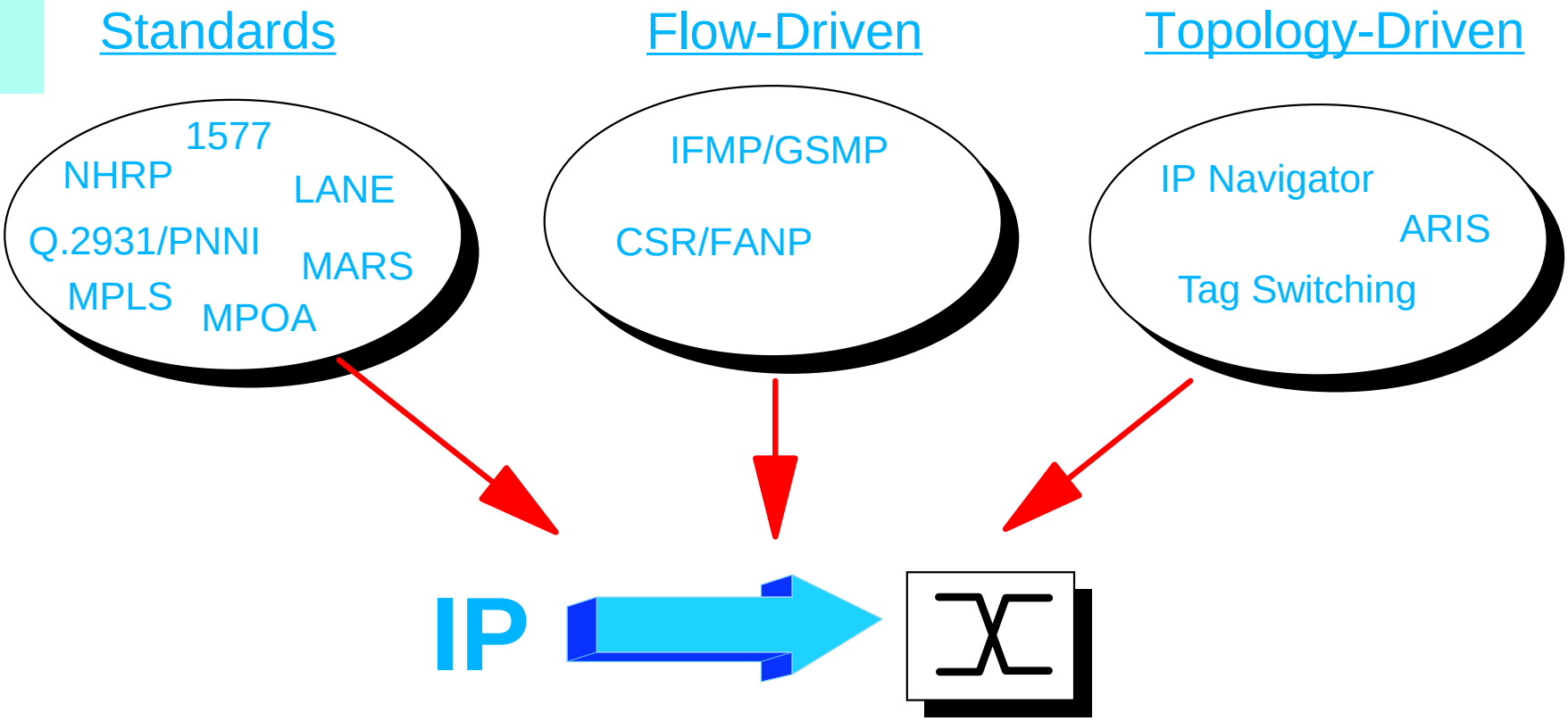


- Use Label Distribution Protocol (LDP) to distribute routing/label mappings to LSRs along routed path
- Classify and label at the ingress
- Label swap over layer-2 path to the egress
- Remove label and forward at the egress

IBM IP Switching Efforts

- MSS
  - SuperELAN, LANE and NHRP shipping today
  - NHRP extended for support on LANs and ELANs
  - positioned for MPOA
- PNNI Phase I shipping on 8260/8265 ATM switch
- GSMP agent and extensions implemented on 8265
- Integrated Switch/Router (ISR)
  - pre-standard MPLS compliant (ARIS)
  - Built on MSS/8265 platform
  - will leverage IBM strengths in ATM signaling/routing
- MPLS
  - ARIS Proposal
  - co-chair working group

The Goal of the many IP Switching Efforts



Support IP internetworking services on top of hardware-based switching

Summary

- IP Switching technology is needed in the Internet and corporate intranets
  - scalability
  - price/performance
  - special services
- Flow vs Topology
  - Flow-driven solutions are suitable for the campus LAN and corporate intranets
  - Topology-driven solutions are designed for large intranets and the Internet
- IP Switching solutions will work over other frame-based switching technologies
  - e.g. Frame Relay, Gigabit Ethernet, IEEE 802.1p switches, etc.
- Implementations differ based on coexistence and interoperability with existing standards

References

- ARIS Documentation
  - http://www.networking.ibm.com/nsw/aris.html
- "MSS and IP Switching", whitepaper, Chris Metz, 5/97
- "Aggregate Route-Based IP Switching", whitepaper, Chris Metz, 8/97
- Multi-Layer Switching Web-site
  - http://www.csl.sony.co.jp/person/demizu/inet/mlr.html
- MPLS
  - draft-ietf-mpls-framework-01.txt
  - draft-rosen-mpls-arch-00.txt
  - http://www.employees.org:80/~mpls/

References

- Ipsilon IP Switching
  - http://www.ipsilon.com/index.htm
- Tag Switching
  - http://www.cisco.com/warp/public/732/tag/index.html
- MPOA
  - ftp://ftp.atmforum.com/pub/approved-specs/af-mpoa-0087.000.pdf
- "Performance Issues in VC-Merge Capable MPLS Switches", Indra Widjaja and Anwar Elwalid
  - draft-widjaja-mpls-vc-merge-00.txt
- Sackett and Metz, "ATM and Multiprotocol Networking", McGraw-Hill, 1997
- Metz, C., "IP Switching: Protocols and Architectures", McGraw-Hill, 1998

Acronyms

- ARIS - Aggregate Route-Based IP Switching
- GSMP - General Switch Management Protocol
- IFMP - Ipsilon Flow Management Protocol
- ISP - Internet Service Provider
- ISR - Integrated Switch Router
- LANE - ATM LAN Emulation
- LDP - Label Distribution Protocol
- LSR - Label Switch Router
- MPLS - Multiprotocol Label Swapping
- MPOA - Multiprotocol over ATM
- MSS - Multiprotocol Switched Services
- NHRP - Next Hop Resolution Protocol
- OSPF - Open Shortest Path First
- PNNI - Private Network-to-Network Interface
- SVC - switched virtual connection
- TDP - Tag Distribution Protocol
- TSR - Tag Switch Router
- VC - Virtual Connection
- VP - Virtual Path