The Way to IP Switching

Scalability Issues and Comparison of Traditional Solutions
Ipsilon's IP Switching, Cisco's Tag Switching

Agenda

- **Comparison of IP over ATM methods**
  - static address resolution
  - ARP server
  - MARS and MCS
  - LANE
  - NHRP
- **IP Switching**
- **Tag Switching**

© 2005, D.I. Manfred Lindner

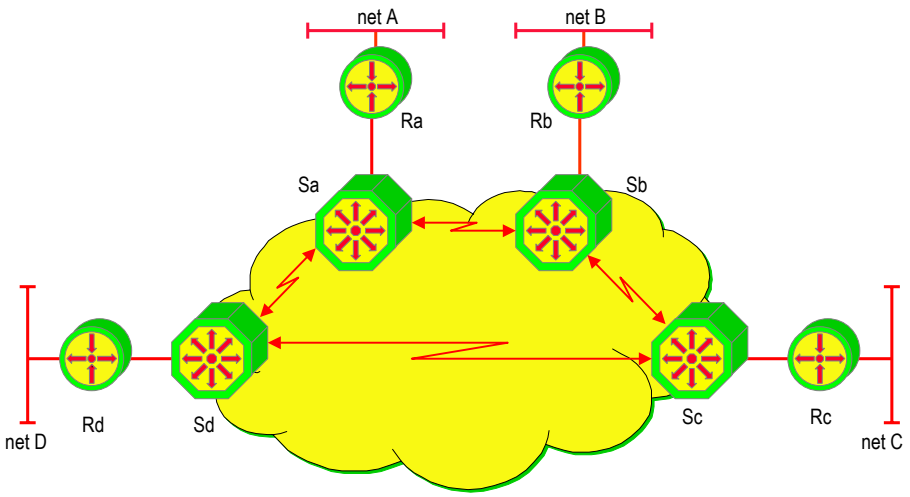
© 2005, D.I. Manfred Lindner

IP over ATM

- **IP routers connected to ATM switches**
  - virtual circuits of the ATM network are used as point-to-point links for IP transport
    - overlay technique of IP
- **ATM network is not a broadcast medium**
  - LAN-like address resolution is not possible
  - LAN-like routing updates are not possible
  - Non Broadcast Multiple Access (NBMA) network
    - for IP a NBMA situation is given if the physical ATM interfaces or the logical ATM sub-interfaces of all routers are in the same IP subnet
    - if all logical ATM sub-interface pairs are in different IP subnets then the ATM network appears as a number of conventional point-to-point interfaces to IP

© 2005, D.I. Manfred Lindner

Physical Topology for Comparison



© 2005, D.I. Manfred Lindner

© 2005, D.I. Manfred Lindner

Possible Solutions for IP Transport

- **several ways to solve address resolution and routing aspects:**
  - full-mesh VC´s
    - with static address mapping/static routing on PVC´s or SVC´s
  - full-mesh VC´s
    - with static address mapping/dynamic routing on PVC´s
  - partial-mesh VC´s
    - with static address mapping/static routing on PVC´s or SVC´s
  - partial-mesh VC´s
    - with static address mapping/dynamic routing on PVC´s
  - on demand VC´s (ARP server)
    - with dynamic address resolution/static routing on SVC´s

© 2005, D.I. Manfred Lindner

Solutions (cont.)

  - on demand VC´s (MARS and MCS)
    - with dynamic address resolution/dynamic routing on point-to-multipoint distribution circuits
  - LANE emulation
  - on demand VC´s (NHRP)
    - with dynamic routing hop by hop and establshing of a cut-through circuits on demand
  - MPOA

© 2005, D.I. Manfred Lindner

© 2005, D.I. Manfred Lindner

Agenda

- **Comparison of IP over ATM methods**
  - static address resolution
  - ARP server
  - MARS and MCS
  - LANE
  - NHRP
- **IP Switching**
- **Tag Switching**

© 2005, D.I. Manfred Lindner

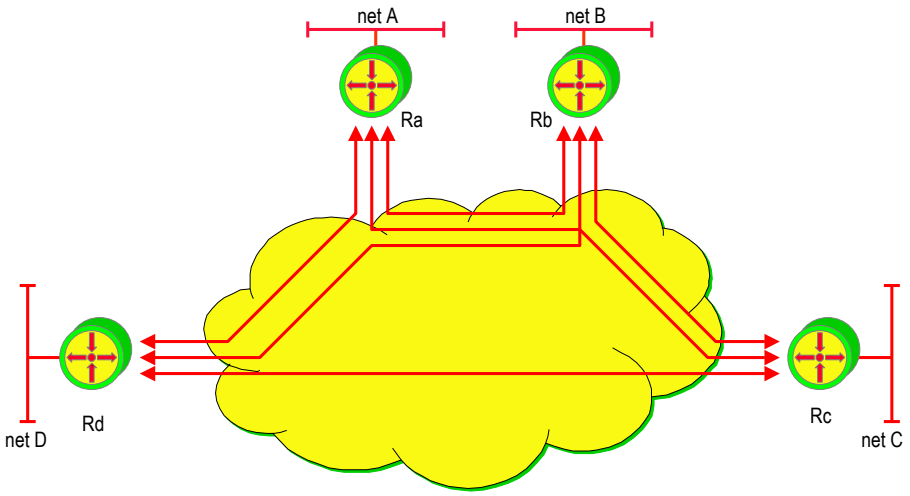
IP Connectivity through Full-mesh VC´s

- **one way to achieve full connectivity between all IP networks**
  - full mesh of ATM virtual circuits
- **address resolution**
  - static mapping ⇨ administrative overhead
  - does not scale well
- **routing**
  - static routing ⇨ administrative overhead
  - dynamic routing ⇨ bandwidth problem
    - duplicates of broadcasts are sent out by a router on every VC
  - dynamic routing ⇨ discovery of neighbors problem
    - broadcasts can be sent on PVC´s only

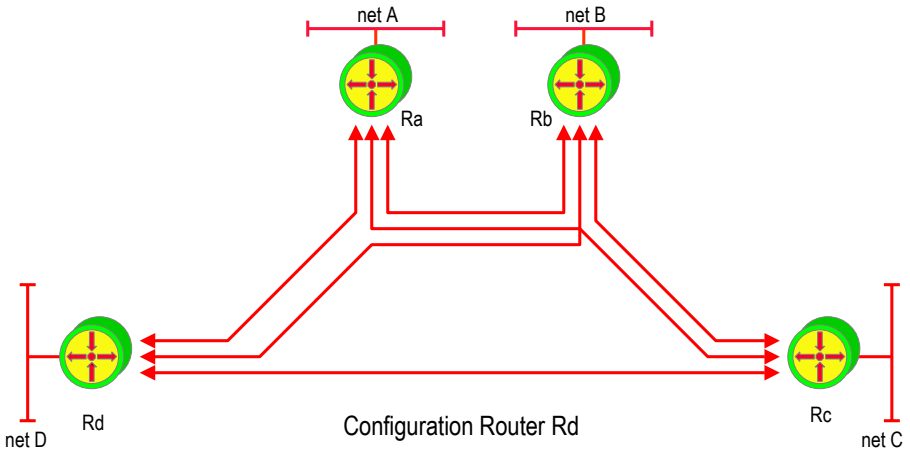
© 2005, D.I. Manfred Lindner

© 2005, D.I. Manfred Lindner

IP Connectivity through Full-mesh VC´s



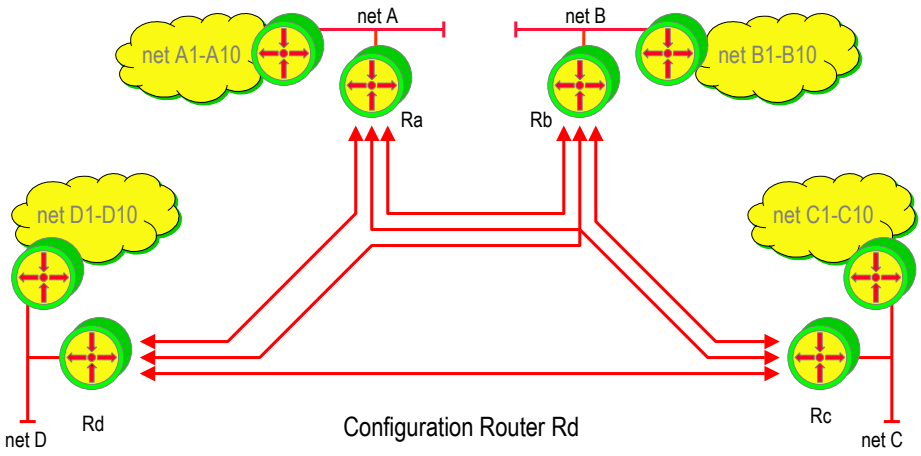
Static Routing/No Routing Broadcasts



static routing	address resolution PVC	address resolution SVC
net A via next hopRa	Ra map VPI/VCI Rd ⇨ Ra	Ra map ATM addr. Ra
net B via next hopRb	Rb map VPI/VCI Rd ⇨ Rb	Rb map ATM addr. Rb
net C via next hopRc	Rc map VPI/VCI Rd ⇨ Rc	Rc map ATM addr. Rc

© 2005, D.I. Manfred Lindner

Static Routing/No Routing Broadcasts



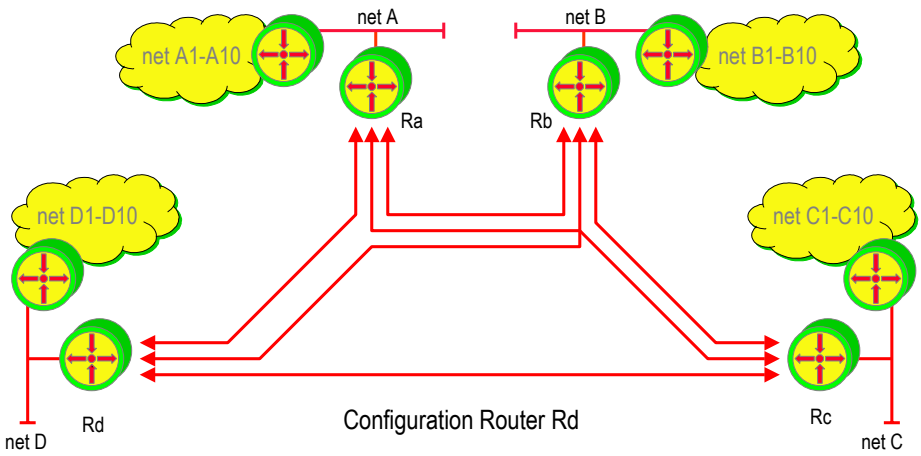
static routing
net A via next hopRa
net B via next hopRb
net C via next hopRc
every remote network listed here!

address resolution PVC
Ra map VPI/VCI Rd ⇨ Ra
Rb map VPI/VCI Rd ⇨ Rb
Rc map VPI/VCI Rd ⇨ Rc

address resolution SVC
Ra map ATM addr. Ra
Rb map ATM addr. Rb
Rc map ATM addr. Rc

© 2005, D.I. Manfred Lindner Way to IP Switching, v3.5 11

Dynamic Routing/Routing Broadcasts



dynamic routing on PVC
VPI/VCI Rd ⇨ Ra broadcast
VPI/VCI Rd ⇨ Rb broadcast
VPI/VCI Rd ⇨ Rc broadcast
note: SVCs are not possible because of discovery problem

address resolution PVC
Ra map VPI/VCI Rd ⇨ Ra
Rb map VPI/VCI Rd ⇨ Rb
Rc map VPI/VCI Rd ⇨ Rc

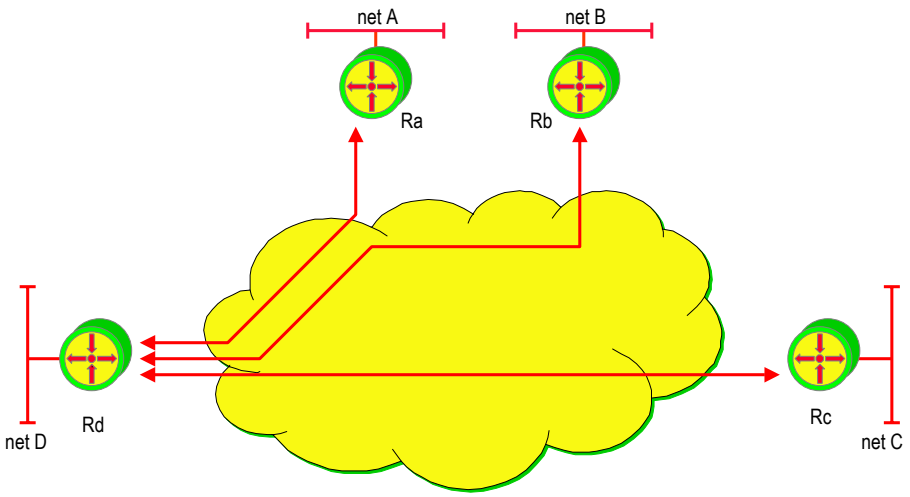
© 2005, D.I. Manfred Lindner Way to IP Switching, v3.5 12

© 2005, D.I. Manfred Lindner

IP Connectivity through Partial-mesh VC´s

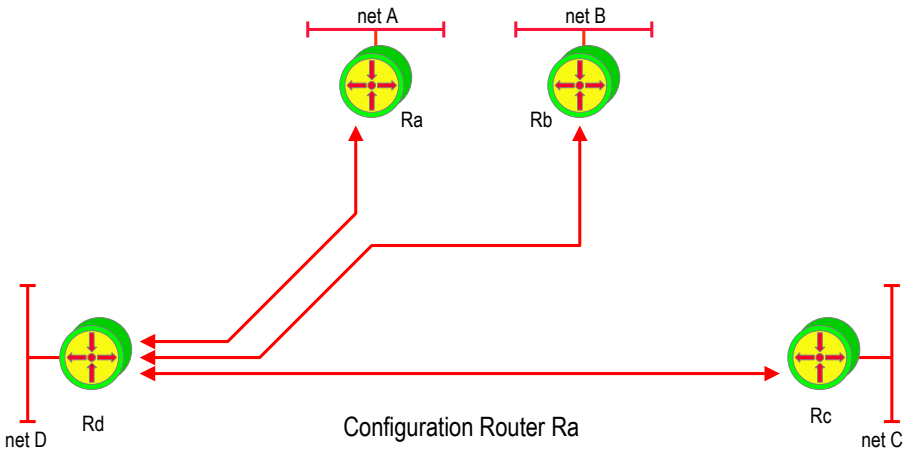
- **another way to achieve full connectivity between all IP networks**
  - partial mesh of ATM virtual circuits
  - hop by hop routing
    - means additional delay and SAR overhead on intermediate router
- **address resolution**
  - same problems as for full mesh VC´s
- **routing**
  - same problems as for full mesh VC´s
  - dynamic routing ⇨ additional split horizon problem
    - in case of NBMA and Distance Vector routing split horizon must be disabled

IP Connectivity through Partial-mesh VC´s



© 2005, D.I. Manfred Lindner

Static Routing/No Routing Broadcasts



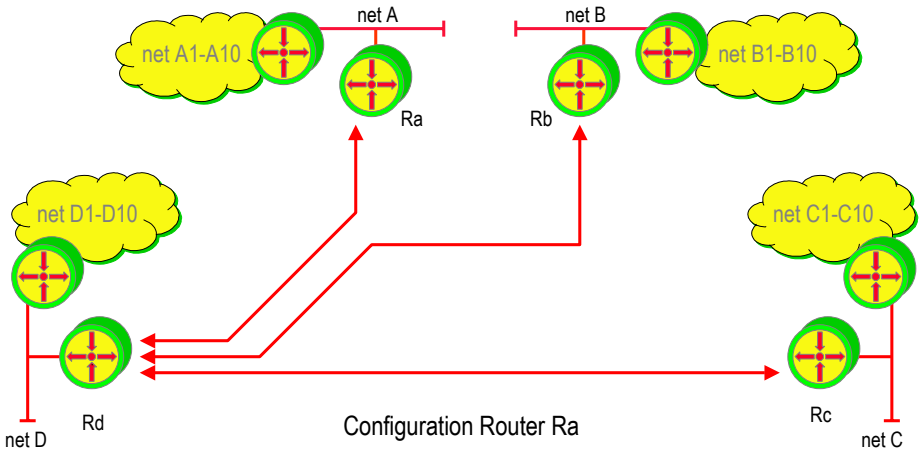
static routing
net D via next hopRd
net B via next hopRd
net C via next hopRd

address resolution PVC
Rd map VPI/VCI Ra ⇨ Rd

address resolution SVC
Rd map ATM addr. Rd

© 2005, D.I. Manfred Lindner Way to IP Switching, v3.5 15

Static Routing/No Routing Broadcasts



static routing
net A via next hopRd
net B via next hopRd
net C via next hopRd
every remote network listed here!

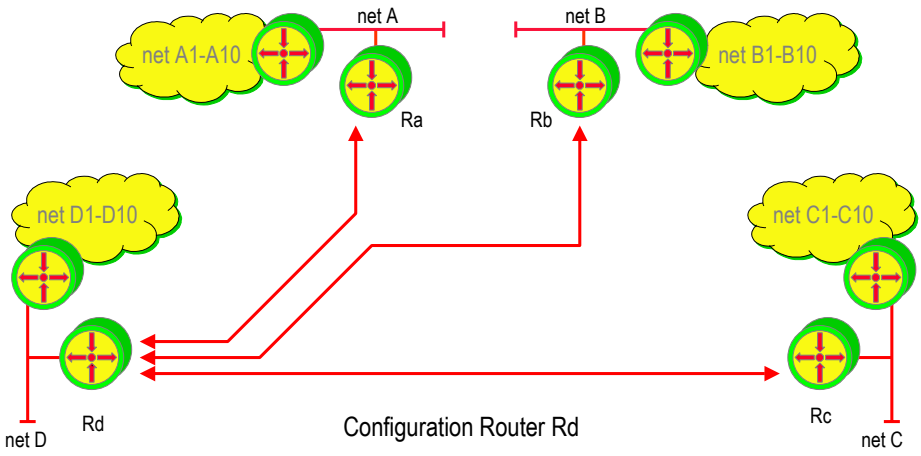
address resolution PVC
Rd map VPI/VCI Ra ⇨ Rd

address resolution SVC
Rd map ATM addr. Rd

© 2005, D.I. Manfred Lindner Way to IP Switching, v3.5 16

© 2005, D.I. Manfred Lindner

Static Routing/No Routing Broadcasts



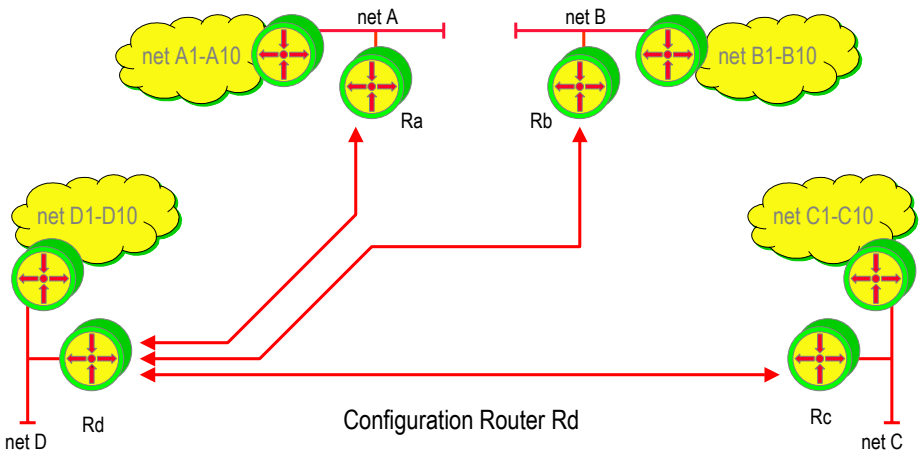
static routing
net A via next hopRa
net B via next hopRb
net C via next hopRc
every remote network listed here!

address resolution PVC
Ra map VPI/VCI Rd ⇨ Ra
Rb map VPI/VCI Rd ⇨ Rb
Rc map VPI/VCI Rd ⇨ Rc

address resolution SVC
Ra map ATM addr. Ra
Rb map ATM addr. Rb
Rc map ATM addr. Rc

© 2005, D.I. Manfred Lindner Way to IP Switching, v3.5 17

Dynamic Routing/Routing Broadcasts



dynamic routing on PVC
VPI/VCI Rd ⇨ Ra broadcast
VPI/VCI Rd ⇨ Rb broadcast
VPI/VCI Rd ⇨ Rc broadcast

address resolution PVC
Ra map VPI/VCI Rd ⇨ Ra
Rb map VPI/VCI Rd ⇨ Rb
Rc map VPI/VCI Rd ⇨ Rc

in case of NBMA: split horizon disabled! (note: SVCs are not possible because of discovery problem)

© 2005, D.I. Manfred Lindner Way to IP Switching, v3.5 18

© 2005, D.I. Manfred Lindner

Agenda

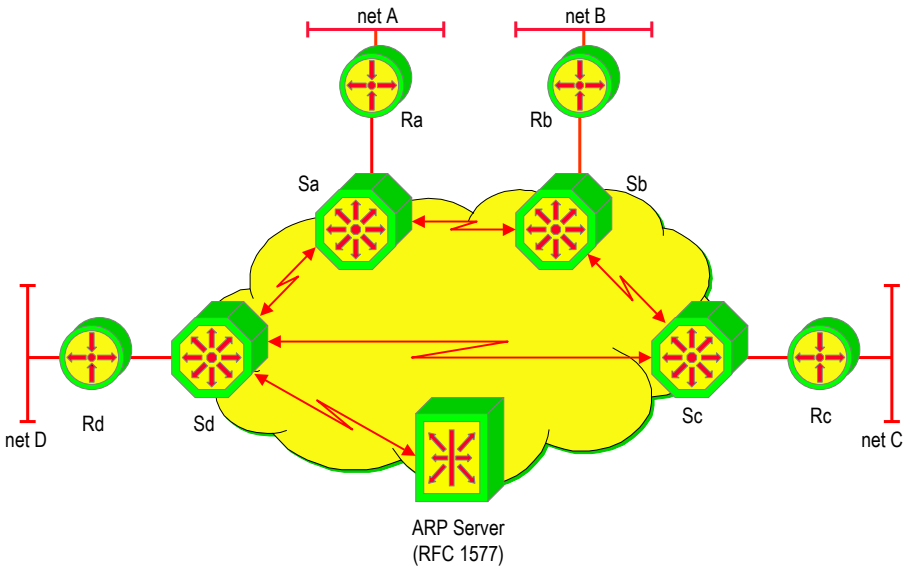
- **Comparison of IP over ATM methods**
  - static address resolution
  - ARP server
  - MARS and MCS
  - LANE
  - NHRP
- **IP Switching**
- **Tag Switching**

NBMA and ARP Server

- **a way to avoid static mapping of addresses**
  - every IP system within an IP subnet is configured with ATM address of ARP server
    - a PVC or SVC is used for the communication between
  - IP systems register themselves at ARP server
  - IP systems ask ARP server for ATM address of other already registered IP systems
  - SVC´ s for user traffic are established on demand
- **routing broadcast problem not addressed by ARP server**
  - static routing required
  - dynamic routing not possible because of discovery problem

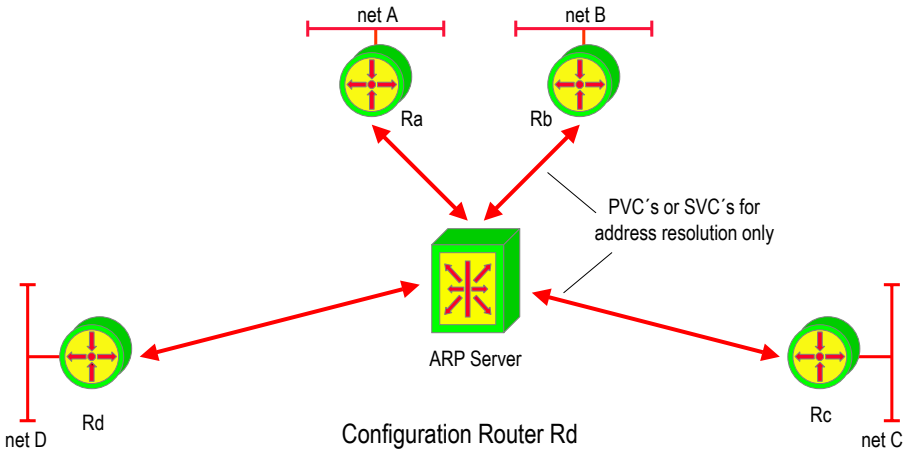
© 2005, D.I. Manfred Lindner

NBMA and ARP Server



© 2005, D.I. Manfred Lindner Way to IP Switching, v3.5 21

Dynamic Address Resolution in NBMA



Configuration Router Rd

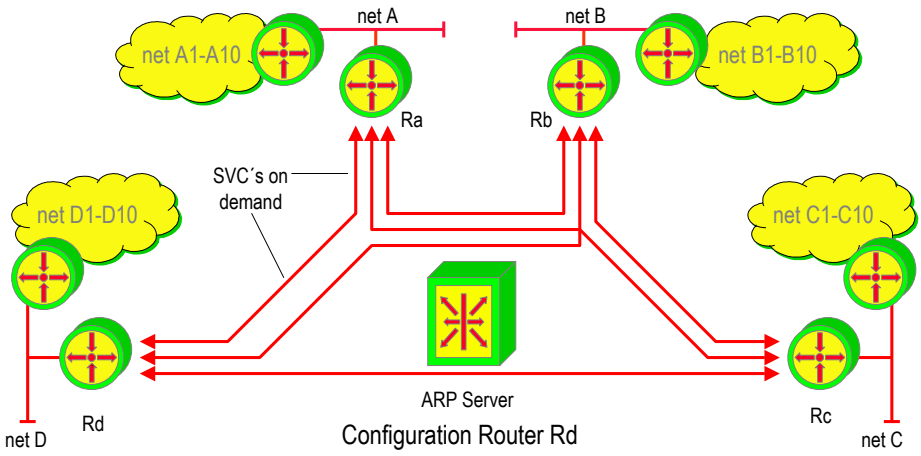
static routing on SVC
net A via next hopRa
net B via next hopRb
net C via next hopRc

address resolution configuration
ATM address of ARP Server

© 2005, D.I. Manfred Lindner Way to IP Switching, v3.5 22

© 2005, D.I. Manfred Lindner

Static Routing/SVC´s on Demand



static routing on SVC
net A via next hopRa
net B via next hopRb
net C via next hopRc
every remote network listed here!

address resolution configuration
ATM address of ARP Server

Agenda

- **Comparison of IP over ATM methods**
  - static address resolution
  - ARP server
  - MARS and MCS
  - LANE
  - NHRP
- **IP Switching**
- **Tag Switching**

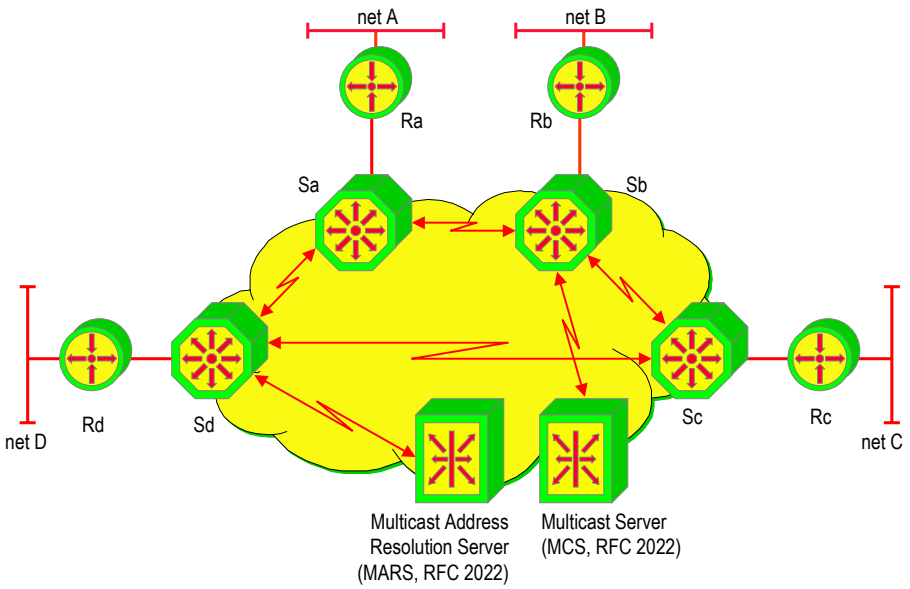
© 2005, D.I. Manfred Lindner

MARS/MCS

- **to enable multicasting and hence dynamic routing based on multicasts**
  - a multicast address resolution server (MARS) is used
    - resolving ATM destination addresses for a given layer 3 multicast group (broadcast)
  - a multicast server (MCS) is used
    - direct distribution using a mesh of ATM point-to-multipoint circuits
  - see details in corresponding RFC´s
- **SVC´s for user traffic are established on demand**

© 2005, D.I. Manfred Lindner

MARS/MCS Server

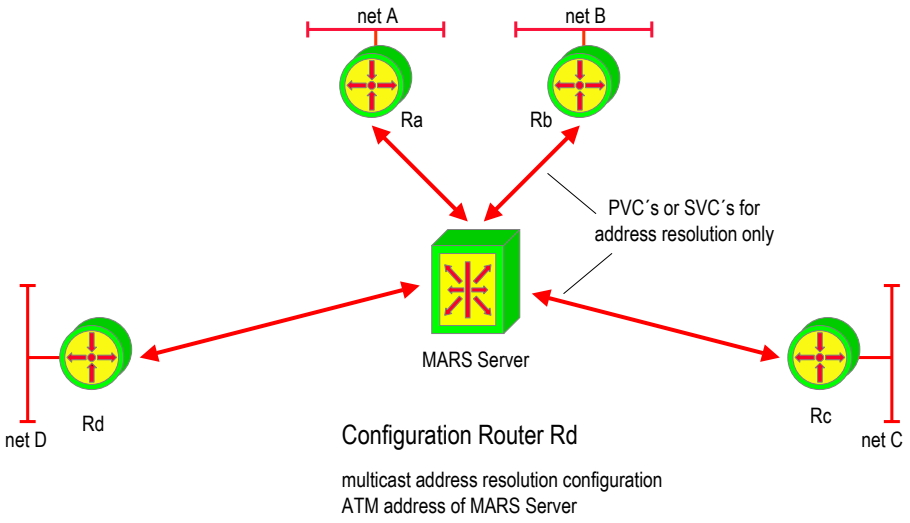


© 2005, D.I. Manfred Lindner

© 2005, D.I. Manfred Lindner

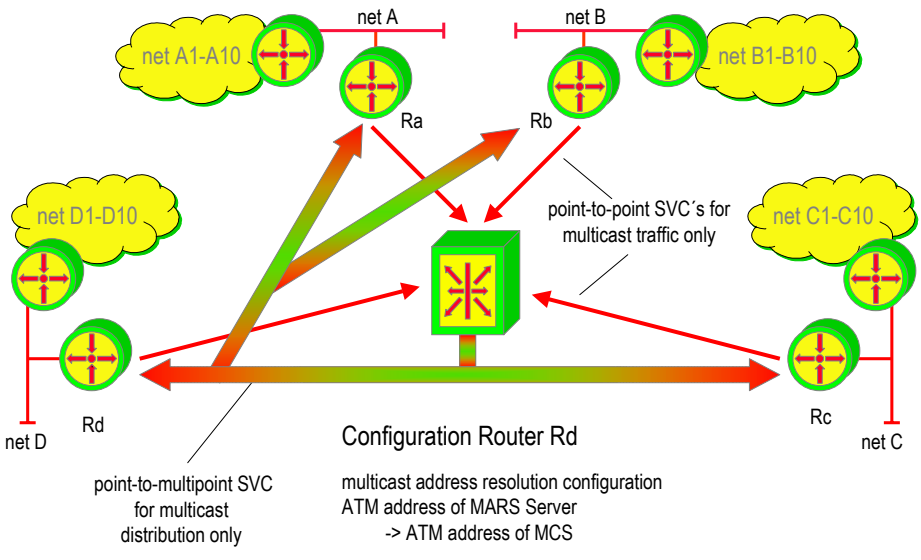
L83 - The Way to IP Switching

MARS



© 2005, D.I. Manfred Lindner

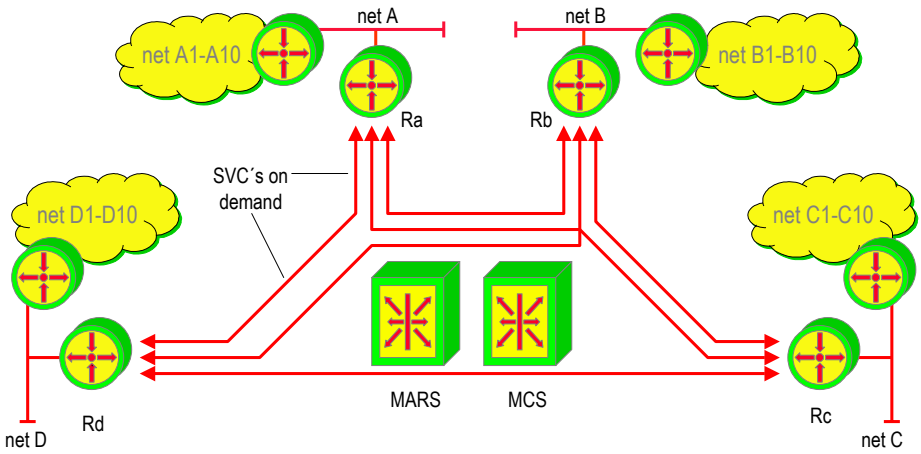
MCS



© 2005, D.I. Manfred Lindner

© 2005, D.I. Manfred Lindner

Dynamic Routing/SVC´s on Demand



Configuration Router Rd

multicast address resolution configuration
ATM address of MARS Server

© 2005, D.I. Manfred Lindner

Agenda

- **Comparison of IP over ATM methods**
  - static address resolution
  - ARP server
  - MARS and MCS
  - LANE
  - NHRP
- **IP Switching**
- **Tag Switching**

© 2005, D.I. Manfred Lindner

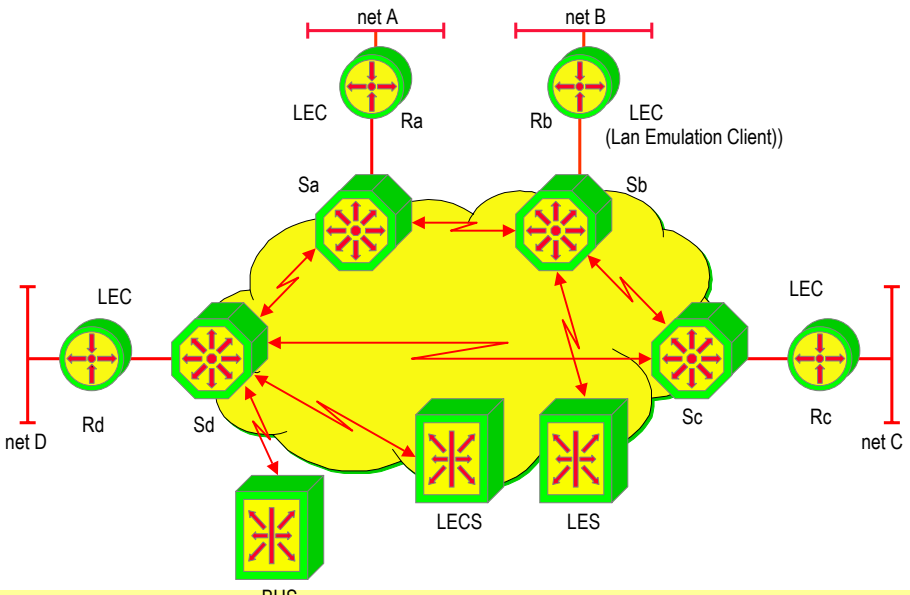
© 2005, D.I. Manfred Lindner

LANE

- **ATM looks like a LAN using MAC addresses for communication**
  - address resolution between MAC address and ATM address is done by LES
  - LEC is configured with ATM address of LECS in order to connect to a certain virtual LAN
  - Broadcasts (routing messages) are sent to BUS which distribute them using point-to-multipoint ATM circuits
  - IP-ARP used to resolve MAC-address
  - LANE-ARP used to resolve ATM-address
- **SVC´s for user traffic are established on demand**

© 2005, D.I. Manfred Lindner

ATM LAN Emulation

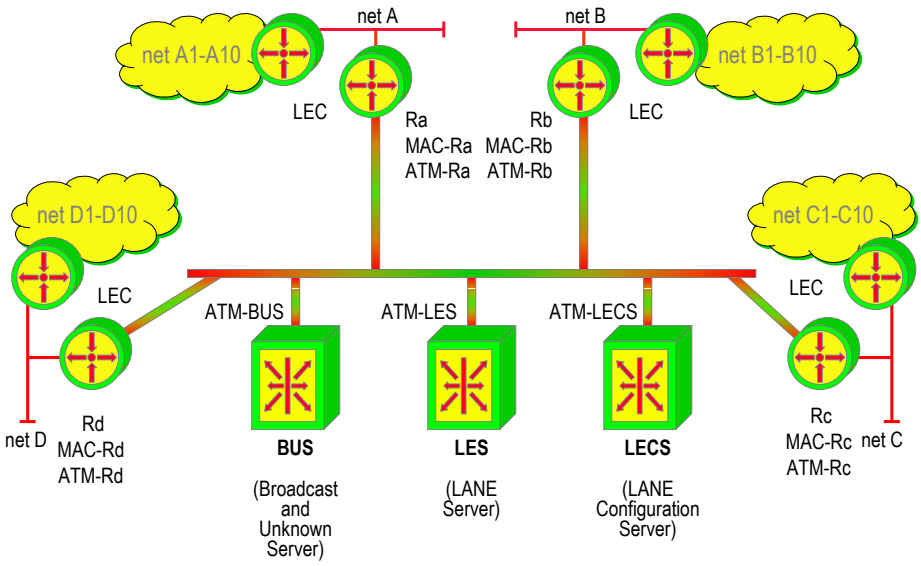


© 2005, D.I. Manfred Lindner

© 2005, D.I. Manfred Lindner

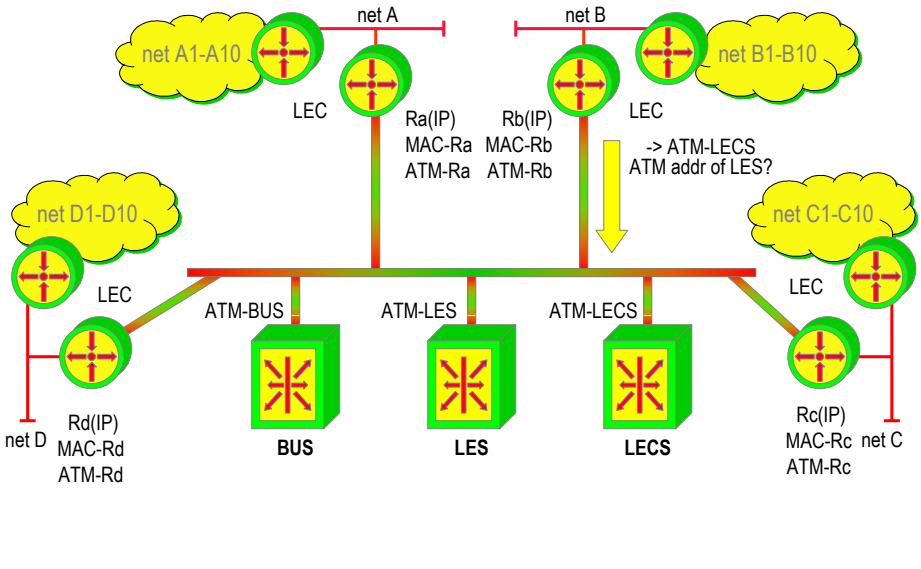
L83 - The Way to IP Switching

LANE



© 2005, D.I. Manfred Lindner

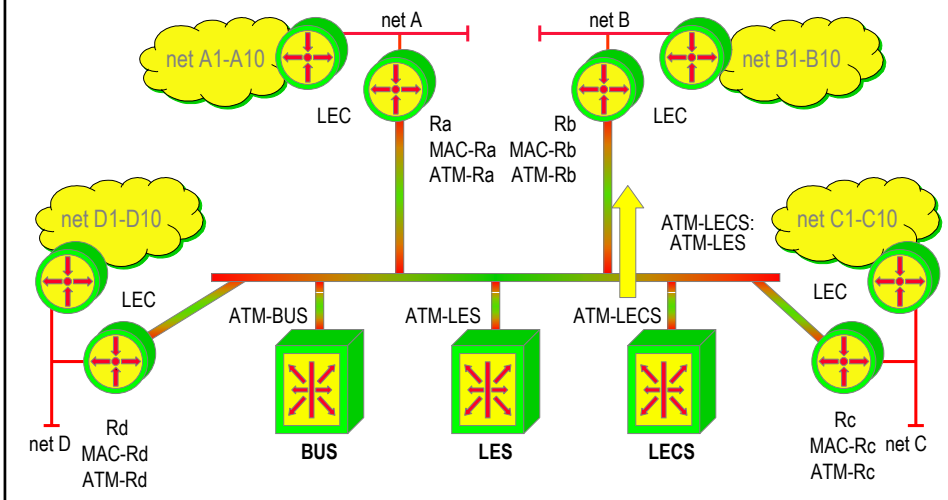
Registration LEC Rb 1



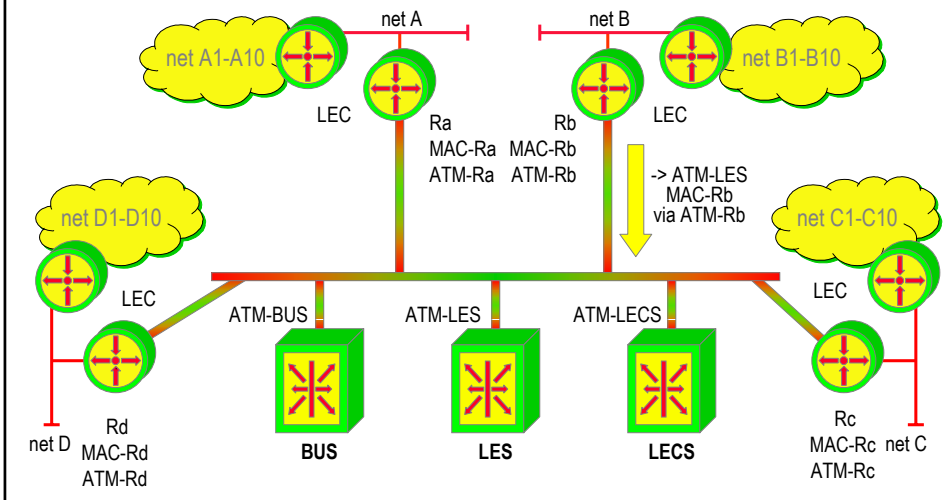
© 2005, D.I. Manfred Lindner

© 2005, D.I. Manfred Lindner

Registration LEC Rb 2

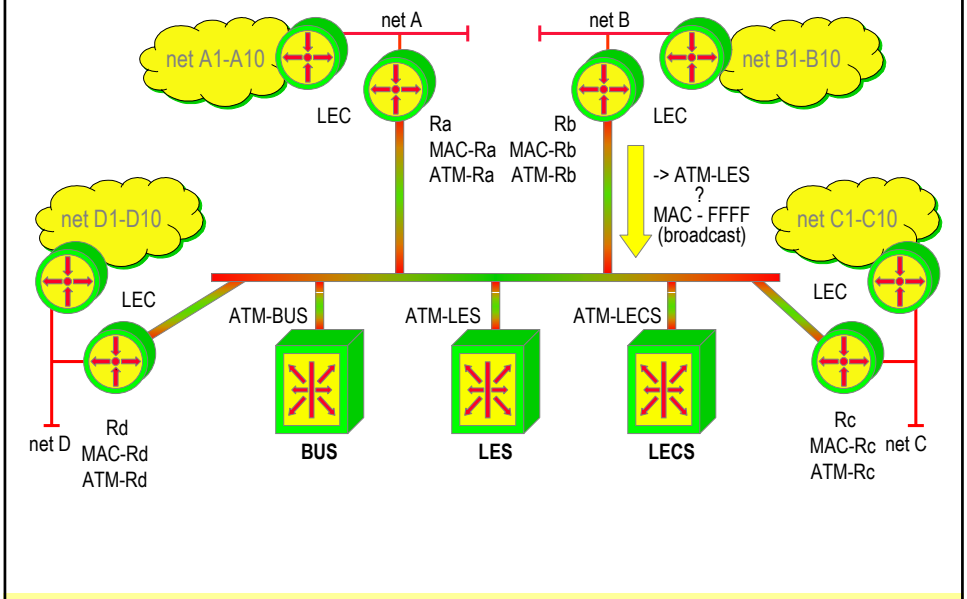


Registration LEC Rb 3

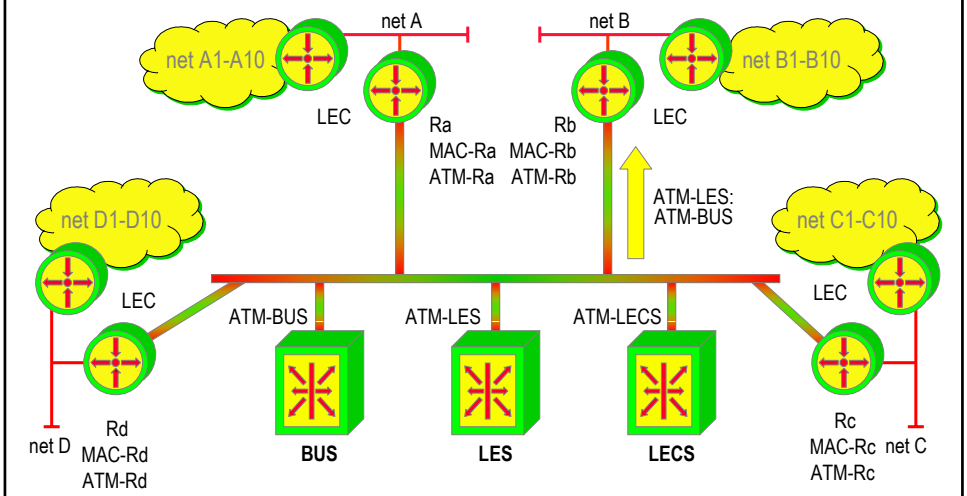


© 2005, D.I. Manfred Lindner

Registration LEC Rb 4

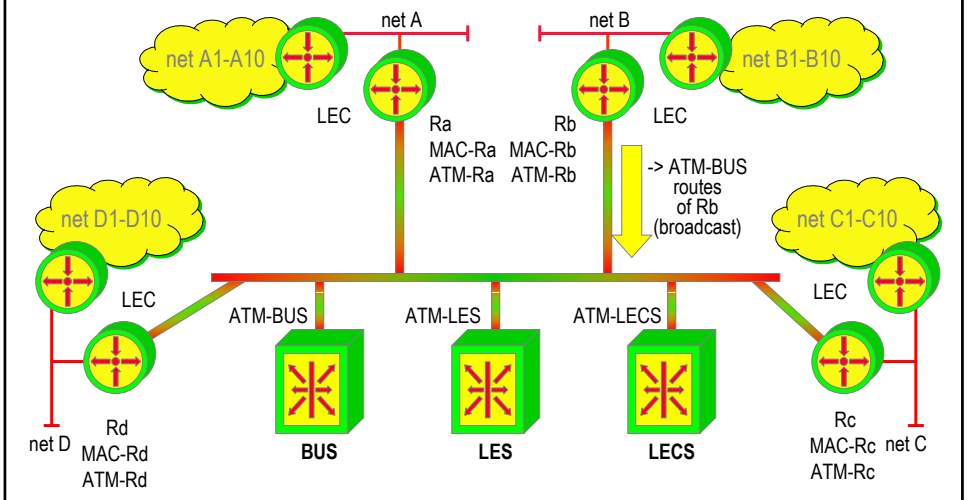


Registration LEC Rb 5



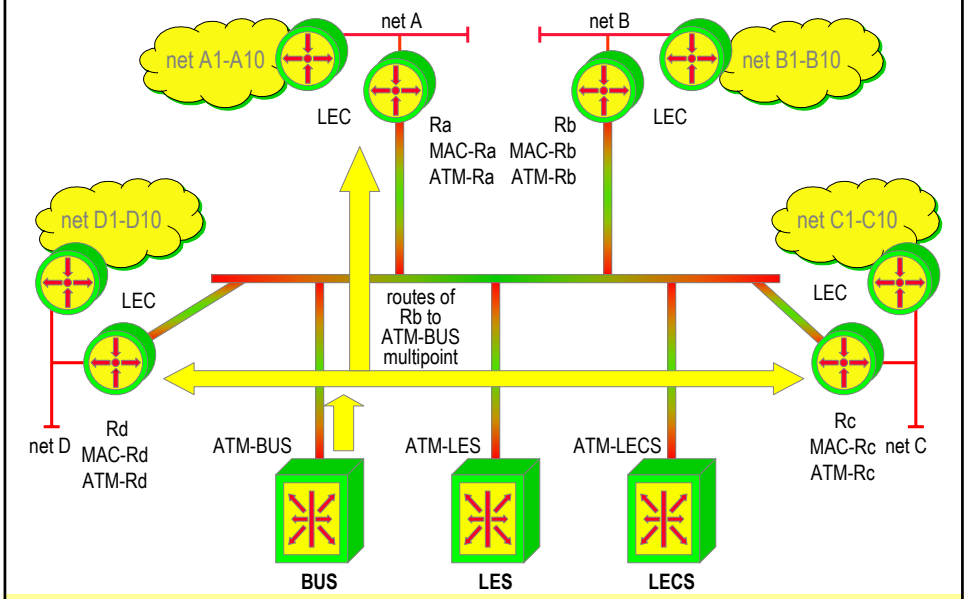
© 2005, D.I. Manfred Lindner

Routing Broadcast Rb



© 2005, D.I. Manfred Lindner Way to IP Switching, v3.5 39

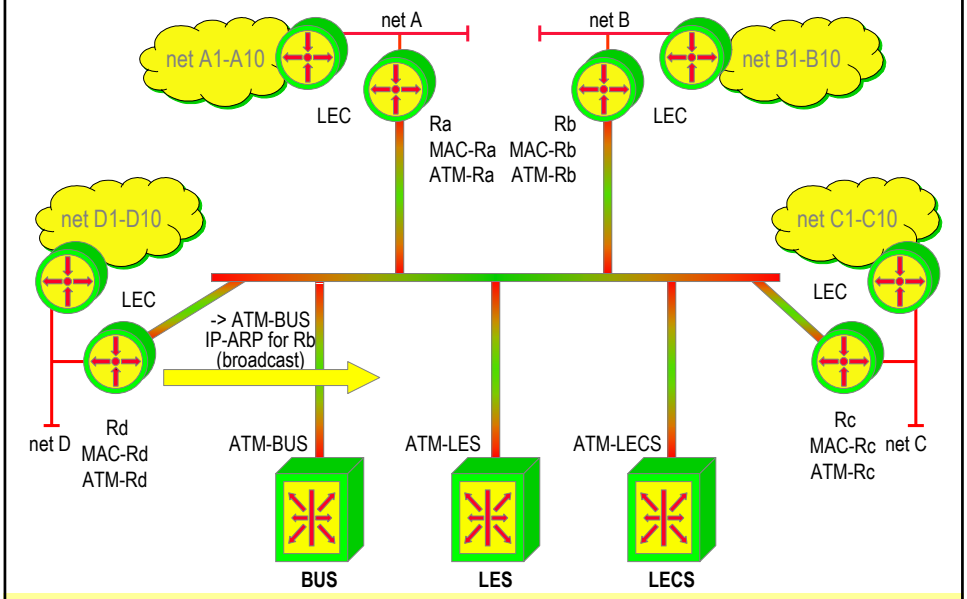
Routing Broadcast Distribution BUS



© 2005, D.I. Manfred Lindner Way to IP Switching, v3.5 40

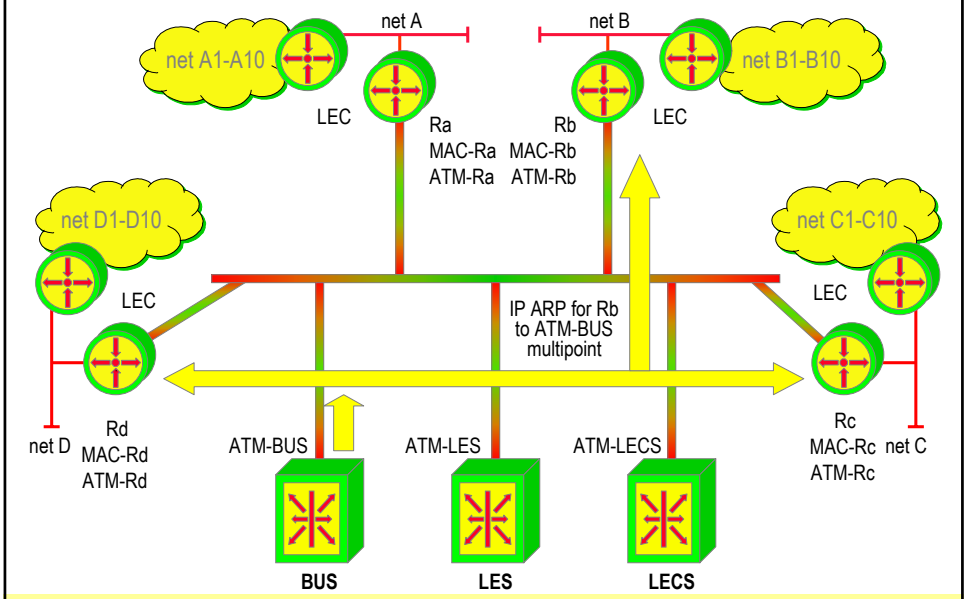
© 2005, D.I. Manfred Lindner

IP-ARP Request Rd



© 2005, D.I. Manfred Lindner Way to IP Switching, v3.5 41

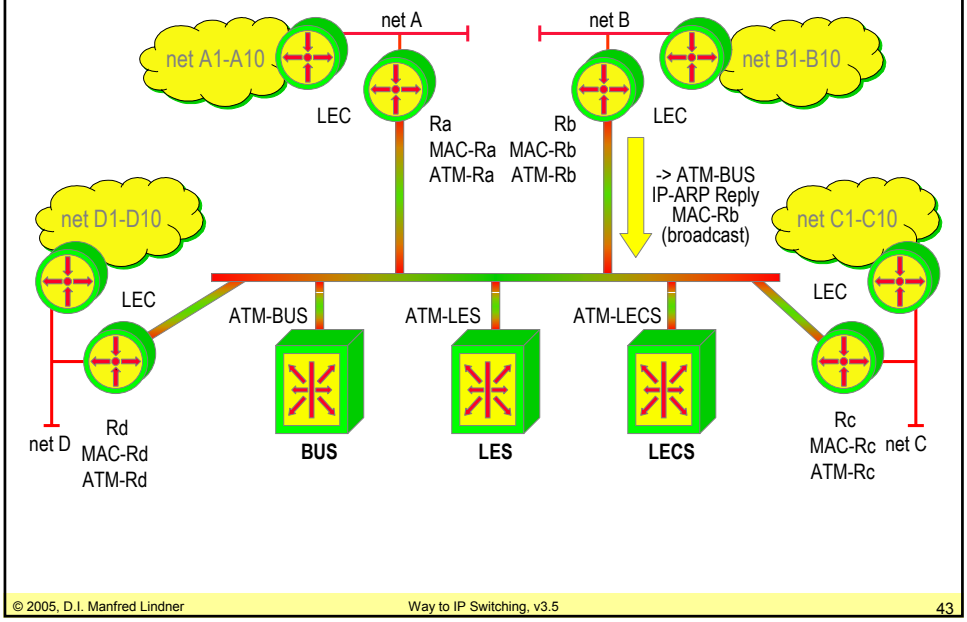
IP-ARP Request Distribution BUS



© 2005, D.I. Manfred Lindner Way to IP Switching, v3.5 42

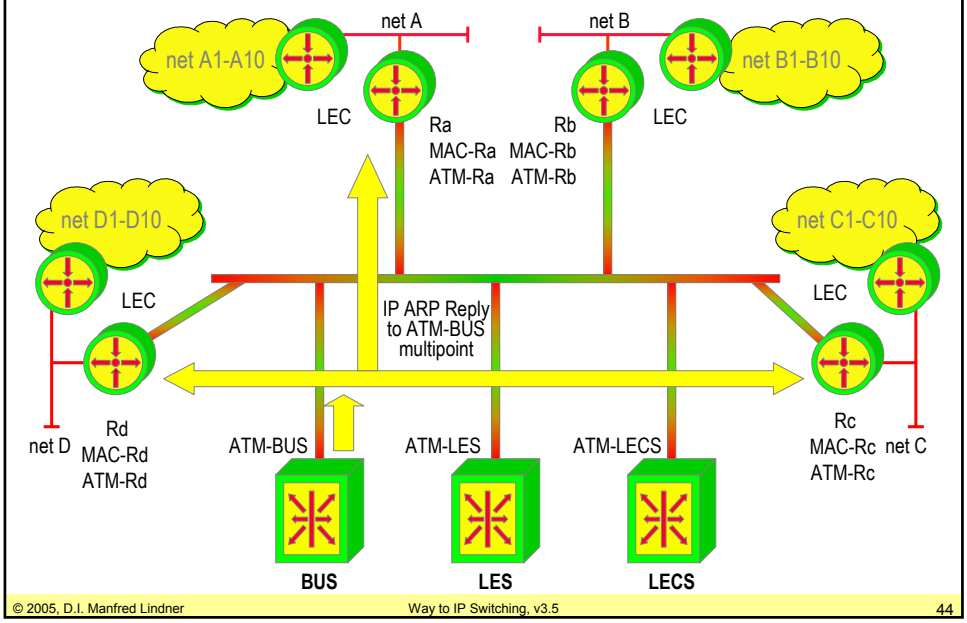
© 2005, D.I. Manfred Lindner

IP-ARP Reply Rb



© 2005, D.I. Manfred Lindner

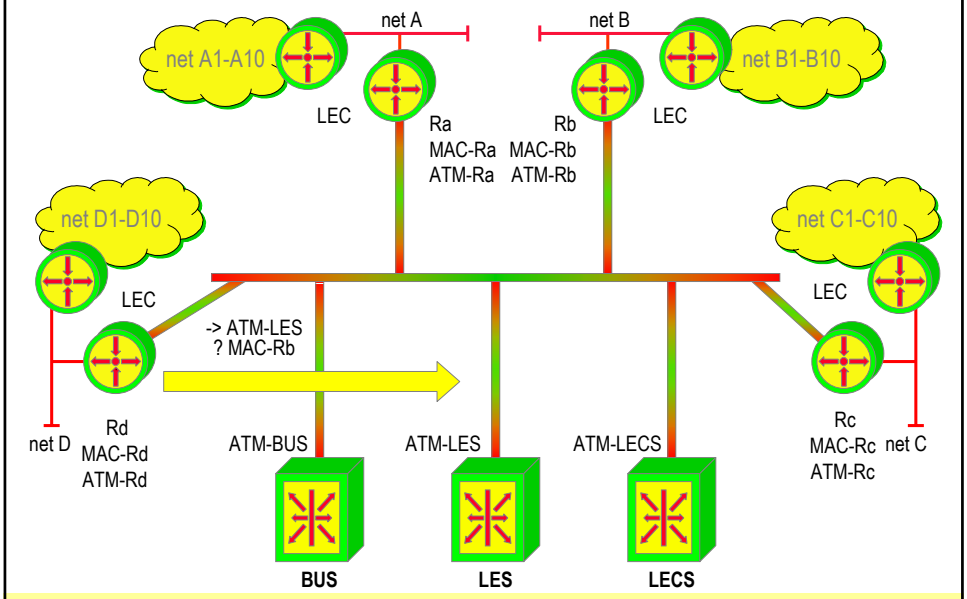
IP-ARP Reply Distribution BUS



© 2005, D.I. Manfred Lindner

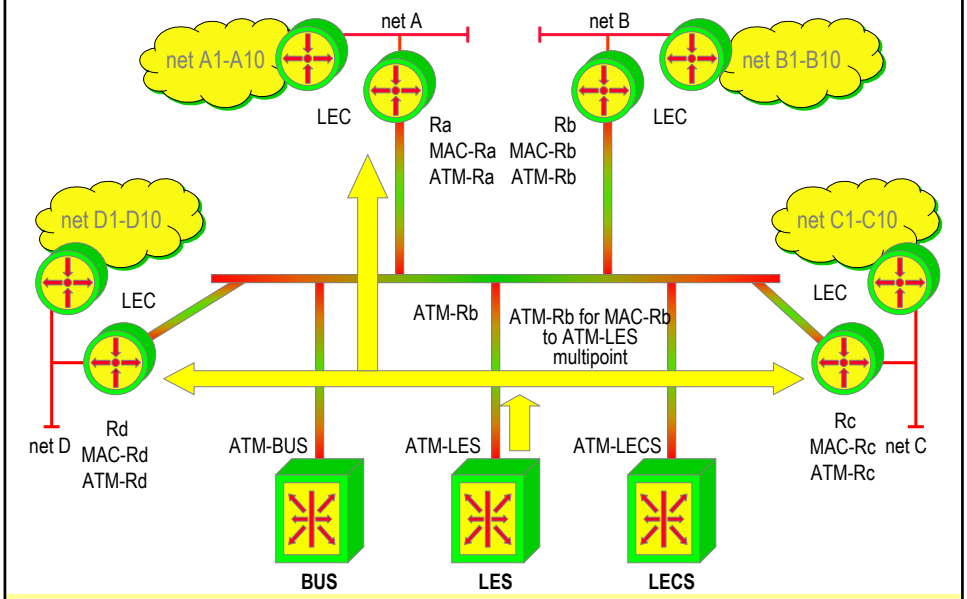
© 2005, D.I. Manfred Lindner

LANE-ARP Request Rd



© 2005, D.I. Manfred Lindner Way to IP Switching, v3.5 45

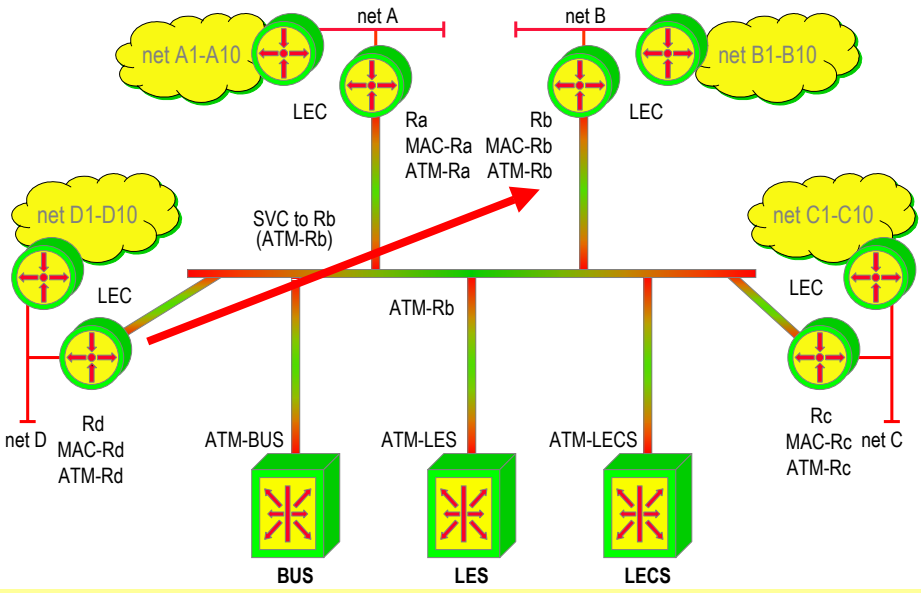
LANE-ARP Reply Distribution LES



© 2005, D.I. Manfred Lindner Way to IP Switching, v3.5 46

© 2005, D.I. Manfred Lindner

SVC on Demand Rd -> Rb



© 2005, D.I. Manfred Lindner

Agenda

- **Comparison of IP over ATM methods**
  - static address resolution
  - ARP server
  - MARS and MCS
  - LANE
  - NHRP
- **IP Switching**
- **Tag Switching**

© 2005, D.I. Manfred Lindner

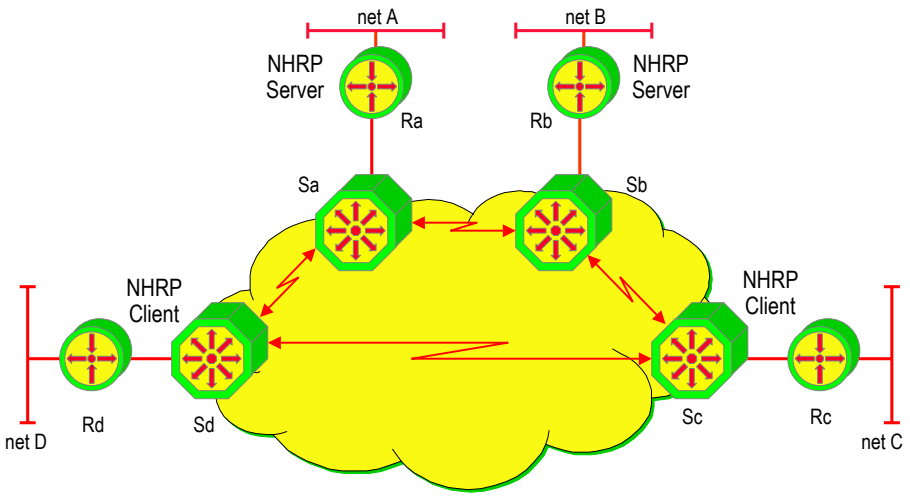
© 2005, D.I. Manfred Lindner

NHRP Server

- **a way to establish a cut-through circuit on demand by using NHRP protocol**
  - NHRP server provides address resolution for clients of the same LIS like RFC 1577
  - NHRP client can ask for a direct path for a given network different to its own LIS and may establish an exclusive, direct SVC for that network
  - dynamic routing is supported in fabric mode
  - routing messages and all default traffic will follow sequence of virtual circuits crossing routers hop-by-hop
- **neighbor discovery problem between routers not addressed by NHRP server**
  - must be still done by PVCs

© 2005, D.I. Manfred Lindner

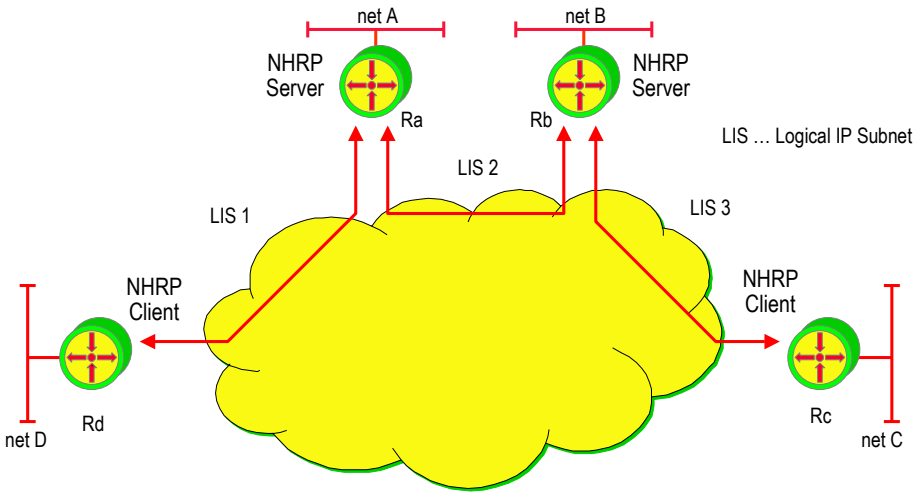
NHRP physical Topology



© 2005, D.I. Manfred Lindner

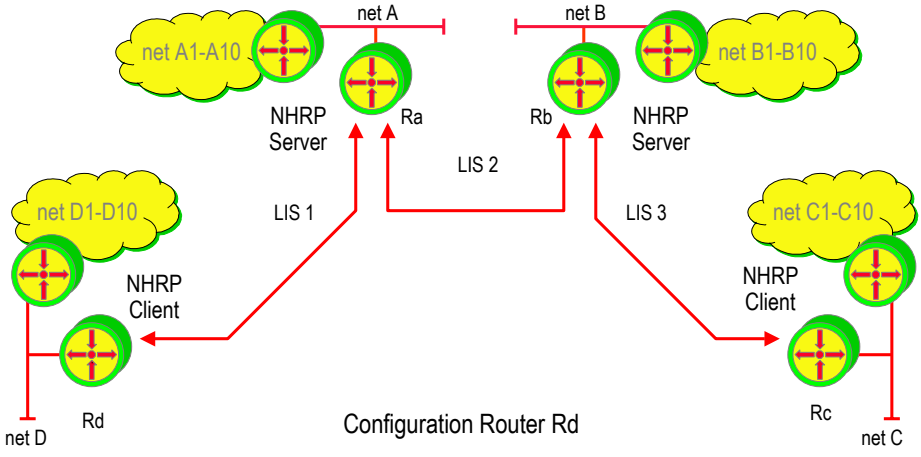
© 2005, D.I. Manfred Lindner

IP Basic Connectivity Hop-by-Hop



© 2005, D.I. Manfred Lindner Way to IP Switching, v3.5 51

Routing Broadcasts and Traffic Hop-by-Hop



dynamic routing on PVC
VPI/VCI Rd ⇨ Ra broadcast

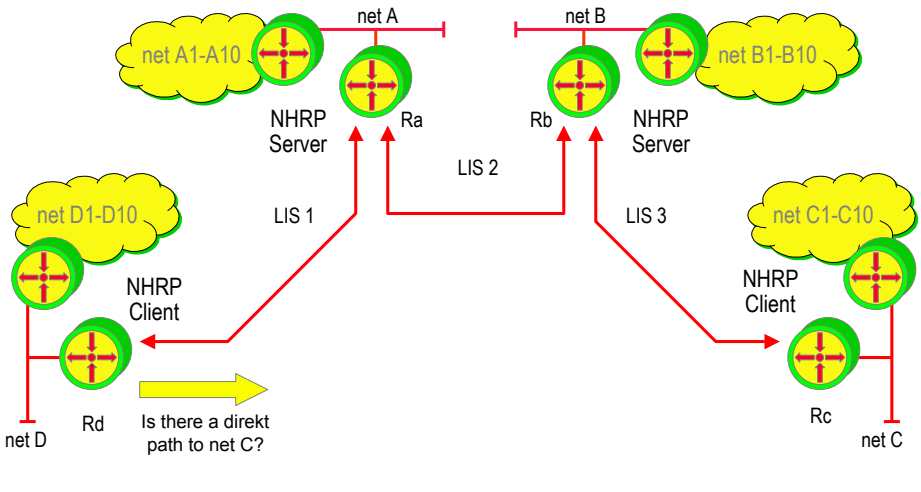
address resolution for next hop router by static mapping
Ra map VPI/VCI Rd ⇨ Ra

address of NHRP Server
ATM address of Ra

© 2005, D.I. Manfred Lindner Way to IP Switching, v3.5 52

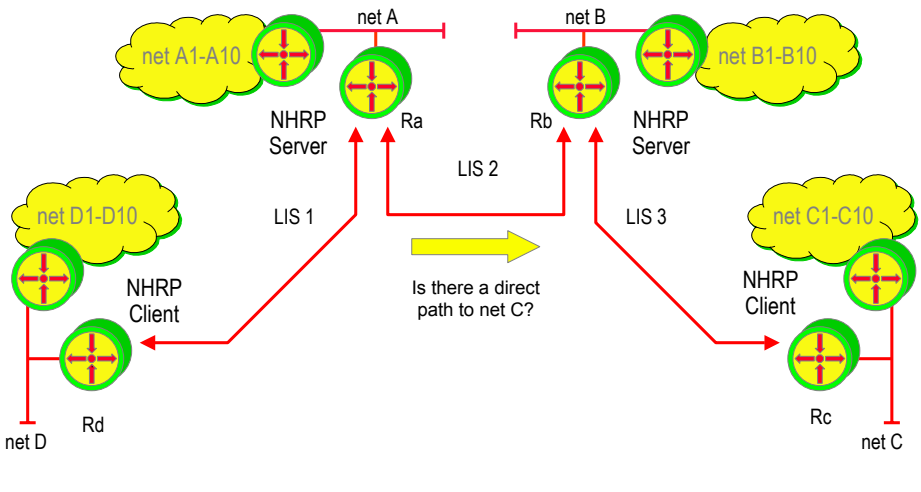
© 2005, D.I. Manfred Lindner

Routing Broadcasts and Traffic Hop-by-Hop



© 2005, D.I. Manfred Lindner

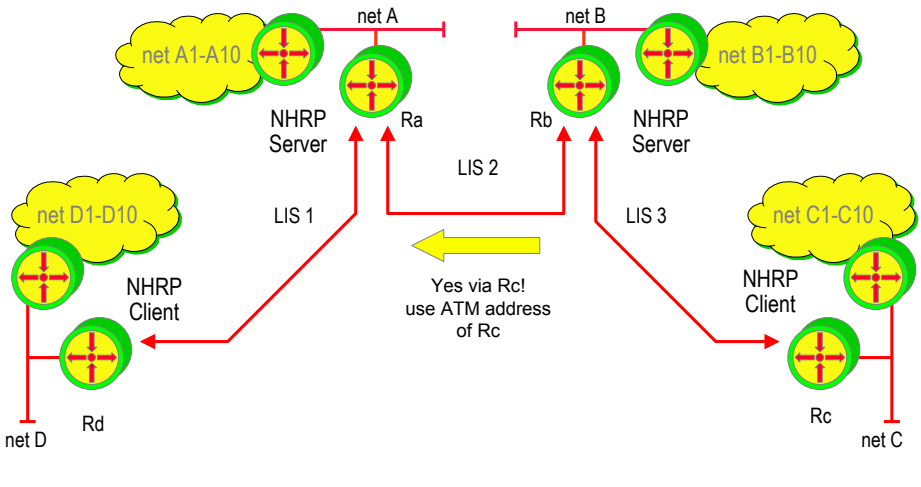
Routing Broadcasts and Traffic Hop-by-Hop



© 2005, D.I. Manfred Lindner

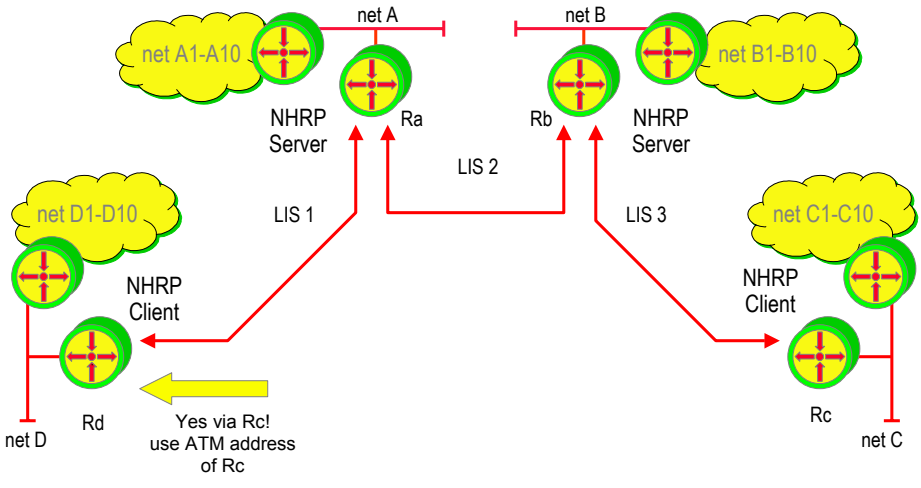
© 2005, D.I. Manfred Lindner

Routing Broadcasts and Traffic Hop-by-Hop



© 2005, D.I. Manfred Lindner

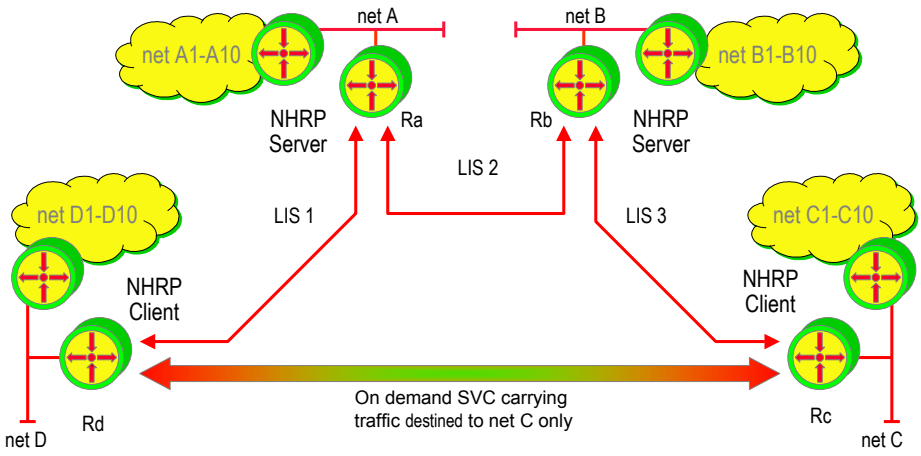
Routing Broadcasts and Traffic Hop-by-Hop



© 2005, D.I. Manfred Lindner

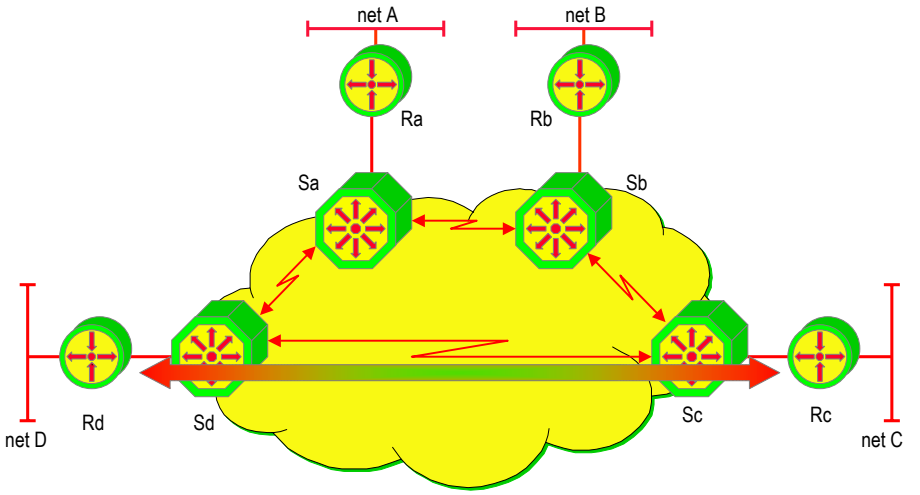
© 2005, D.I. Manfred Lindner

On Demand cut-through circuit



© 2005, D.I. Manfred Lindner

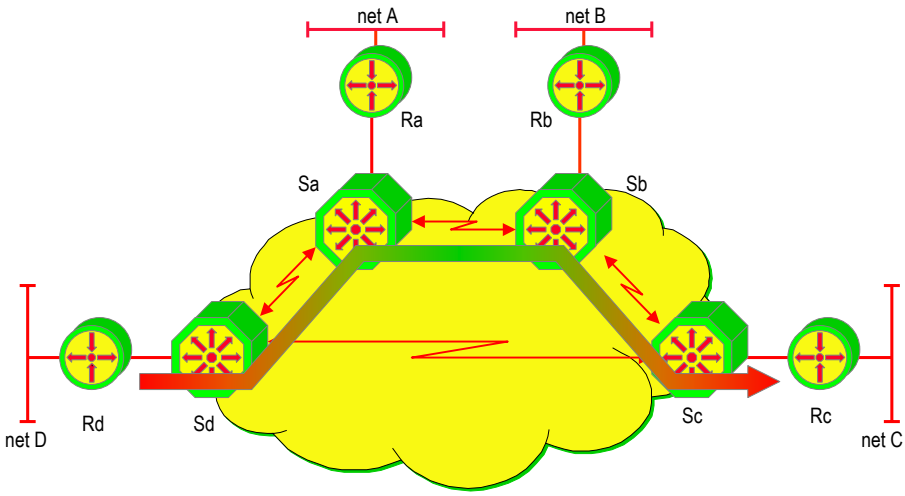
Possible VC Path 1



© 2005, D.I. Manfred Lindner

© 2005, D.I. Manfred Lindner

Possible VC Path 2



© 2005, D.I. Manfred Lindner

Agenda

- **Comparison of IP over ATM methods**
  - static address resolution
  - ARP server
  - MARS and MCS
  - LANE
  - NHRP
- **IP Switching**
- **Tag Switching**

© 2005, D.I. Manfred Lindner

© 2005, D.I. Manfred Lindner

IP Switching

- **proposal of Ipsilon Networks**
  - avoids costly cell reassembly for long-term flows
  - each ATM switch should be equipped with routing software
  - based on destination routing
- **requires flow detection**
  - flow-oriented traffic: ftp, telnet, web, multimedia
  - short-lived traffic: DNS query, mail, time sync, management
  - claims: 80% of packets and 90% of bytes are flow-oriented

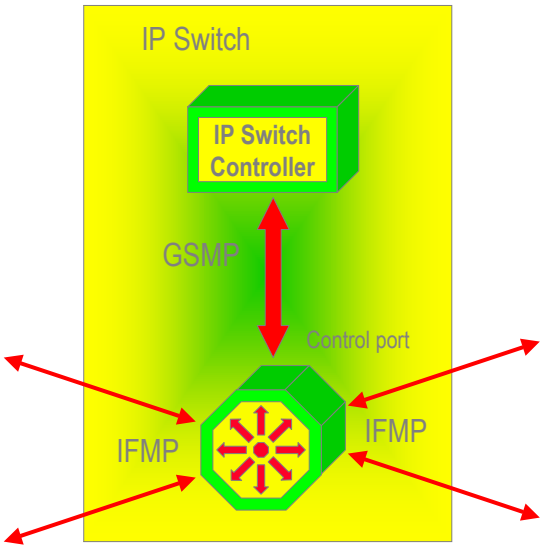
IP Switching

- **uses special protocols**
  - Generic Switch Management Protocol (GSMP)
  - Ipsilon Flow Management Protocol (IFMP)
  - RFC1953
- **relatively simple software should be added to the ATM switch**
  - GSMP - 2000 lines
  - IFMP - 10000 lines
- **multicast flows are mapped into point-multipoint VCs**
- **initially IP only, other protocols must tunneled**

© 2005, D.I. Manfred Lindner

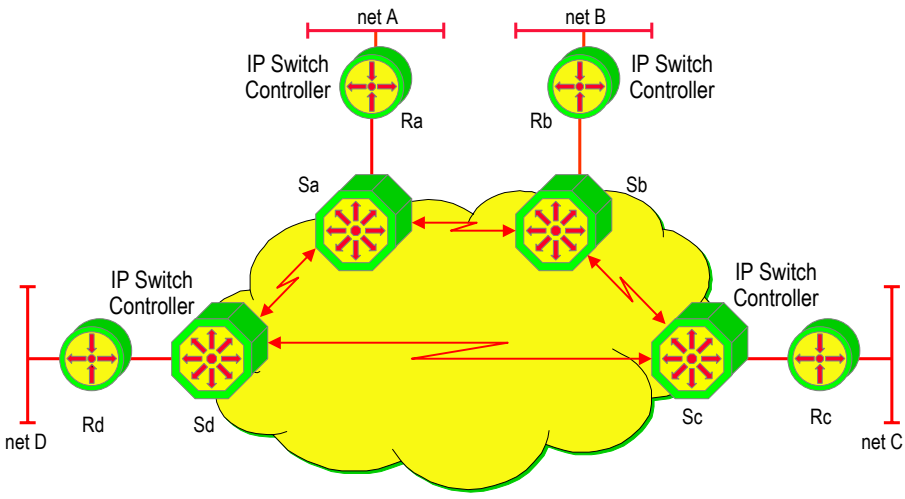
L83 - The Way to IP Switching

IP Switching



© 2005, D.I. Manfred Lindner Way to IP Switching, v3.5 63

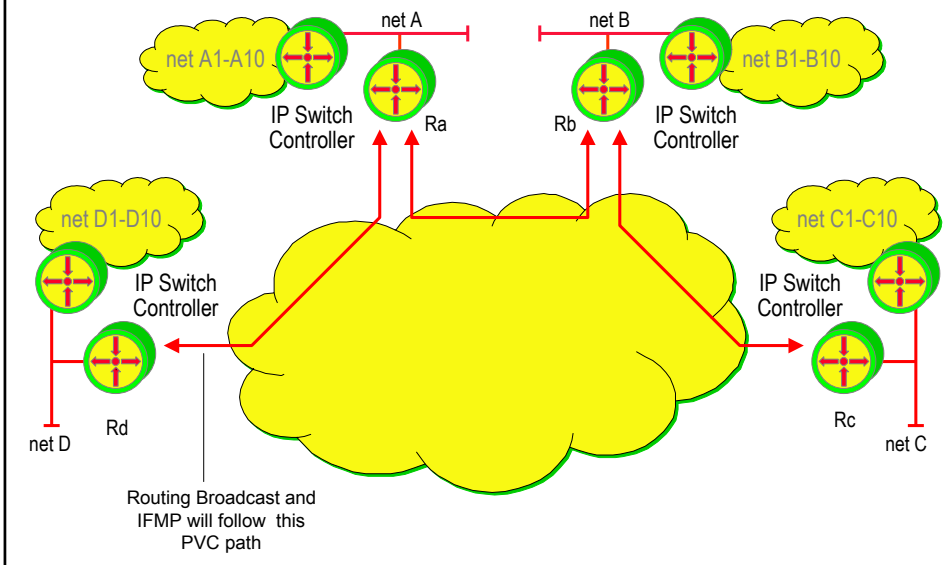
IP Switching



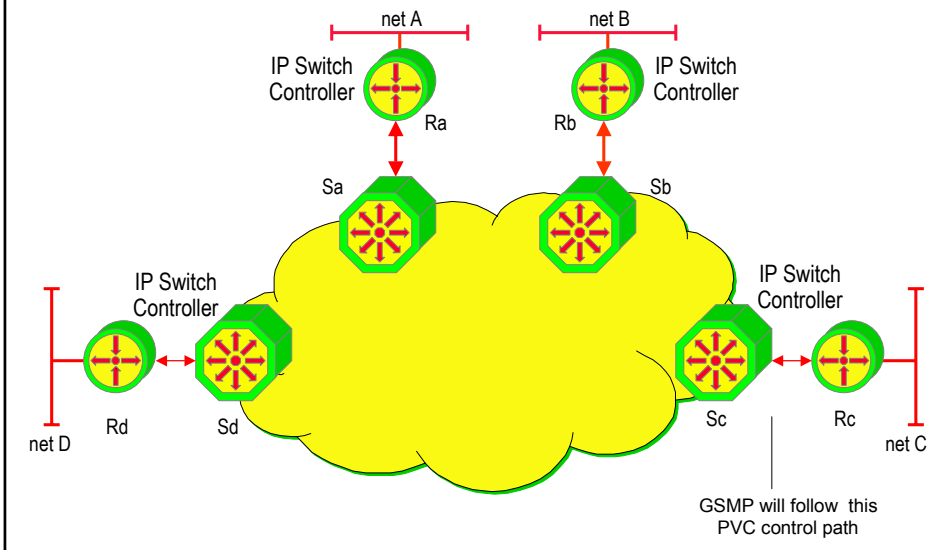
© 2005, D.I. Manfred Lindner Way to IP Switching, v3.5 64

© 2005, D.I. Manfred Lindner

IP Basic Connectivity via normal VC´s

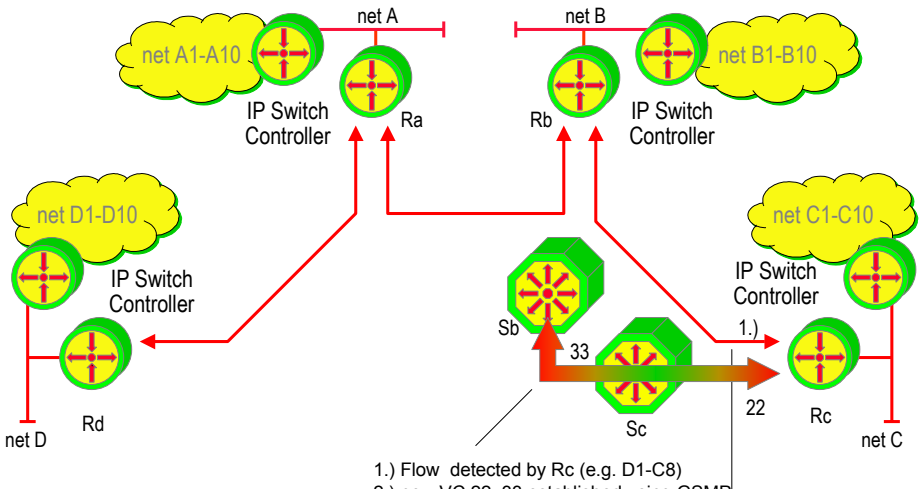


Control Virtual Circuit for GSMP



© 2005, D.I. Manfred Lindner

Flow Detection Rc

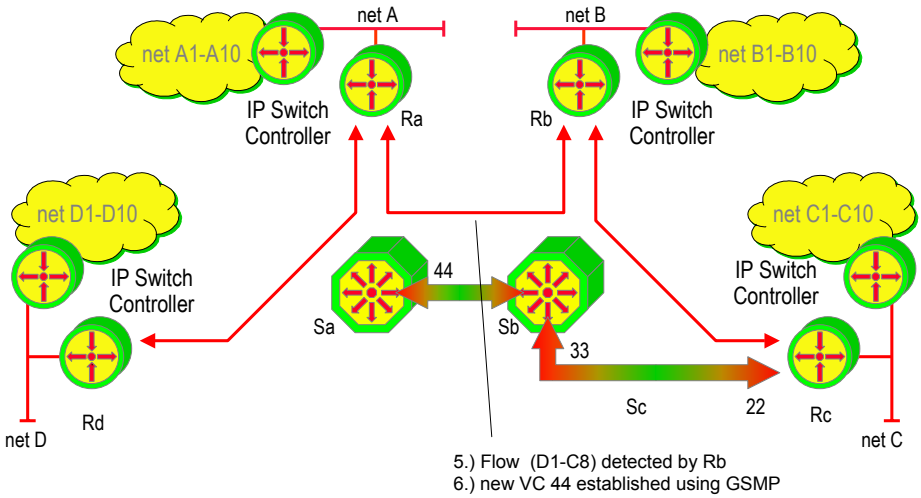


1.) Flow detected by Rc (e.g. D1-C8)
2.) new VC 22, 33 established using GSMP
3.) VC 22 and VC 33 connected via GSMP

4.) new VC 33 and Flow (D1-C8) signaled by IFMP to upstream IP Switch Controller

© 2005, D.I. Manfred Lindner

Flow Detection Rb

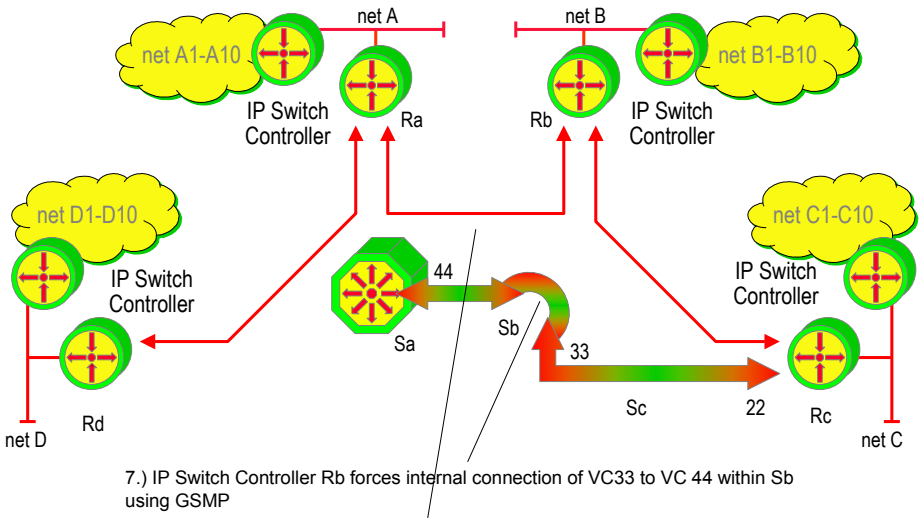


5.) Flow (D1-C8) detected by Rb
6.) new VC 44 established using GSMP

© 2005, D.I. Manfred Lindner

© 2005, D.I. Manfred Lindner

Cut-Through on Switch Sb

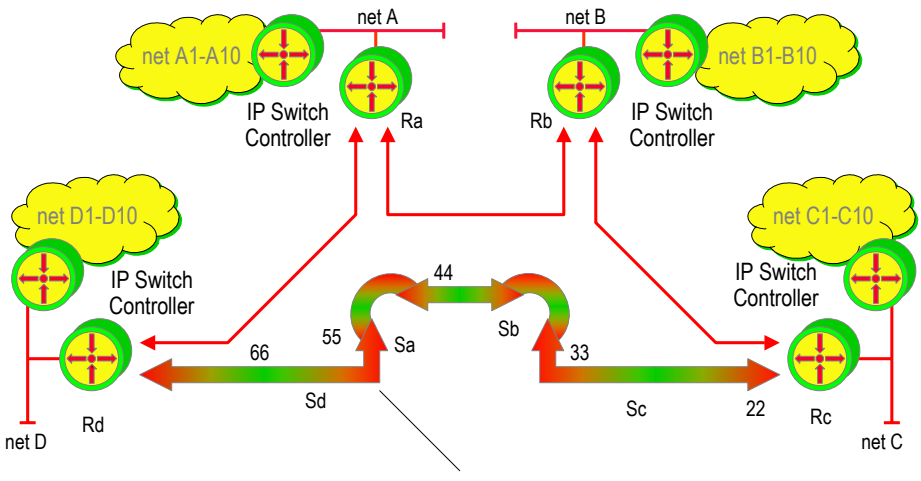


7.) IP Switch Controller Rb forces internal connection of VC33 to VC 44 within Sb using GSMP

8.) new VC 44 signaled by IFMP to upstream IP Switch Controller

© 2005, D.I. Manfred Lindner

Final Picture

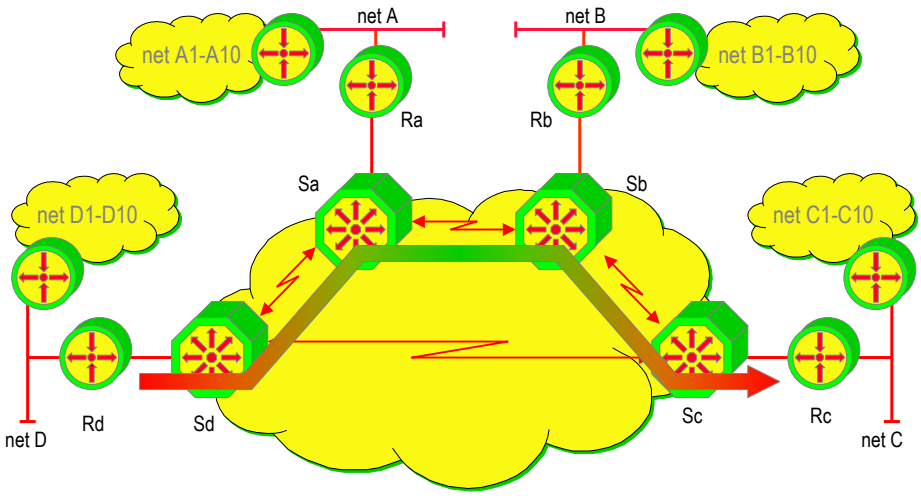


new path bypassing IPSwitch Controllers established.
Flow can now use new VC (separated from basic VC)

© 2005, D.I. Manfred Lindner

© 2005, D.I. Manfred Lindner

Physical Topology



IP Switching Operations

- **Step 1 - classical packet forwarding**
- **Step 2 - flow identification, first hop labeled**
  - mapping input to control
  - redirect request sent upward
- **Step 3 - flow identification, second hop labeled**
  - redirect request might come on output
  - local decision, on accepting connect to control
- **Step 4 - shortcut is created inside the switch for the flow**
  - input and output is connected directly

© 2005, D.I. Manfred Lindner

IP Switching Operations

- **Flow labeling**
  - VCI field is used
- **Refresh timer**
  - because it is not a real connection-oriented technology
  - if there was no activity the VC is closed
- **Uses a point-to-point model**
- **Quality of service**
  - based on local decisions
  - RSVP can be mapped

IP Switching Issues

- **only usable over ATM as a transport**
- **connection setup on demand, so first packet is not switched**
- **scalability is limited by VC explosion**
  - number of VC is greater or equal to number of flows

© 2005, D.I. Manfred Lindner

Agenda

- **Comparison of IP over ATM methods**
  - static address resolution
  - ARP server
  - MARS and MCS
  - LANE
  - NHRP
- **IP Switching**
- **Tag Switching**

Tag Switching

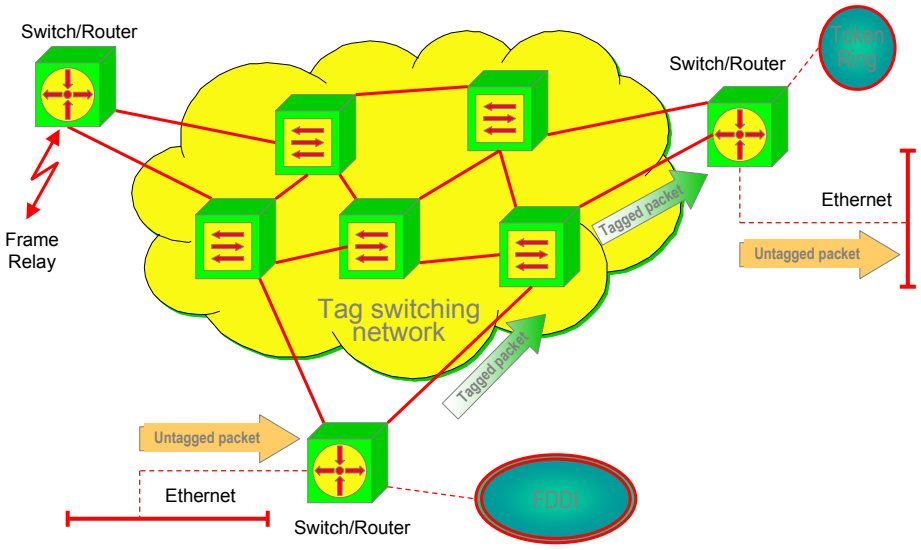
- **continuous growth of Internet and emerging multimedia applications**
  - demand higher bandwidth within the Internet Service Provider (ISP)
  - demand improved scaling properties of the Internet routing system
- **demand for higher bandwidth**
  - requires higher forwarding performance for both multicast and unicast traffic
- **the destination-based forwarding paradigm**
  - is not adequate in all situations
    - we would like to have more flexible control over how traffic is routed

© 2005, D.I. Manfred Lindner

Tag Switching

- **method to combine label-swapping with network layer routing**
  - ingress edge device attaches a label to the packet
  - switches do forwarding decisions based on the labels
    - very quick lookup, lower latency
  - tags have only local significance
  - egress edge device strips off the tag
- **proposed by Cisco**
  - RFC 2105 - “Cisco Systems’ Tag Switching Architecture Overview”

Tag Switching



© 2005, D.I. Manfred Lindner

Tag Switching

- **Tag switching is a technology that promises**
  - an efficient solution to these challenges
- **Tag switching combines**
  - the flexibility and functionality provided by Network Layer routing
  - with the simplicity provided by the label swapping forwarding paradigm
- **the simplicity of label swapping forwarding paradigm**
  - enable improved forwarding performance

© 2005, D.I. Manfred Lindner

Tag Switching

- **by associating a wide range of forwarding granularities with a tag**
  - the same forwarding paradigm can be used to support a wide variety of routing functions
    - such as destination-based routing, multicast, hierarchy of routing knowledge, and flexible routing control
- **is intended to simplify integration of routers and ATM switches**
  - by employing common addressing, routing, and management procedures
    - in fact, competition to Integrated PNNI
  - intended to be used in large environments, not applicable for workgroups and campus environments
  - independent of network layer

© 2005, D.I. Manfred Lindner

© 2005, D.I. Manfred Lindner

Tag Switching

- **forwarding component**
  - uses tags carried by packets plus tag information maintained by a tag switch to perform packet forwarding
- **control component**
  - maintains correct tag forwarding information among a group of tag switches
- **tags could be**
  - inserted between MAC layer and Link layer
  - part of the layer 2 header (for example ATM)
  - part of the network layer header (for example flow label in IPv6)

© 2005, D.I. Manfred Lindner

Tag Switching

- **control component**
  - is responsible for create tag bindings
    - allocation of a tag
    - bind the tag to a route
  - distribute the tag bindings among tag switches
    - use existing routing protocols
    - use tag distribution protocol
- **performance**
  - one VC per routing table entry
  - one memory reference compared to 4-16 in classical router

© 2005, D.I. Manfred Lindner

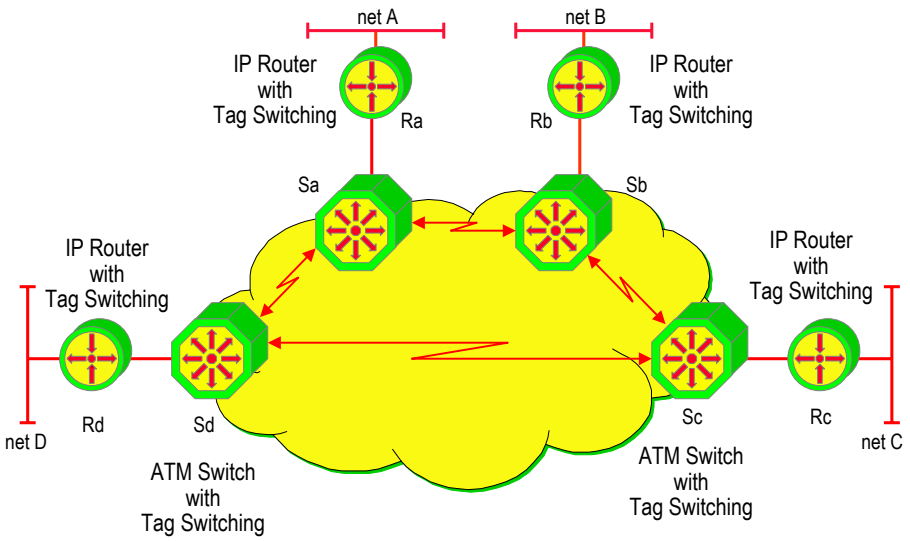
© 2005, D.I. Manfred Lindner

Tag Switching

- **Destination based routing**
  - downstream tag allocation (⇨ example follows)
  - downstream tag allocation on demand
  - upstream tag allocation
- **Tag information distribution**
  - piggybacking existing routing protocols
  - Tag Distribution Protocol (TDP)
- **Separate tags for inter-AS and intra-AS switching**
- **Multicast tagging**
- **Policy based routing is supported**
  - manual or external assignment of tags

© 2005, D.I. Manfred Lindner

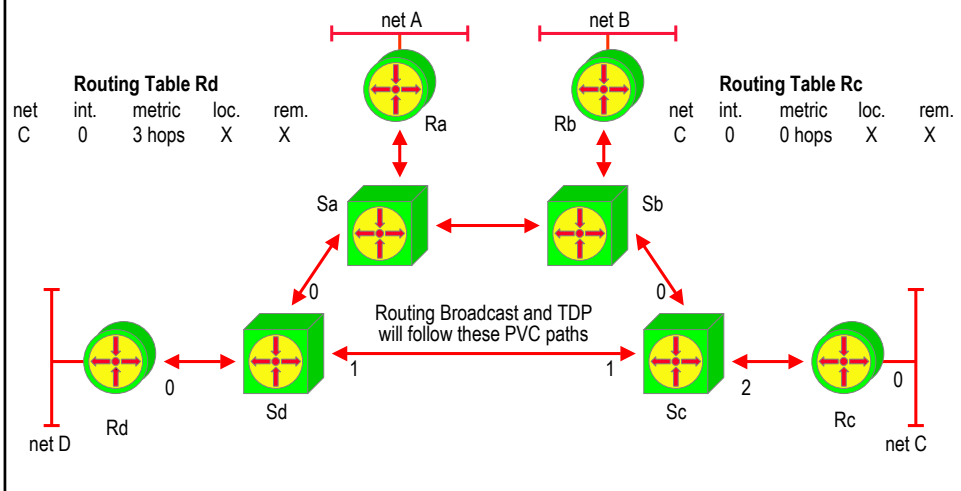
Tag Switching



© 2005, D.I. Manfred Lindner

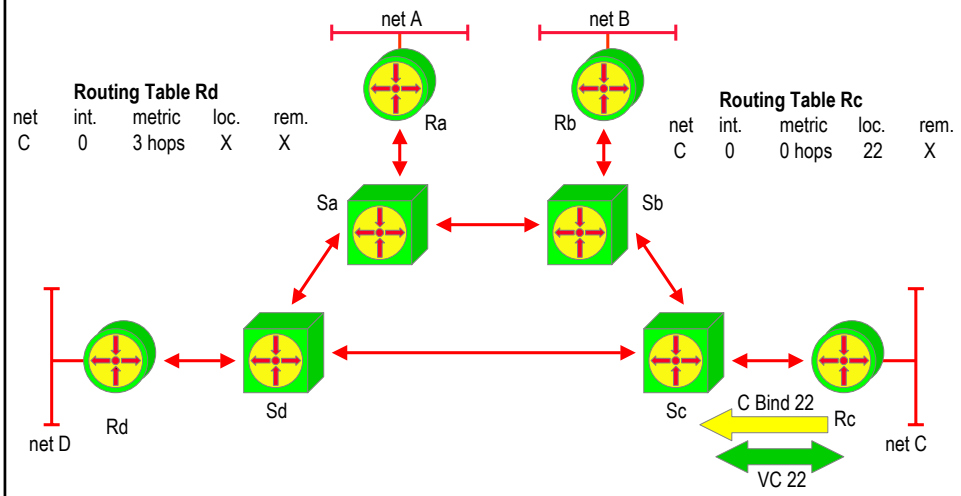
© 2005, D.I. Manfred Lindner

IP Basic Connectivity and Routing



© 2005, D.I. Manfred Lindner Way to IP Switching, v3.5 85

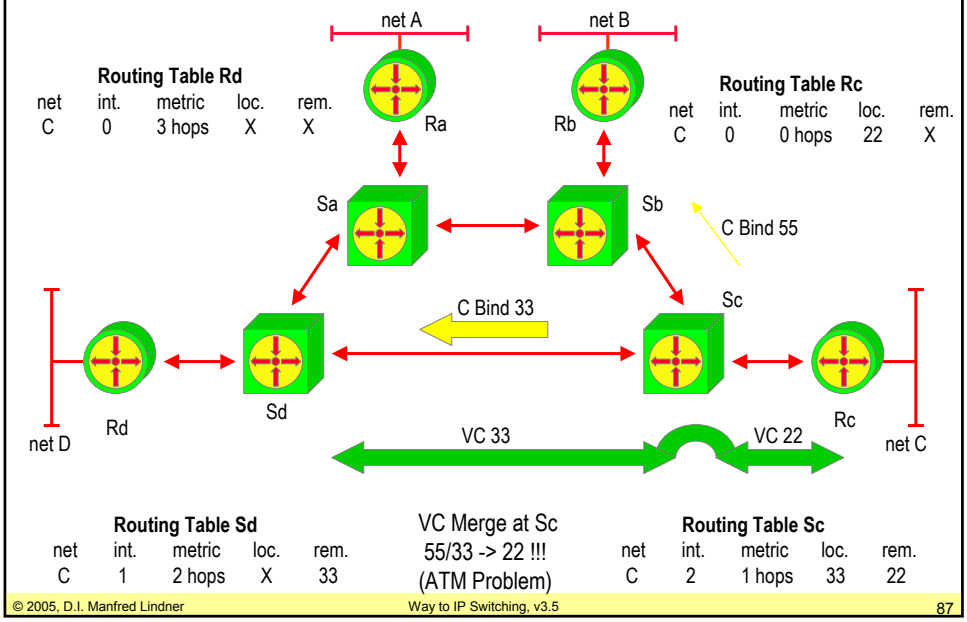
Downstream Allocation 1 (Default)



© 2005, D.I. Manfred Lindner Way to IP Switching, v3.5 86

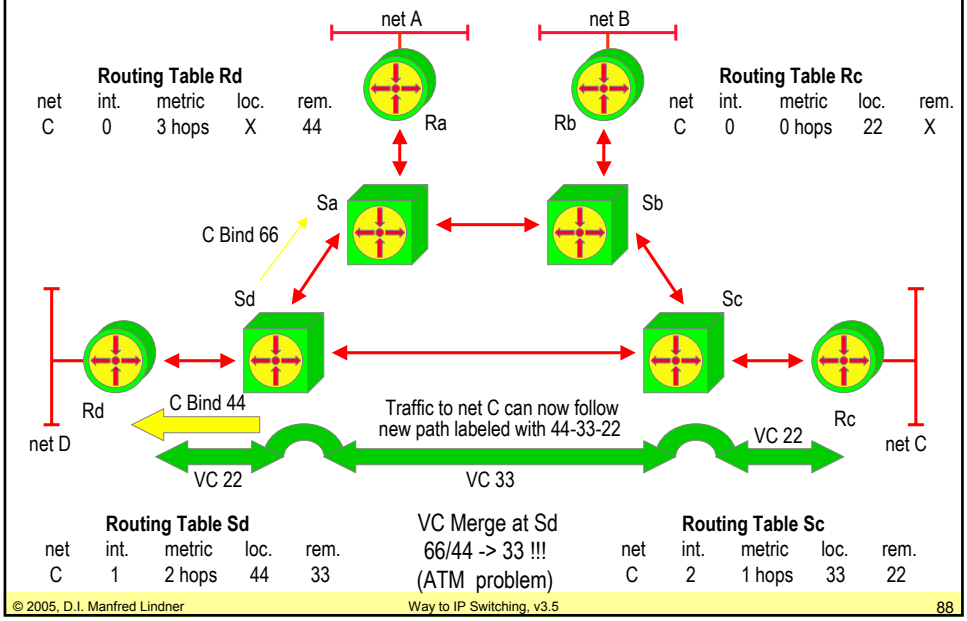
© 2005, D.I. Manfred Lindner

Downstream Allocation 2 (Default)



© 2005, D.I. Manfred Lindner

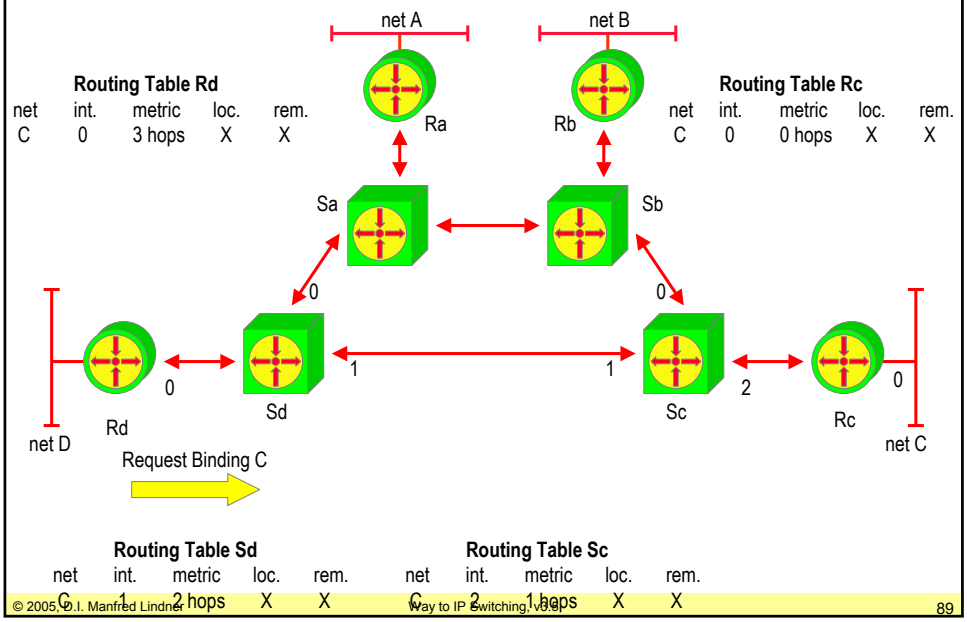
Downstream Allocation 3 (Default)



© 2005, D.I. Manfred Lindner

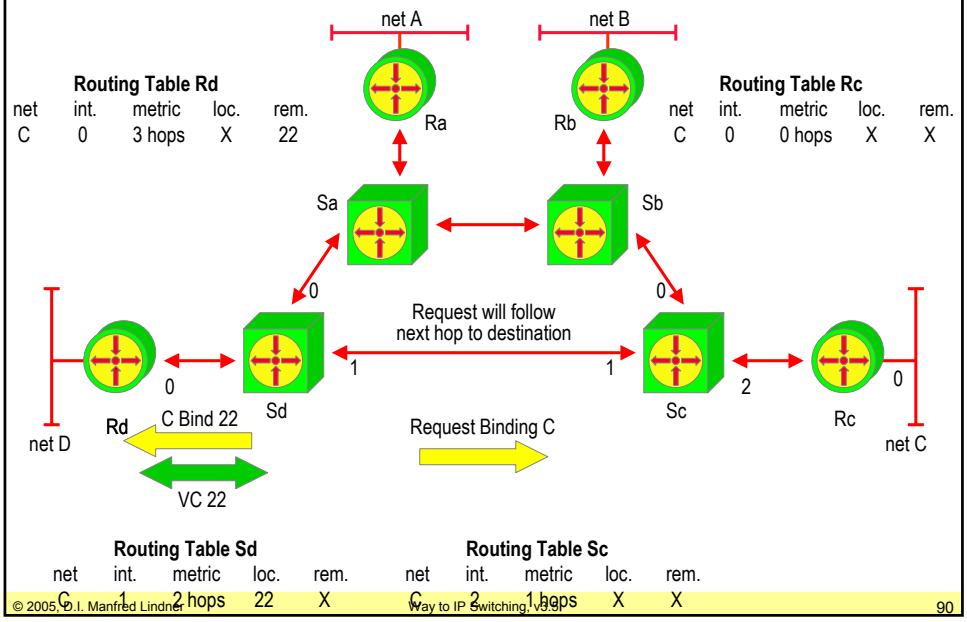
© 2005, D.I. Manfred Lindner

Downstream on Demand 1 (Mode for ATM)



© 2005, D.I. Manfred Lindner

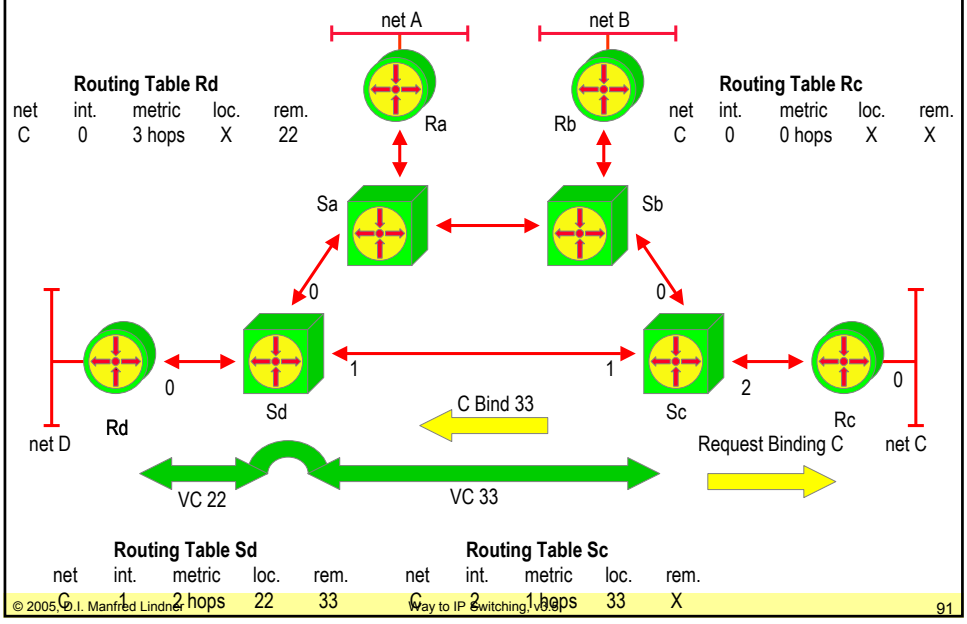
Downstream on Demand 2 (Mode for ATM)



© 2005, D.I. Manfred Lindner

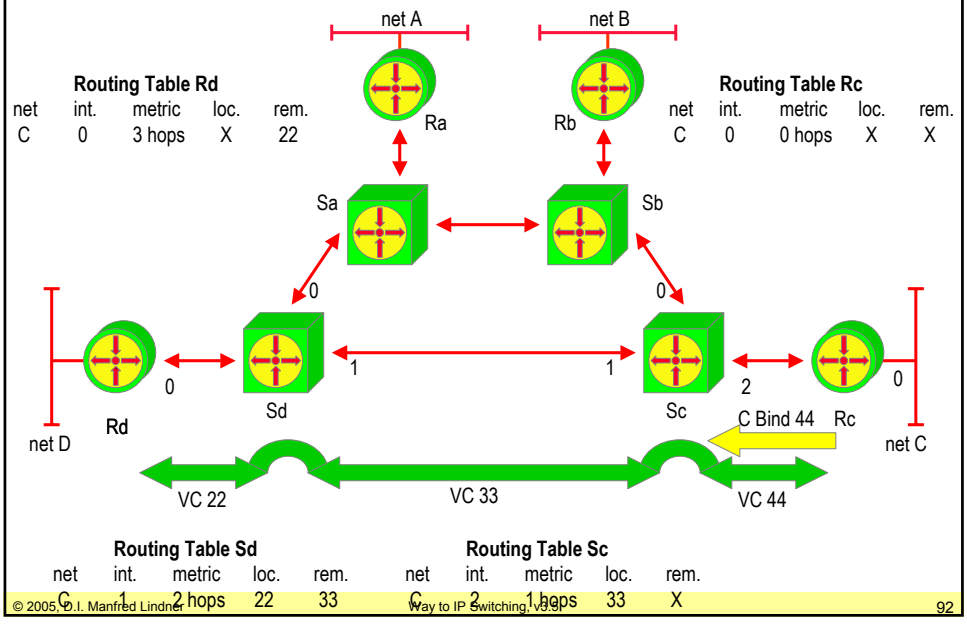
© 2005, D.I. Manfred Lindner

Downstream on Demand 3 (Mode for ATM)



© 2005, D.I. Manfred Lindner

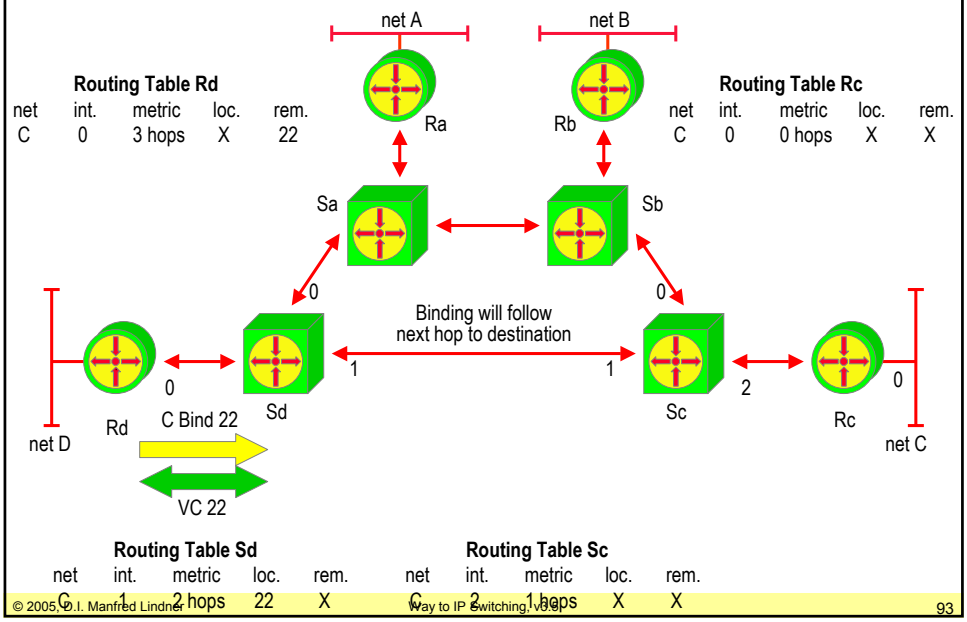
Downstream on Demand 4 (Mode for ATM)



© 2005, D.I. Manfred Lindner

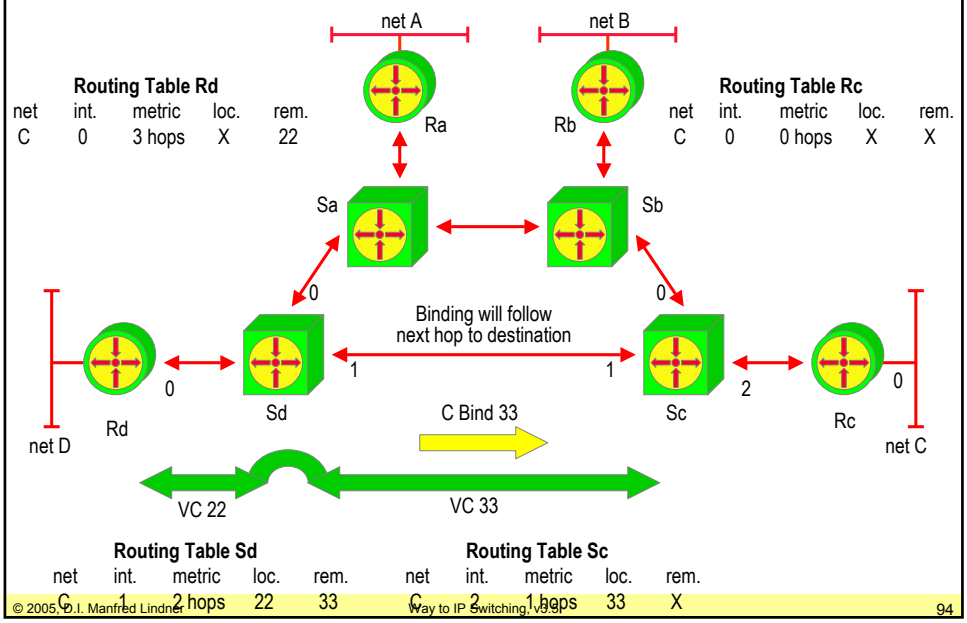
© 2005, D.I. Manfred Lindner

Upstream 1 (Mode for QoS/Traffic Engineering)



© 2005, D.I. Manfred Lindner

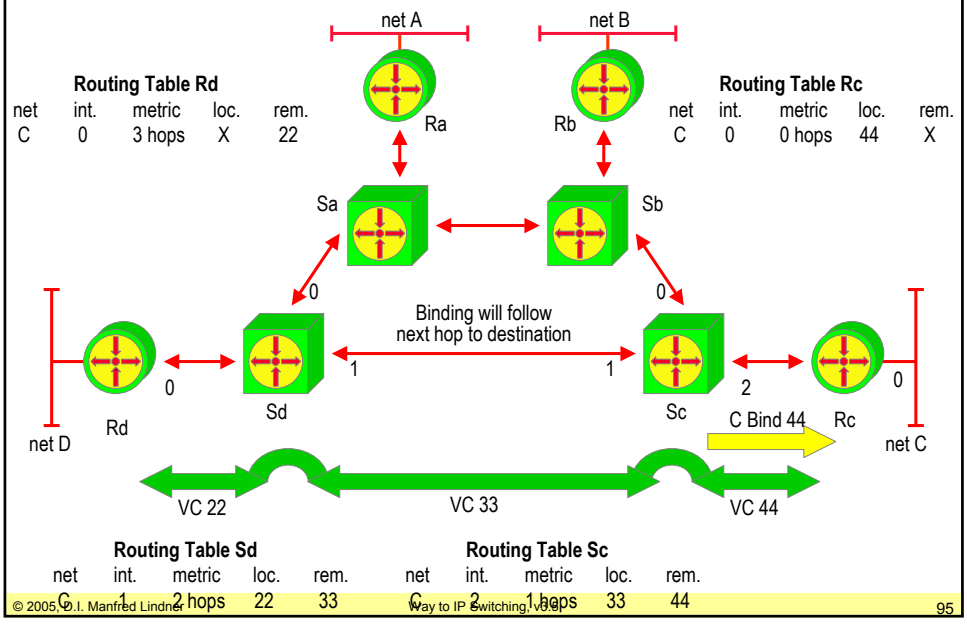
Upstream 2 (Mode for QoS/Traffic Engineering)



© 2005, D.I. Manfred Lindner

© 2005, D.I. Manfred Lindner

Upstream 3 (Mode for QoS/Traffic Engineering)



Tag Switching

- **quality of service support**
  - define tags associated with specific quality of service
  - use explicit routes to guarantee the quality of service
  - destination based routing can´t guarantee because every router can make it´s own decision
- **TAG switching and ATM**
  - can be easily deployed with ATM because ATM uses label swapping
    - VPI/VCI is used for tagging
  - ATM switches needs to implement control component of tag switching
  - ATM attached router peers with ATM switch (TAG switch)
    - exchange tag binding information

© 2005, D.I. Manfred Lindner