Enterasys OneFabric and Data Center Interconnect Solutions

Ivan Pepelnjak (ip@ipSpace.net)
NIL Data Communications

Markus Nispel
Enterasys Networks

Who is Markus Nispel?

- Chief Technology Strategist, Enterasys Networks
- Networking technologist since 1988
- Experience with IT and mobile networks
- Strategist, author & blogger (www.**sdn**central.com and www.enterasys.com)

Focus: Technology Strategy and Solutions Architecture

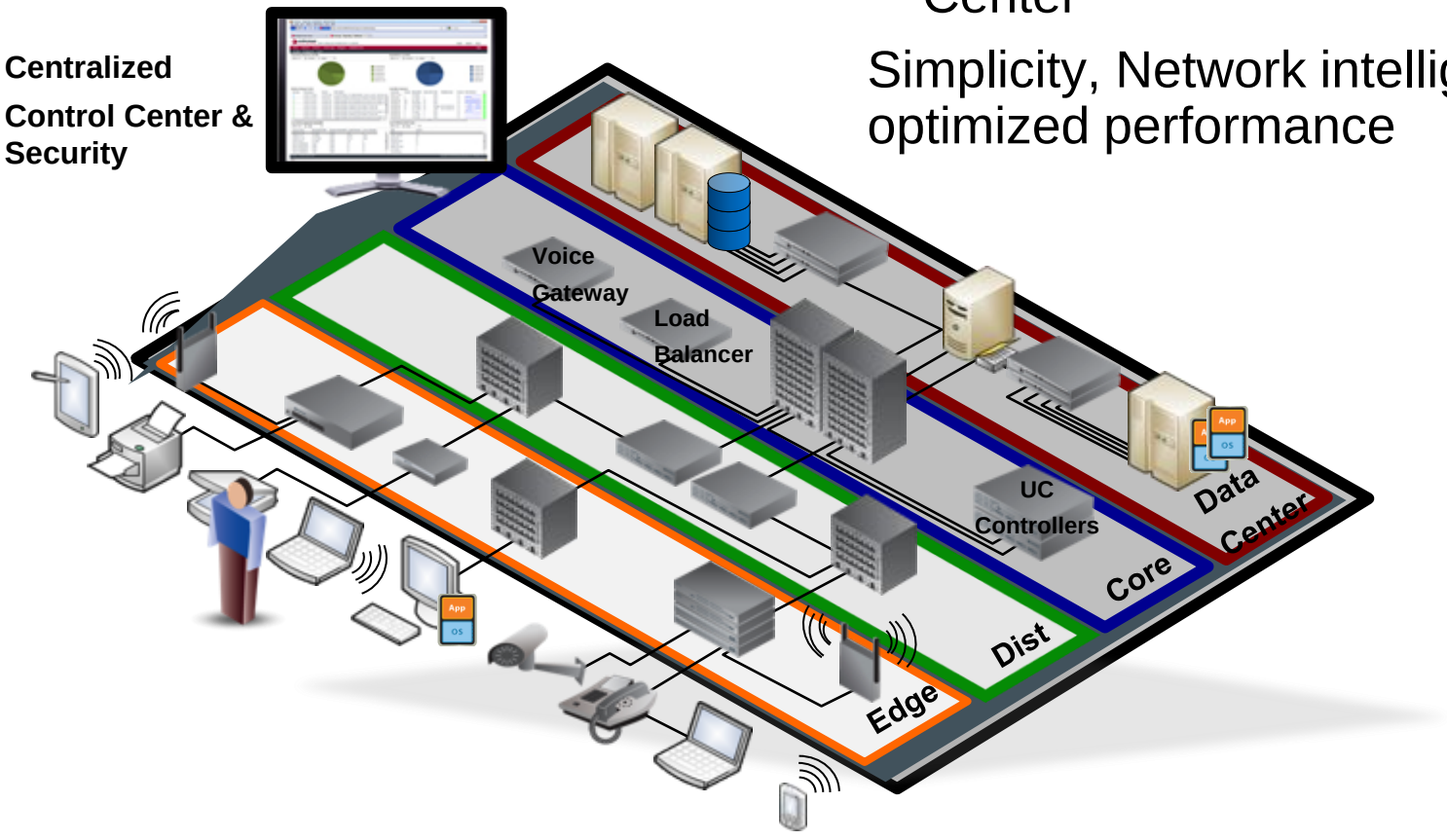
- Management and Security – Campus and Data Center
- Software Defined Networks
- Data Center Architecture
- Unified Access Strategy
- Application Visibility and Control

Enterasys Networks – A Brief Overview

End-to-End Network System Provider

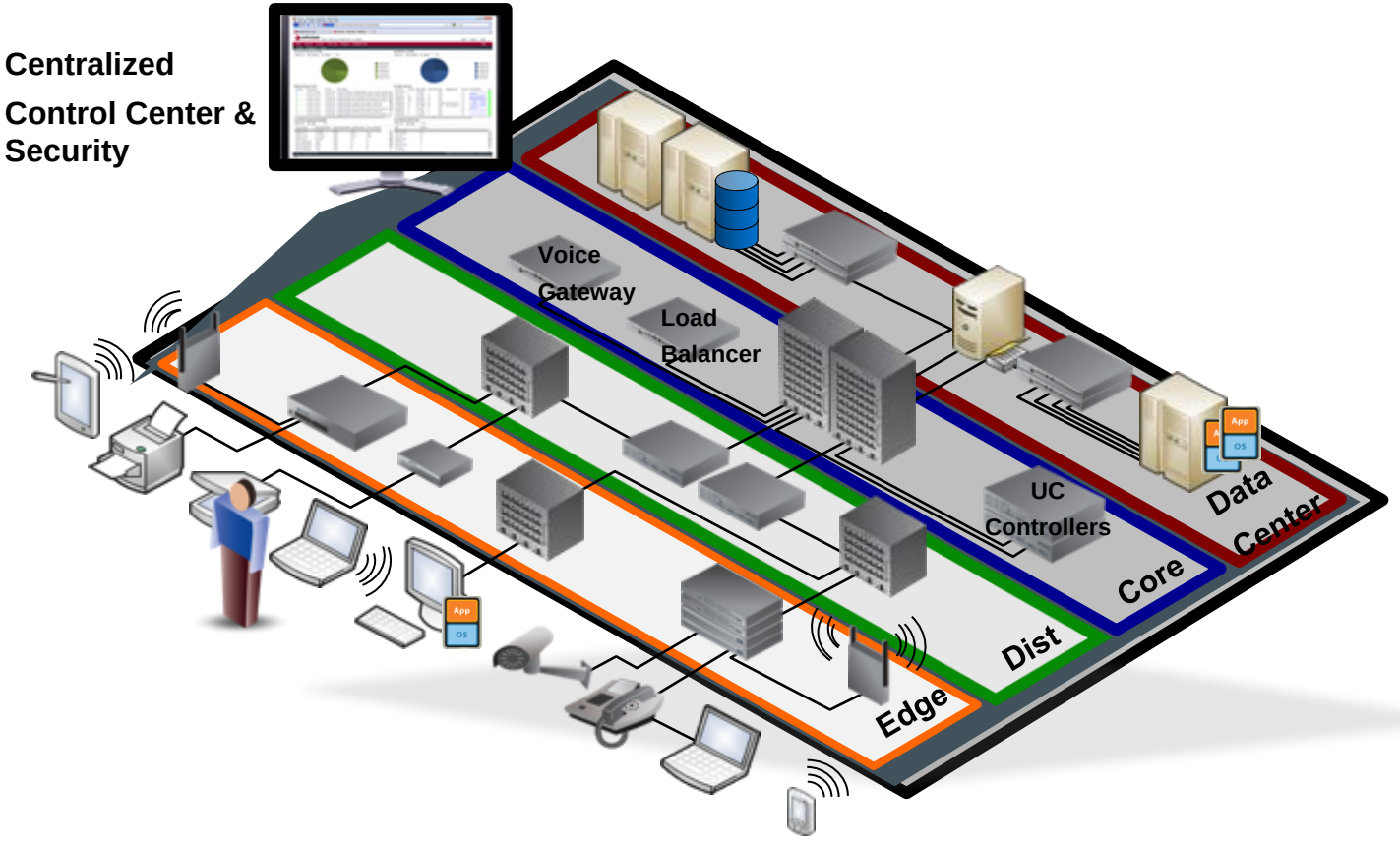
- Unified Fabric- and SDN-Architecture
- Centralized Management and Control
- From Edge (wired and wireless) to Data Center

Simplicity, Network intelligence and optimized performance



Enterasys Networks – A Brief Overview

- Global presence: 5 continents, 90+ countries, 800+ patents
- Mission critical proven: 20,000+ customers
- Sustained > 10% growth in network switching
- Sustained high global customer satisfaction: 95%



Who is Ivan Pepelnjak (@ioshints)

- Networking engineer since 1985
- Technical director, later Chief Technology Advisor @ NIL Data Communications
- Consultant, blogger (blog.ioshints.info), book and webinar author

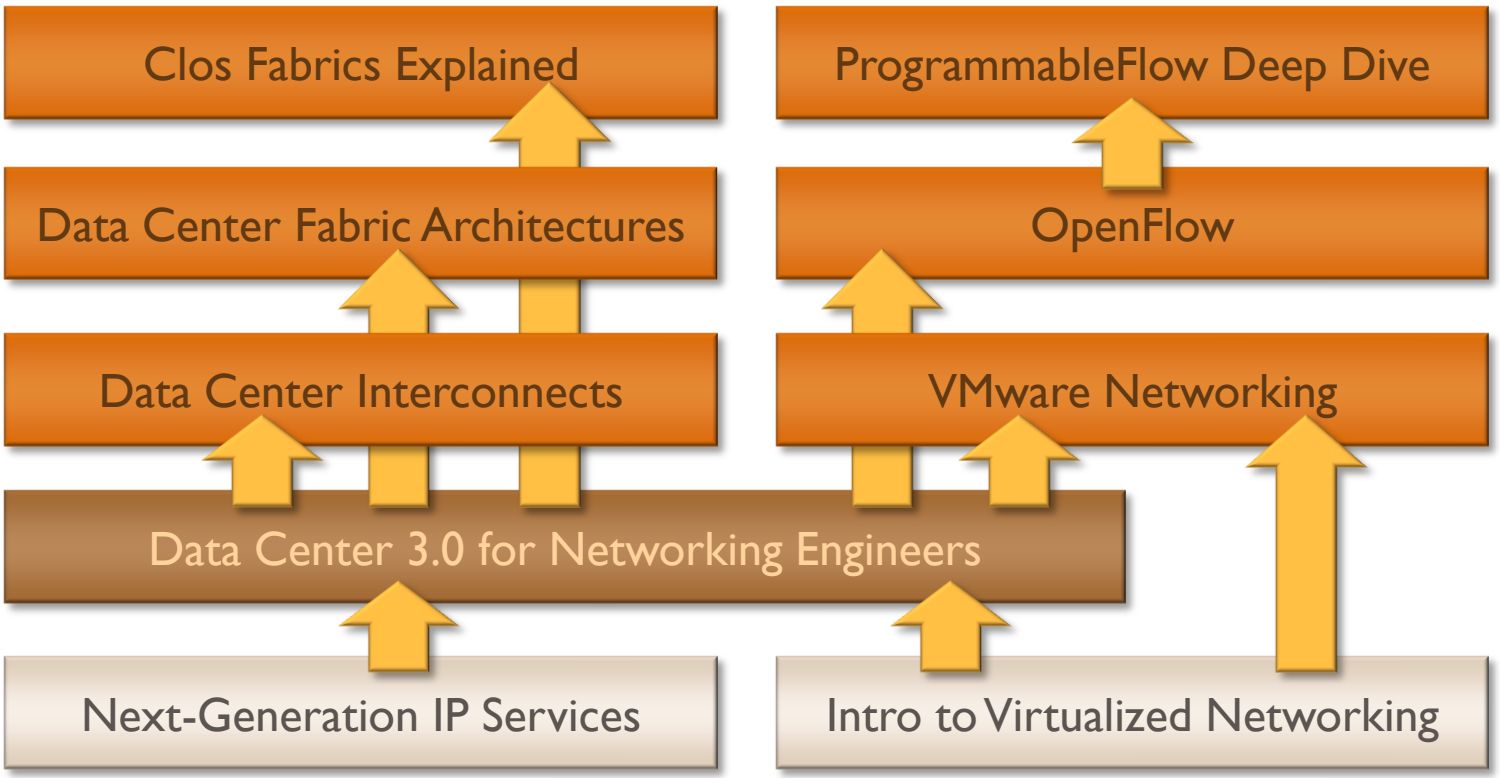
Focus: real-life applications of emerging technologies

- Large-scale data centers and network virtualization
- Networking solutions for cloud computing
- Scalable application design
- Core IP routing/MPLS, IPv6, VPN

CISCO CCIE EMERITUS

vmware vEXPERT

The Bigger Picture: Data Center Webinars



Availability

- Live sessions
- Recordings of individual webinars
- Yearly subscription

Other options

- Customized webinars
- ExpertExpress
- On-site workshops

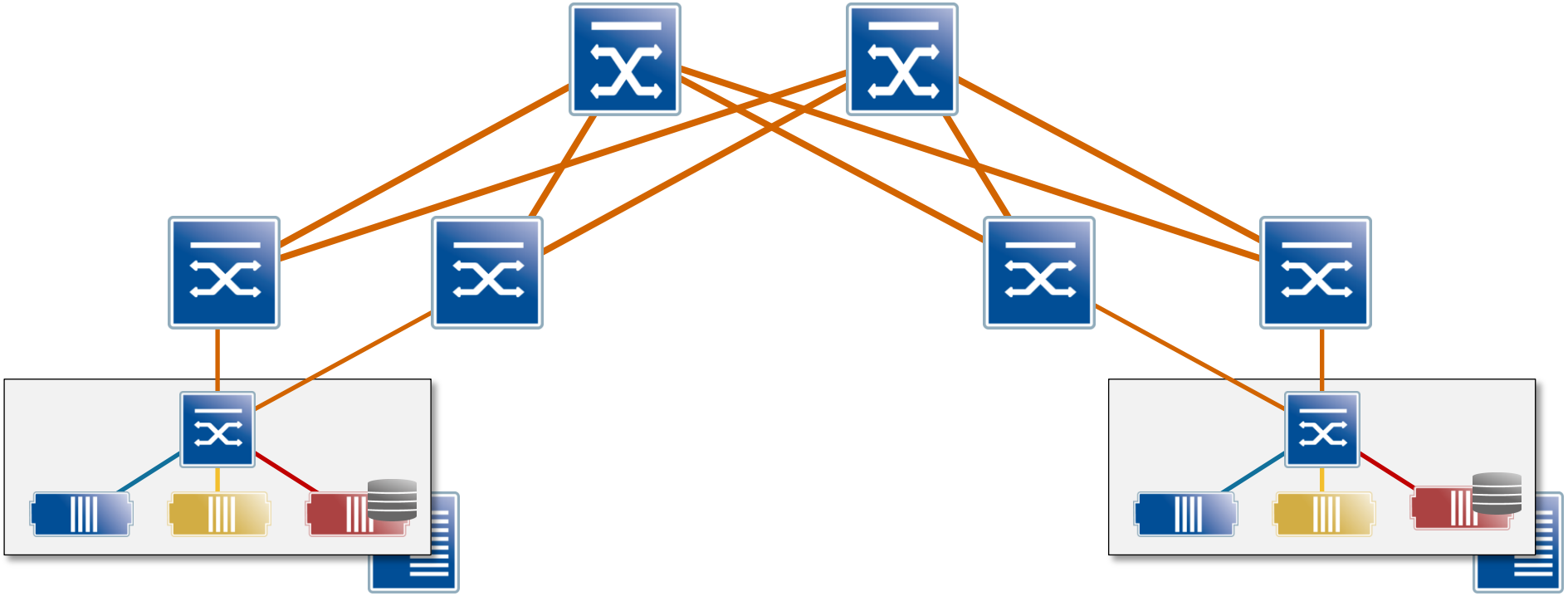
More information @ http://www.ipSpace.net/Roadmap/DC

Agenda

- VM mobility challenges
- Fabric and Host routing
- Data center interconnects
- VM mobility across layer-3 interconnects
- Integration with L4-7 network services
- Integration with overlay virtual networking

Fabric and Host Routing

Typical Enterprise Data Center Scenario



- Leaf & Spine Fabric
- Virtual networks built with VLANs
- Multiple subnets (security zones)
- Packet filters and firewalls

? L3 forwarding in core or ToR?

? Optimal egress and ingress flow?

? What happens after VM move?

? DC interconnects?

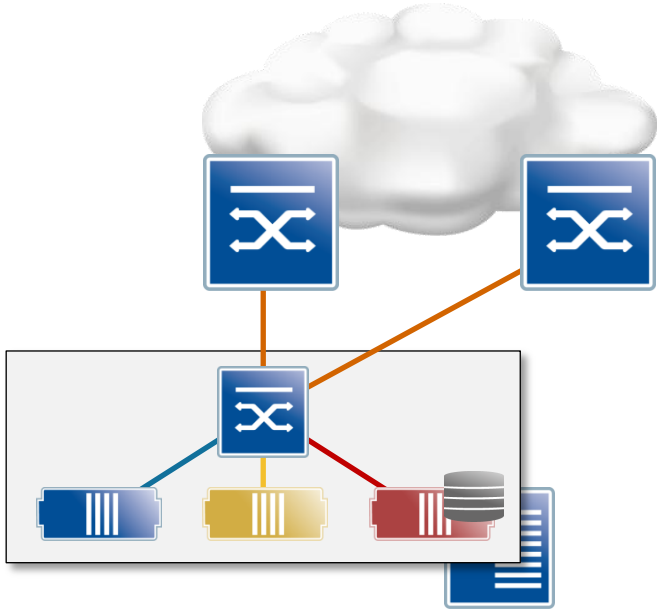
A Closer Look at the VM State

Interfaces

- IP address
- Subnet mask (on/off net)

ARP table

- IP-to-MAC mappings (intra-subnet hosts only)



Routing table

- Prefix-to-next hop
- Next hop must be on-net
- Usually just a default route (created from default gateway information)

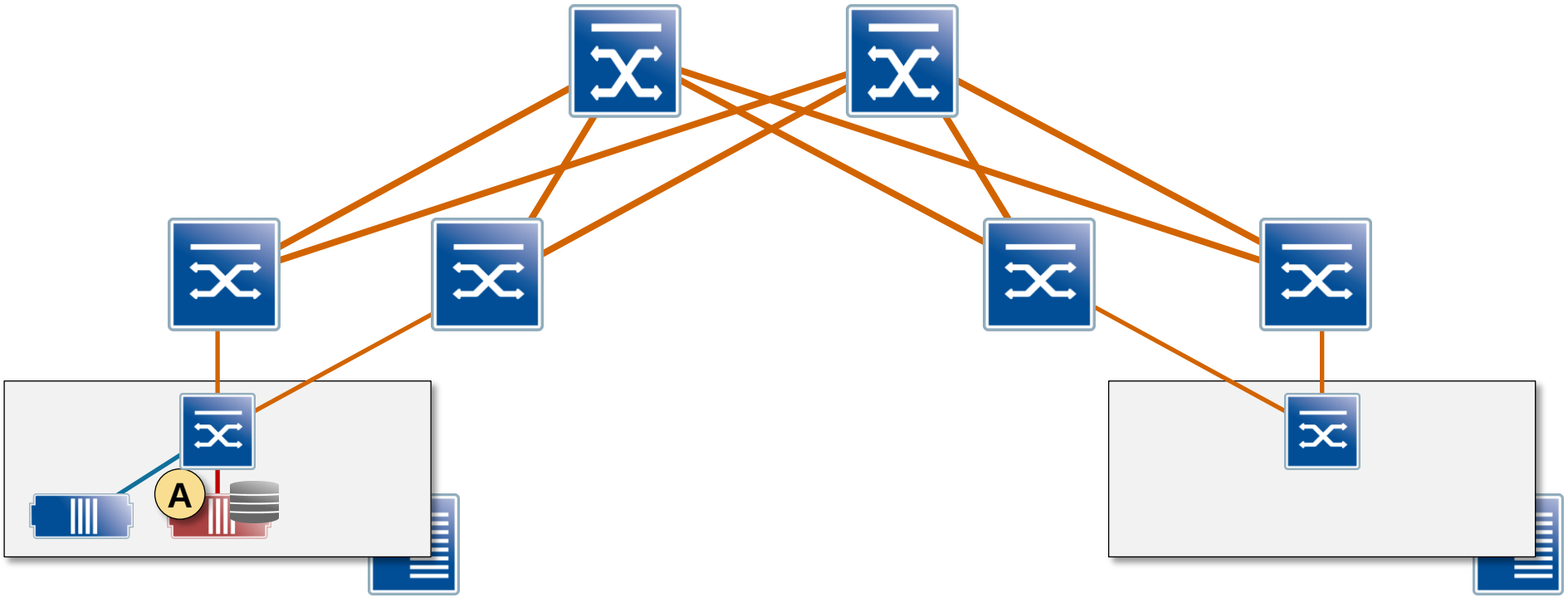
DNS resolution parameters

- DNS server(s), domain prefix, local DNS cache

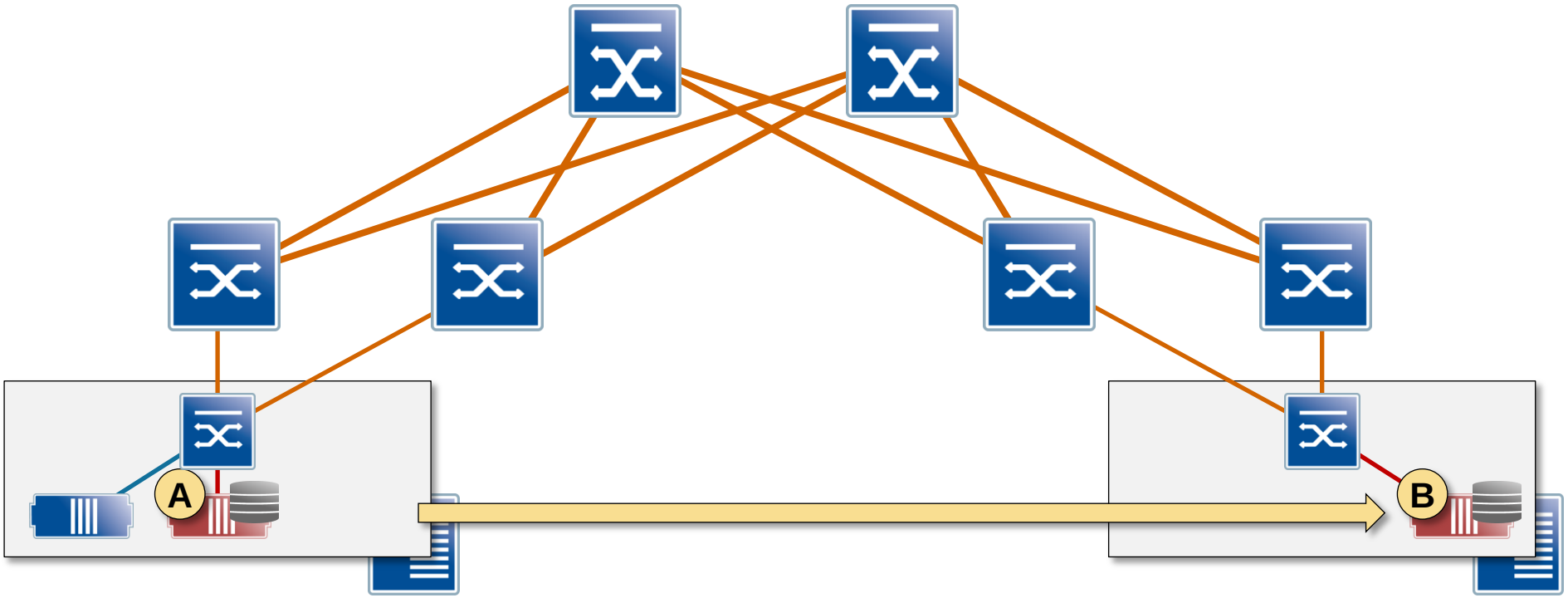
TCP connection table

- 5-tuple (local/remote address and port) + connection state

Live VM Mobility Requires Large VLANs

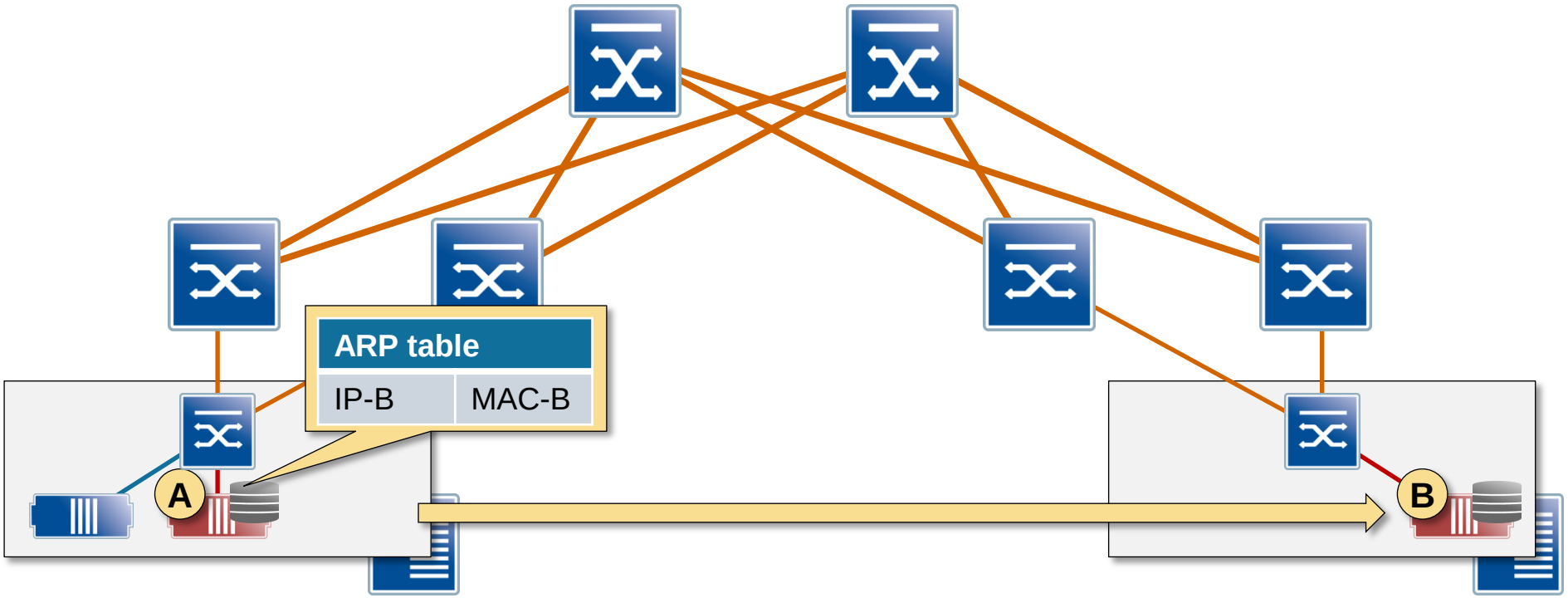


Live VM Mobility Requires Large VLANs



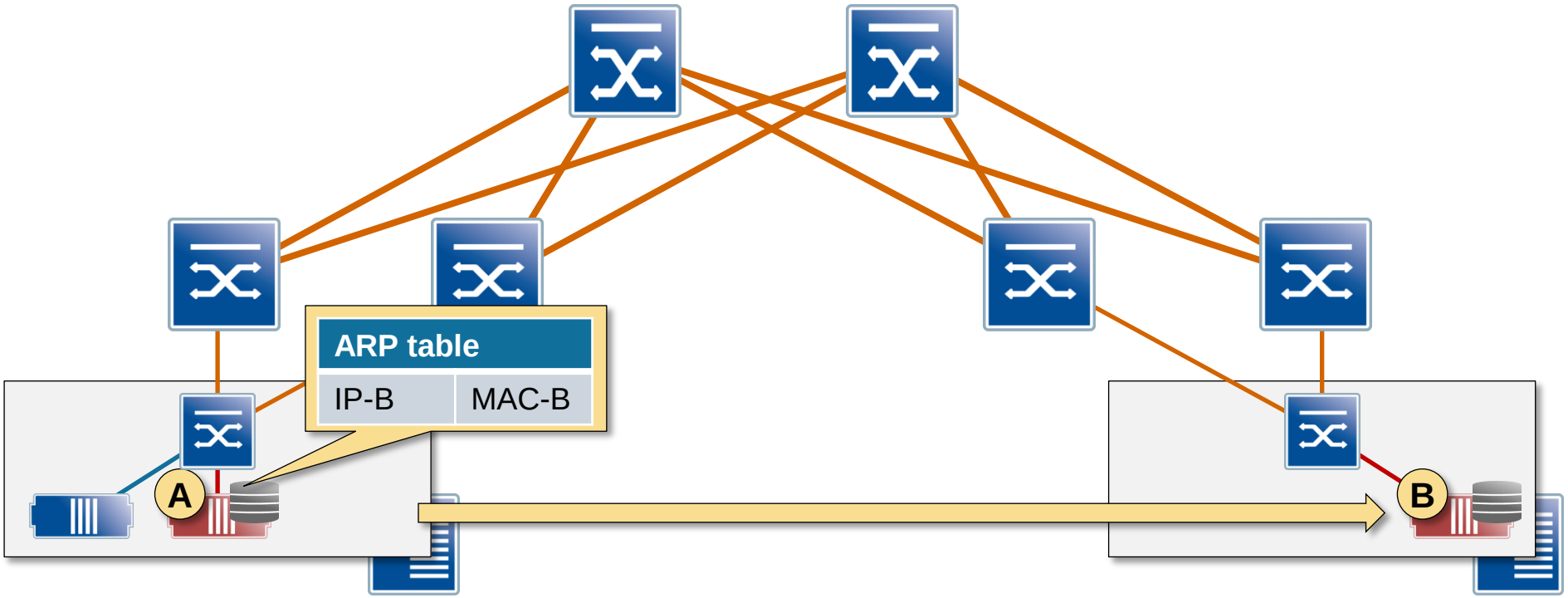
- VM B is moved (while running) to another physical server

Live VM Mobility Requires Large VLANs



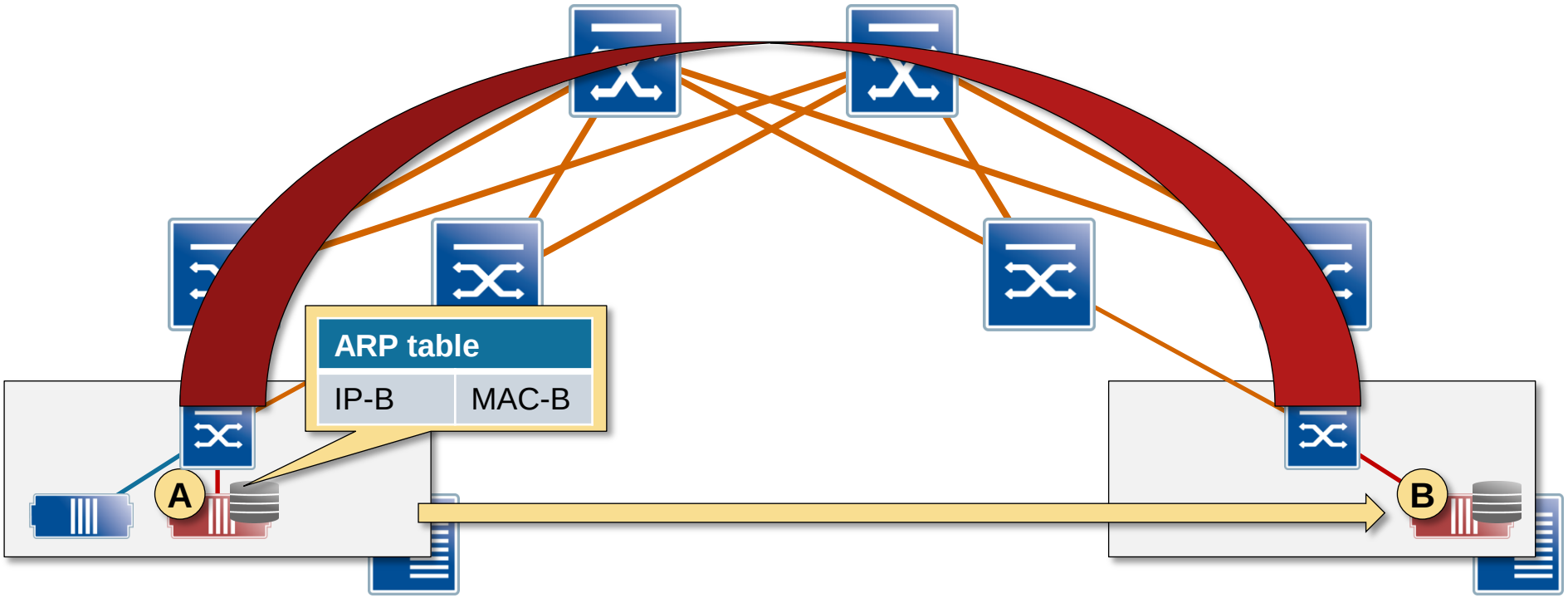
- VM B is moved (while running) to another physical server
- VM A has MAC address of VM B in its ARP cache

Live VM Mobility Requires Large VLANs



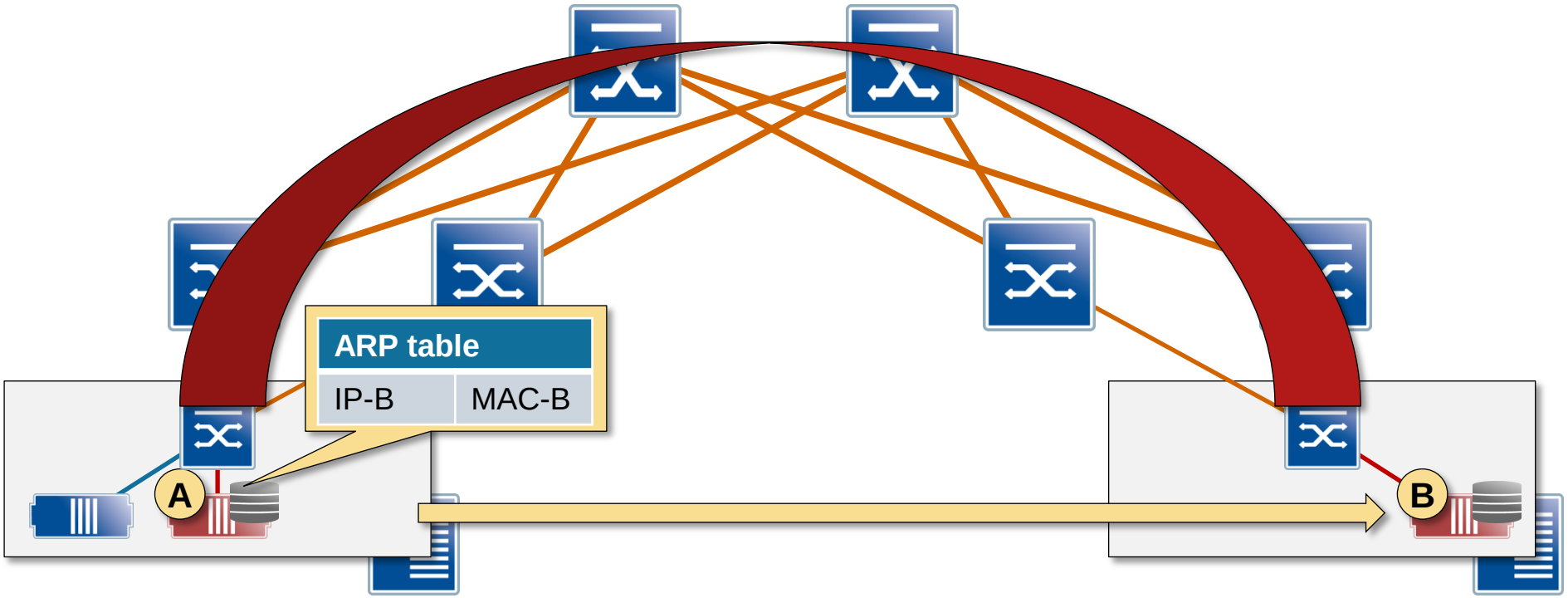
- VM B is moved (while running) to another physical server
- VM A has MAC address of VM B in its ARP cache
- We need direct layer-2 connectivity between A and B

Live VM Mobility Requires Large VLANs



- VM B is moved (while running) to another physical server
- VM A has MAC address of VM B in its ARP cache
- We need direct layer-2 connectivity between A and B
- VLAN (or overlay network) between source and target hypervisor hosts

Live VM Mobility Requires Large VLANs

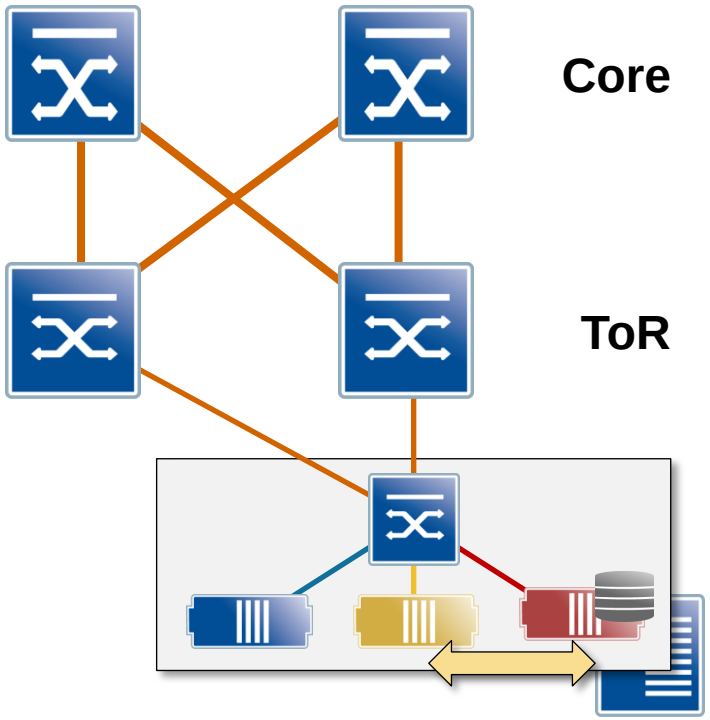


- VM B is moved (while running) to another physical server
- VM A has MAC address of VM B in its ARP cache
- We need direct layer-2 connectivity between A and B
- VLAN (or overlay network) between source and target hypervisor hosts

Remember: Layer-2 network = single failure domain

Layer-3 Forwarding with Large VLANs

- Yellow VM communicates with Red VM
- Different subnets ➔ Layer-3 forwarding

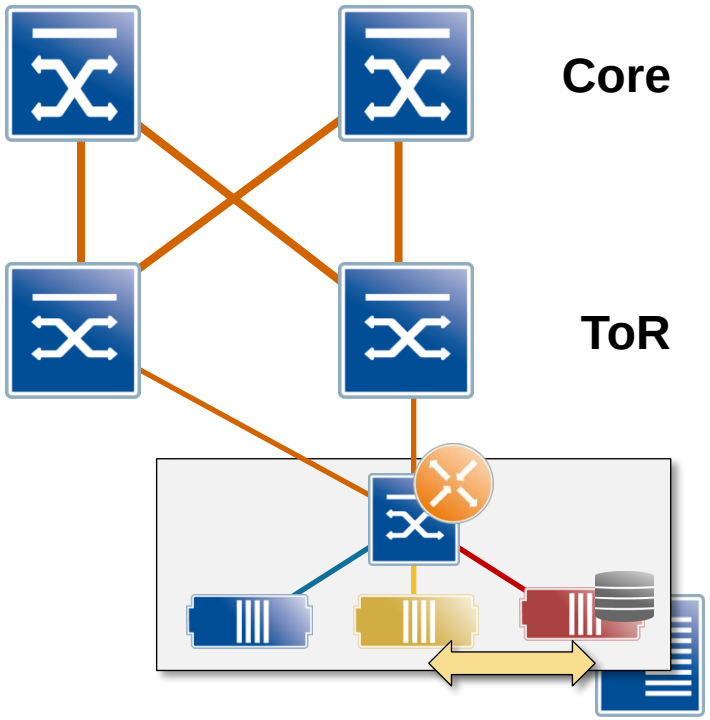


Layer-3 Forwarding with Large VLANs

- Yellow VM communicates with Red VM
- Different subnets ➔ Layer-3 forwarding

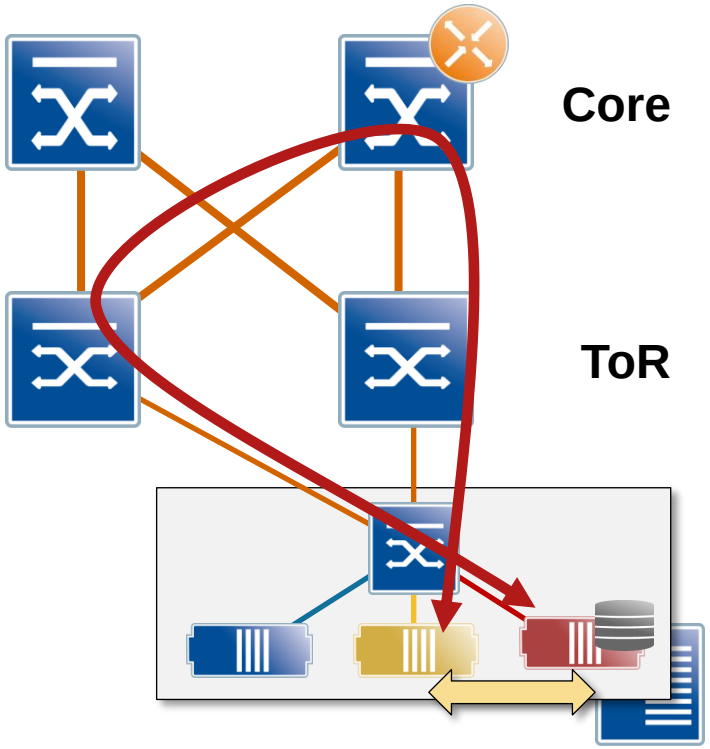
Potential solutions

- L3 forwarding in hypervisor vSwitch
  → not yet available



Layer-3 Forwarding with Large VLANs

- Yellow VM communicates with Red VM
- Different subnets ➔ Layer-3 forwarding

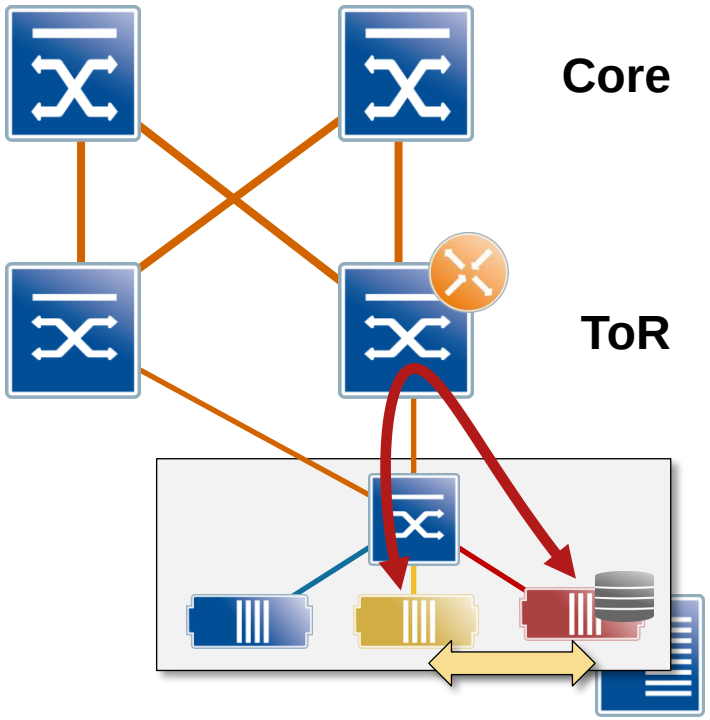


Potential solutions

- L3 forwarding in hypervisor vSwitch
  → not yet available
- L3 forwarding in core switches
  → unnecessary latency
  → ToR-to-Core links wasted

Layer-3 Forwarding with Large VLANs

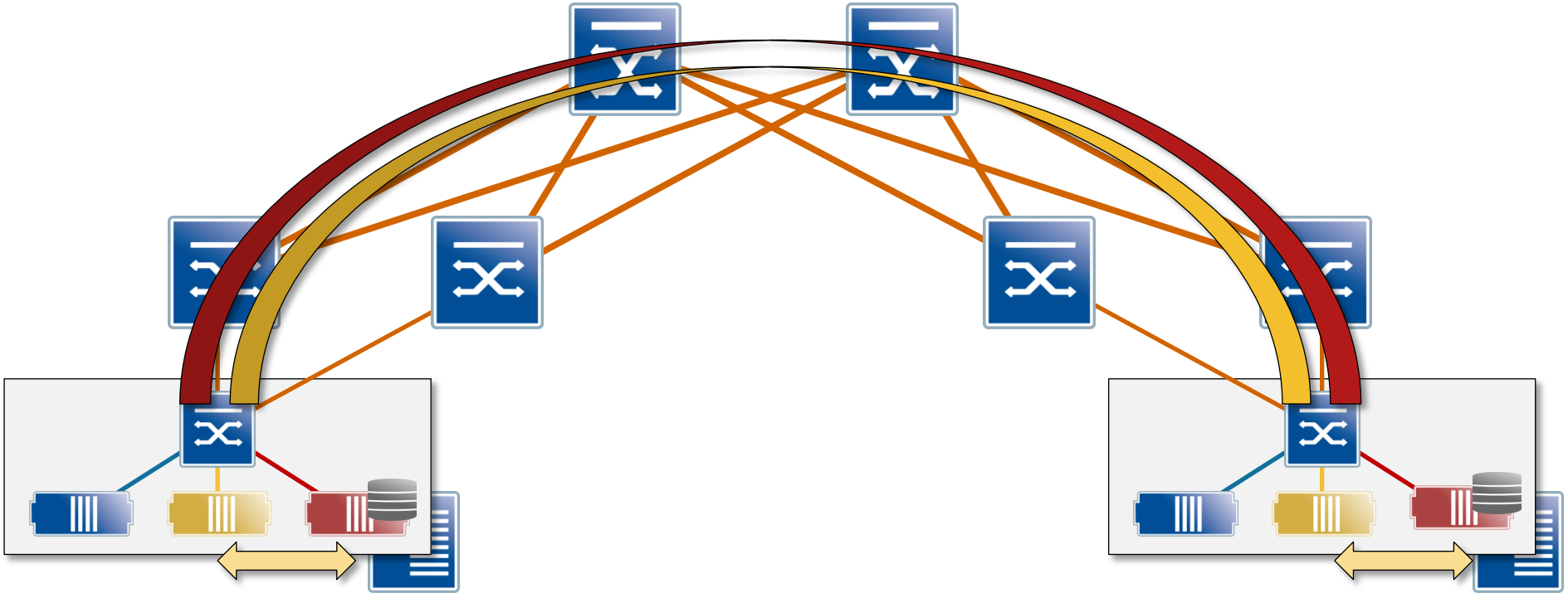
- Yellow VM communicates with Red VM
- Different subnets ➔ Layer-3 forwarding



Potential solutions

- L3 forwarding in hypervisor vSwitch
  → not yet available
- L3 forwarding in core switches
  → unnecessary latency
  → ToR-to-Core links wasted
- L3 forwarding in ToR switches
  → best of both worlds
  → requires **optimal inter-subnet forwarding**

Optimal Inter-Subnet Forwarding Explained

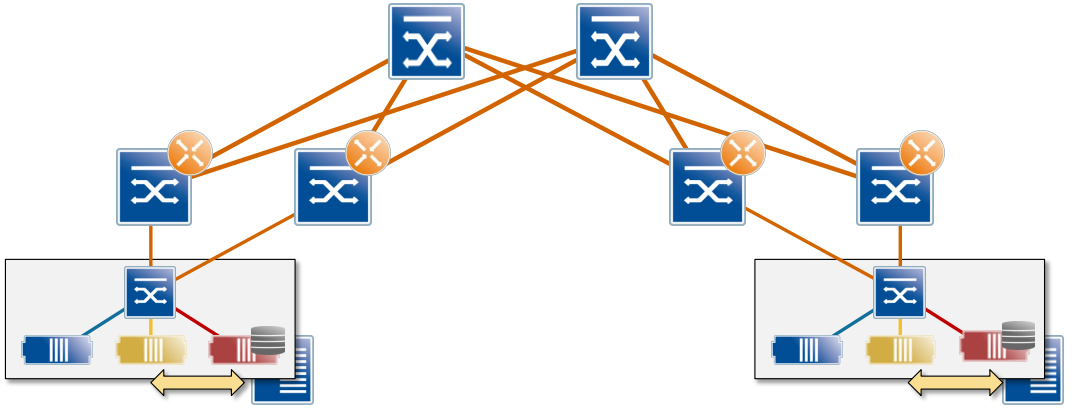


- Red and Yellow VLANs (and IP subnets) are stretched across ToR switches
- Which ToR switch should do L3 forwarding?
- The only good answer: **all of them**
- ToR switches **must share first-hop IP and MAC address**
  (no need to share configuration)

Solution Space

Active-active MLAG forwarding

- L3 forwarding at core or limited to two ToR switches



ToR stacking

- Virtual chassis, IRF ...
- Suboptimal E-W traffic flow (within the stack, not over core)

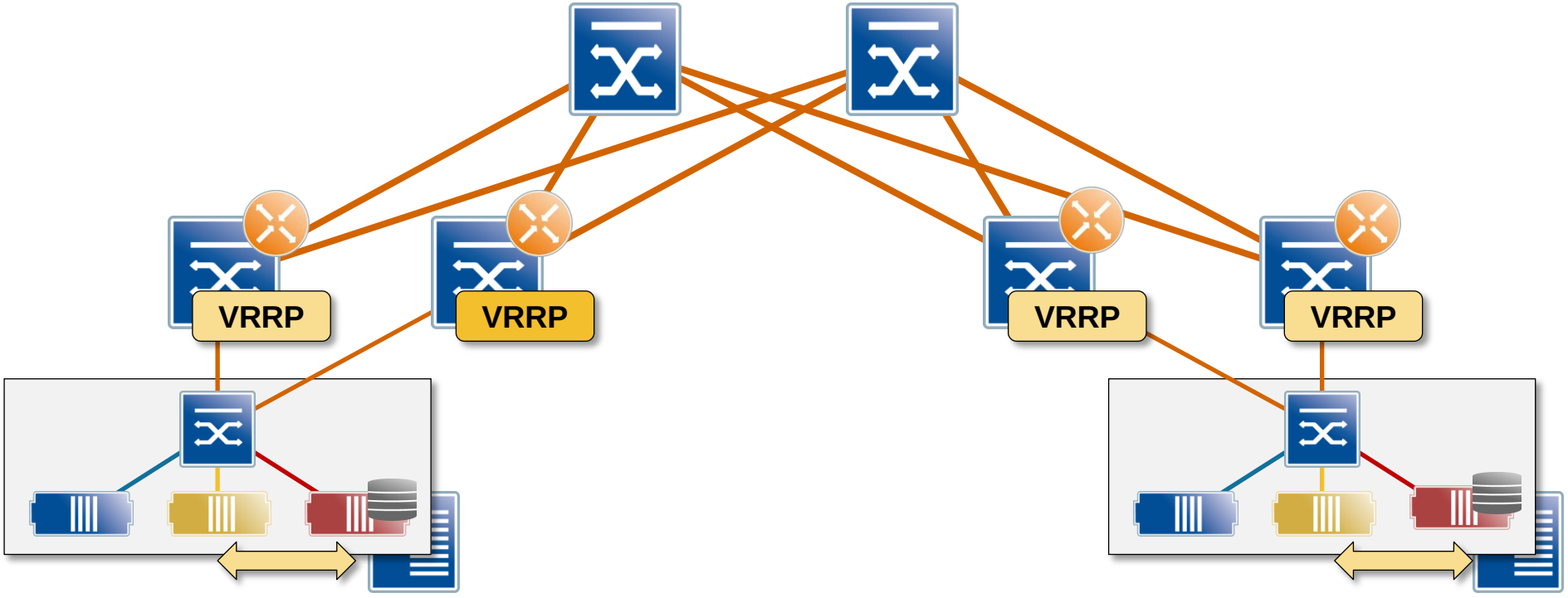
Single logical device

- QFabric, NEC ProgrammableFlow
- Single management point ➔ single failure domain

Just-Do-It

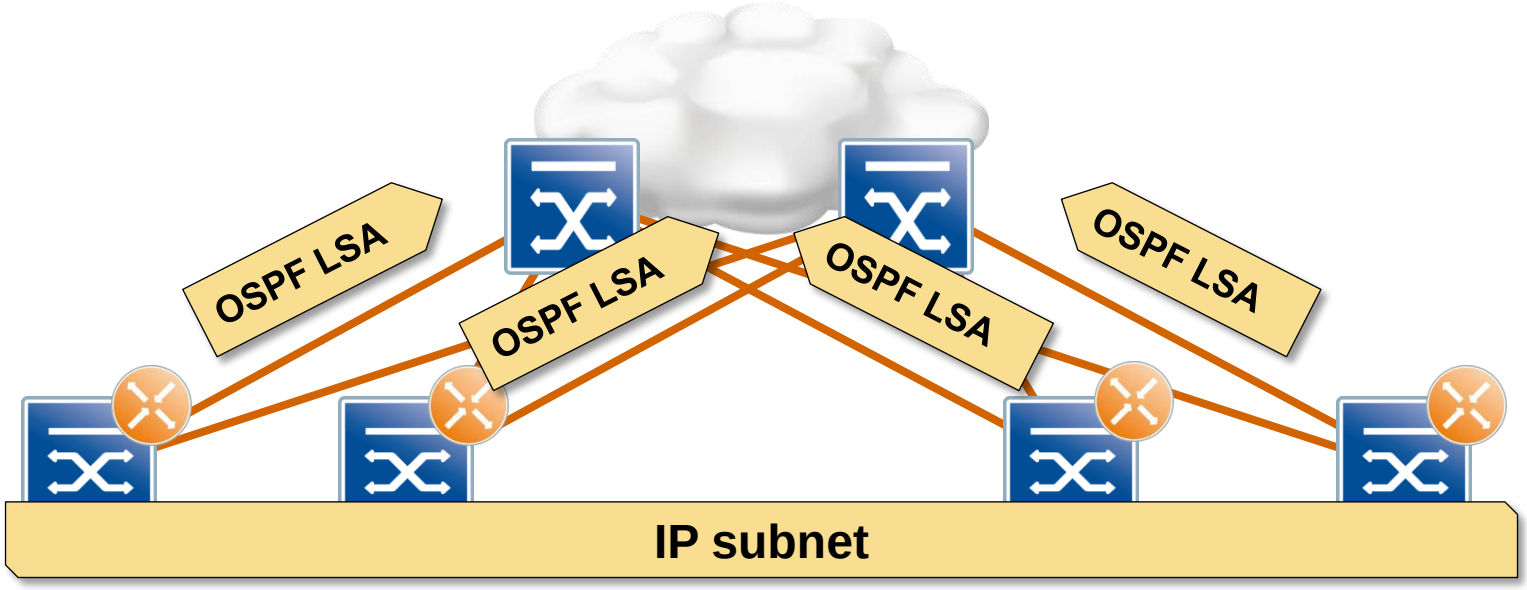
- Virtual ARP: Same IP and MAC address configured on multiple switches
- Requires careful management or automation/orchestration system

Enterasys Fabric Routing



- VRRP is configured on ToR switches (optional: core switches)
- One ToR switch becomes VRRP master
- All ToR switches share VRRP MAC address
  ➔ no MAC learning, active/active VRRP across the whole fabric
- First-hop (ingress) ToR switch performs L3 forwarding ➔ optimal traffic flow

Optimal Network-to-VM Forwarding – The Problem

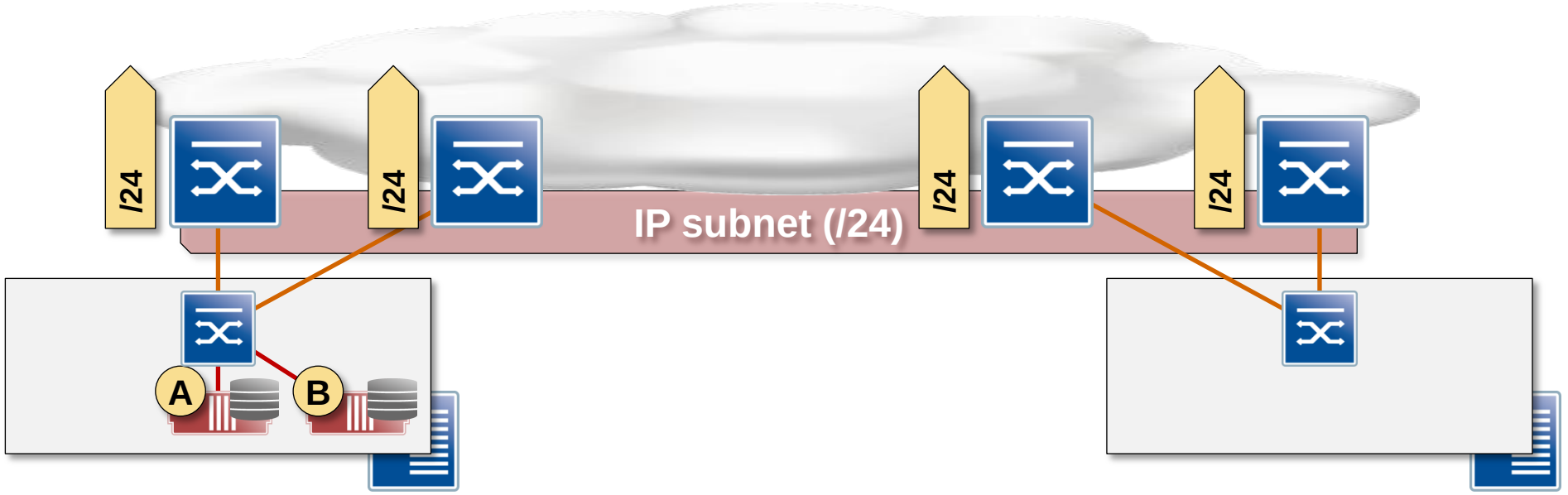


- The same subnet is configured on all ToR switches
- ToR switches advertise the subnet to core (or edge) routers
- Core routes have N equal-cost paths
  → packet toward a host (or VM) could take any one of those paths

Probability of a misrouted flow: $\frac{N-1}{N}$

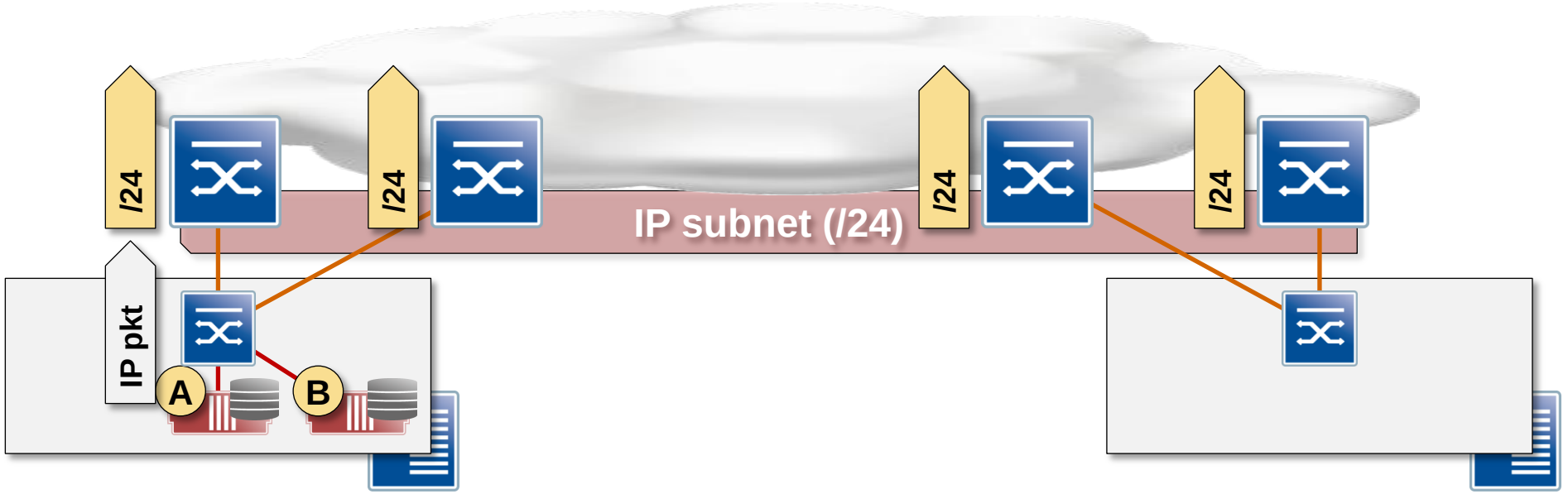
Typical solution: configure the same subnet on core switches

Enterasys Host Routing



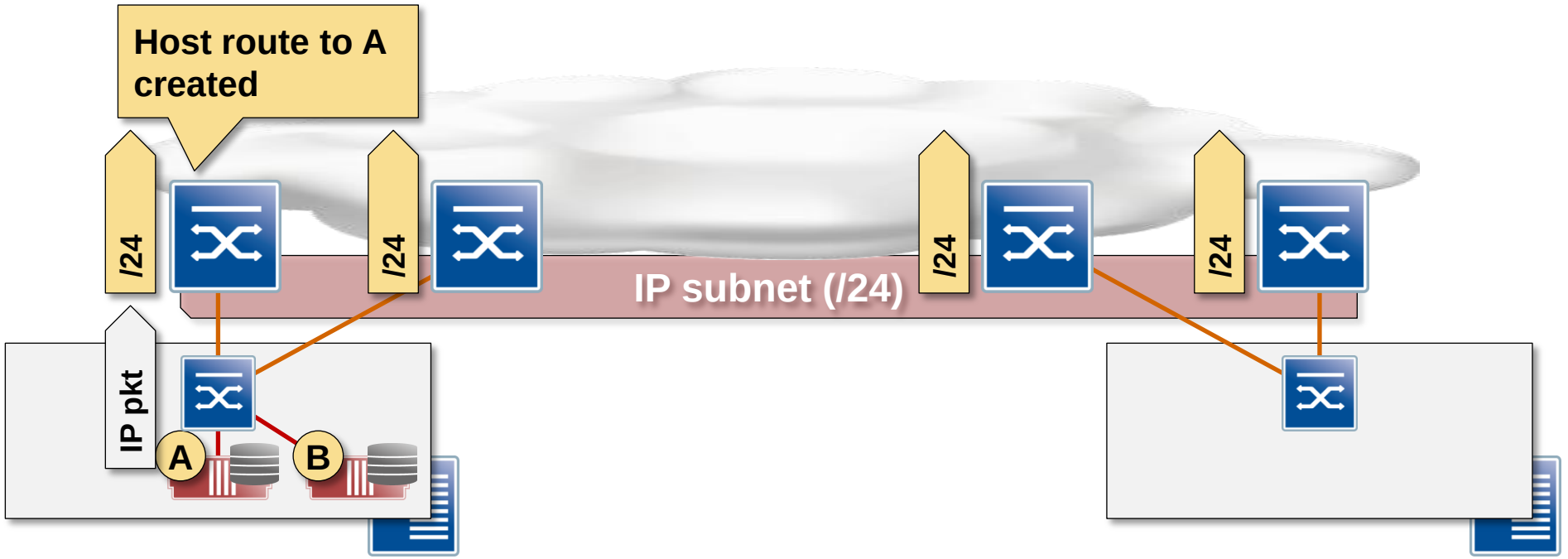
- All ToR switches advertise the prefix to shared subnet

Enterasys Host Routing



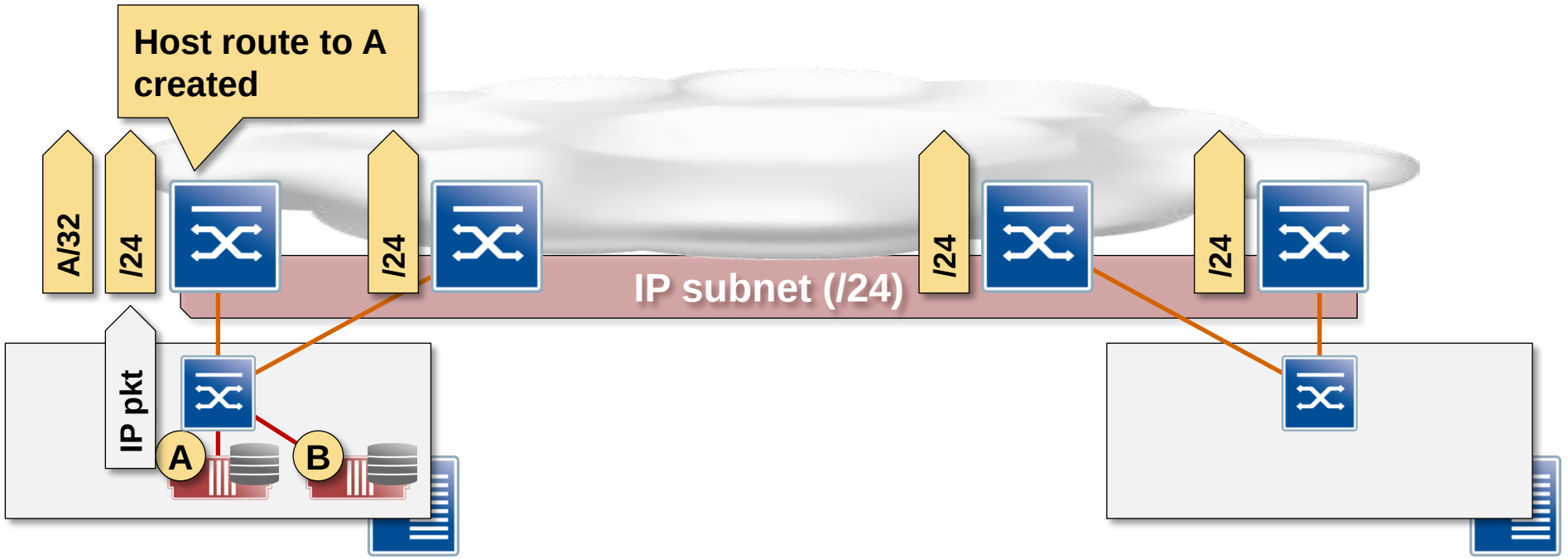
- All ToR switches advertise the prefix to shared subnet

Enterasys Host Routing



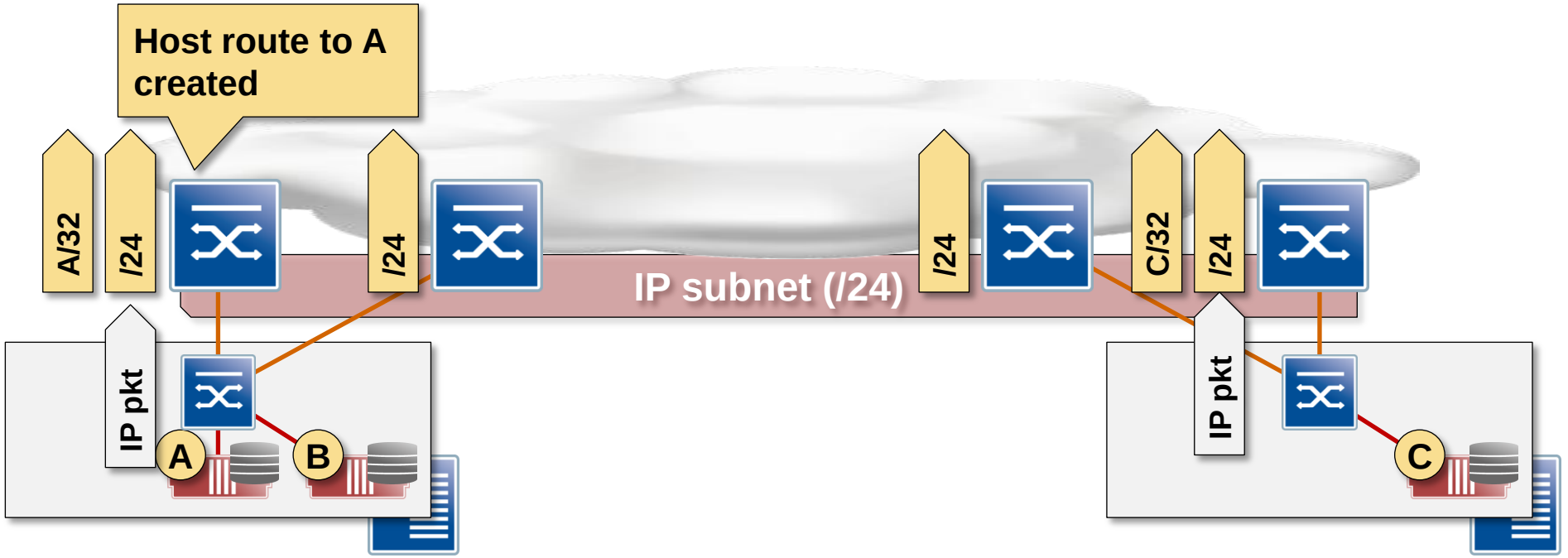
- All ToR switches advertise the prefix to shared subnet
- ToR switch creates a host route toward directly attached host based on inbound traffic and interface configuration

Enterasys Host Routing



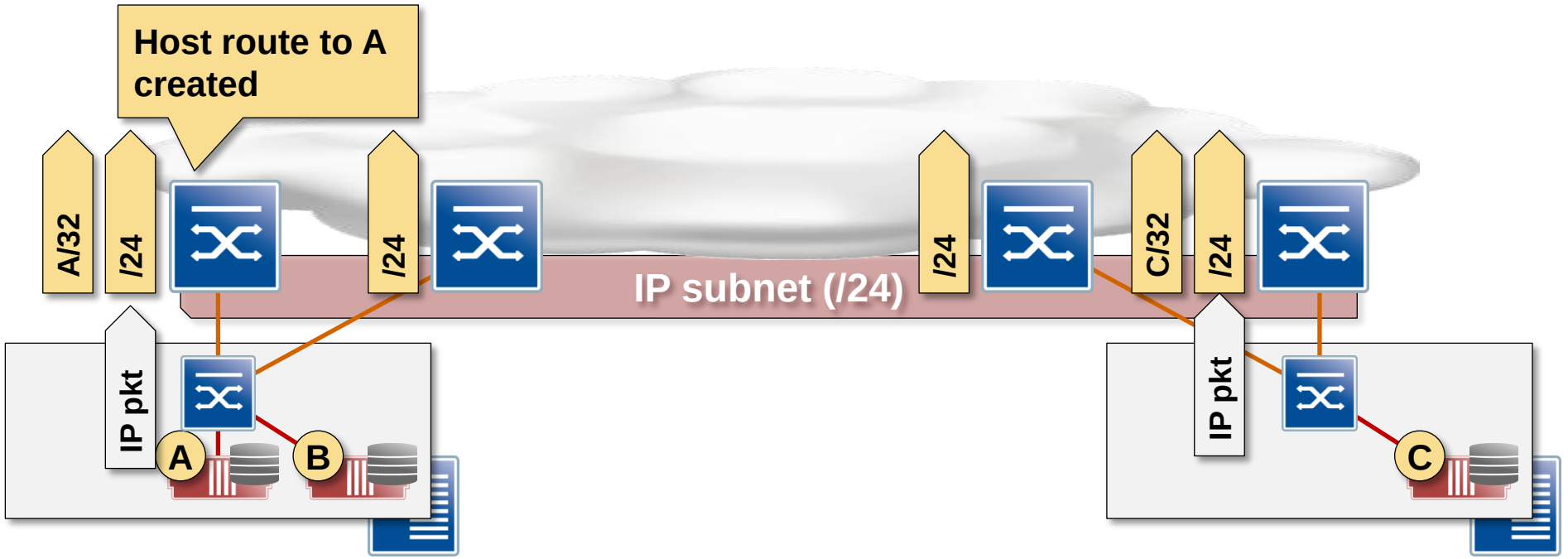
- All ToR switches advertise the prefix to shared subnet
- ToR switch creates a host route toward directly attached host based on inbound traffic and interface configuration
- Host route (/32) is redistributed into routing protocol(s)

Enterasys Host Routing



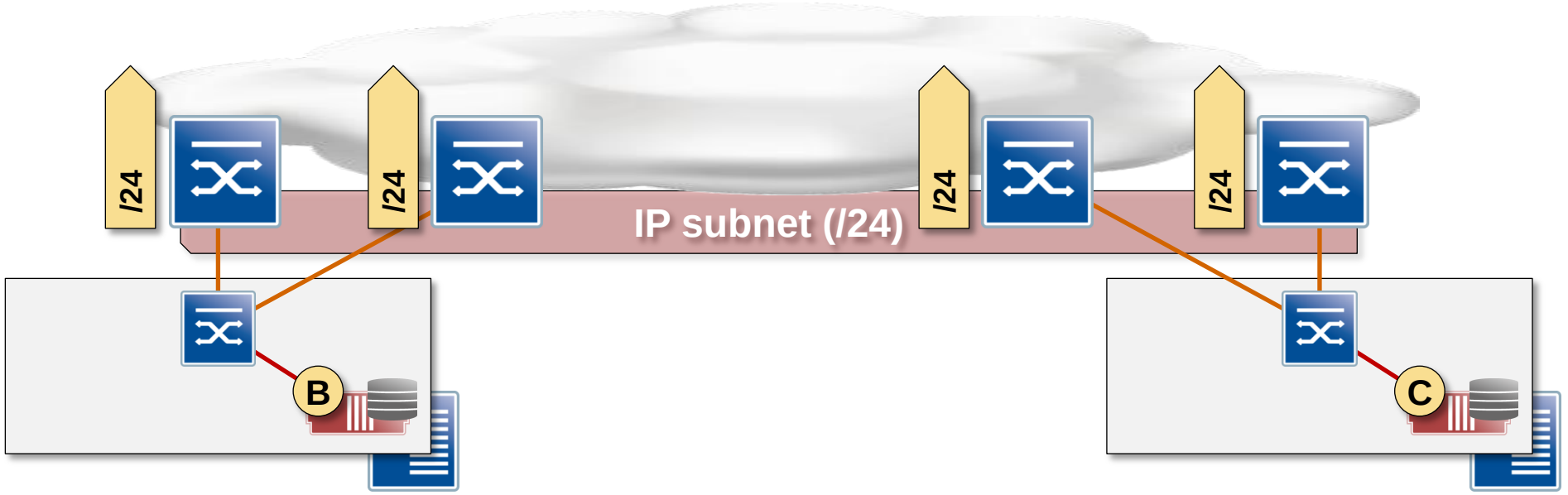
- All ToR switches advertise the prefix to shared subnet
- ToR switch creates a host route toward directly attached host based on inbound traffic and interface configuration
- Host route (/32) is redistributed into routing protocol(s)

Enterasys Host Routing

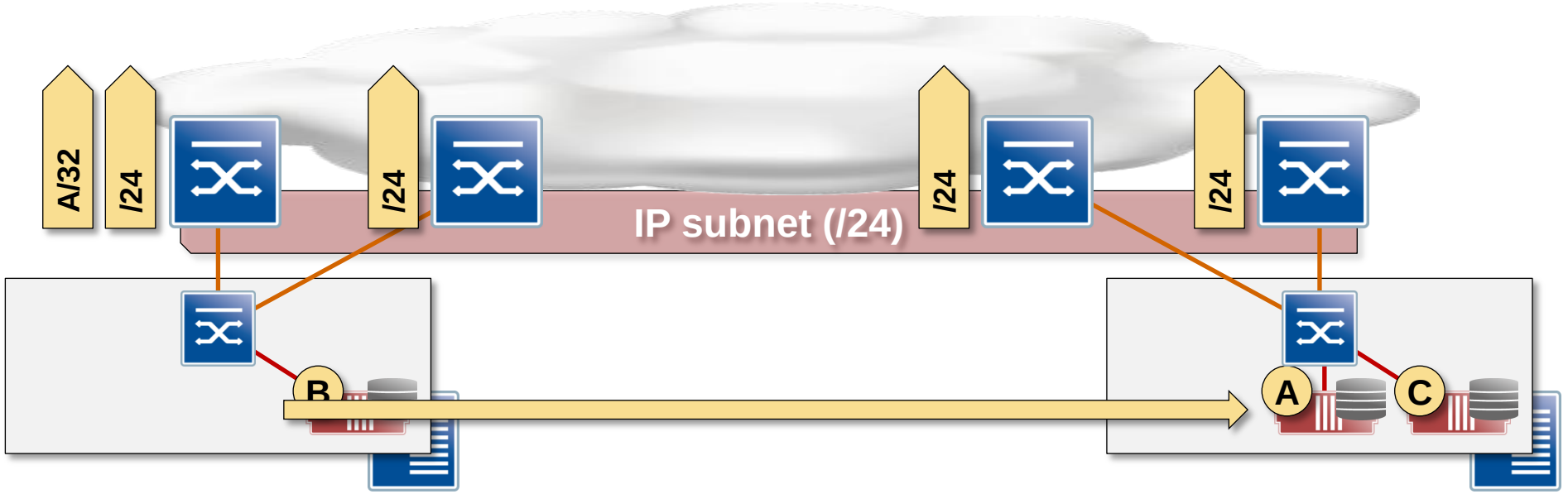


- All ToR switches advertise the prefix to shared subnet
- ToR switch creates a host route toward directly attached host based on inbound traffic and interface configuration
- Host route (/32) is redistributed into routing protocol(s)
- Every L3 switch has optimal path(s) toward all IP hosts

Host Routing and VM Mobility

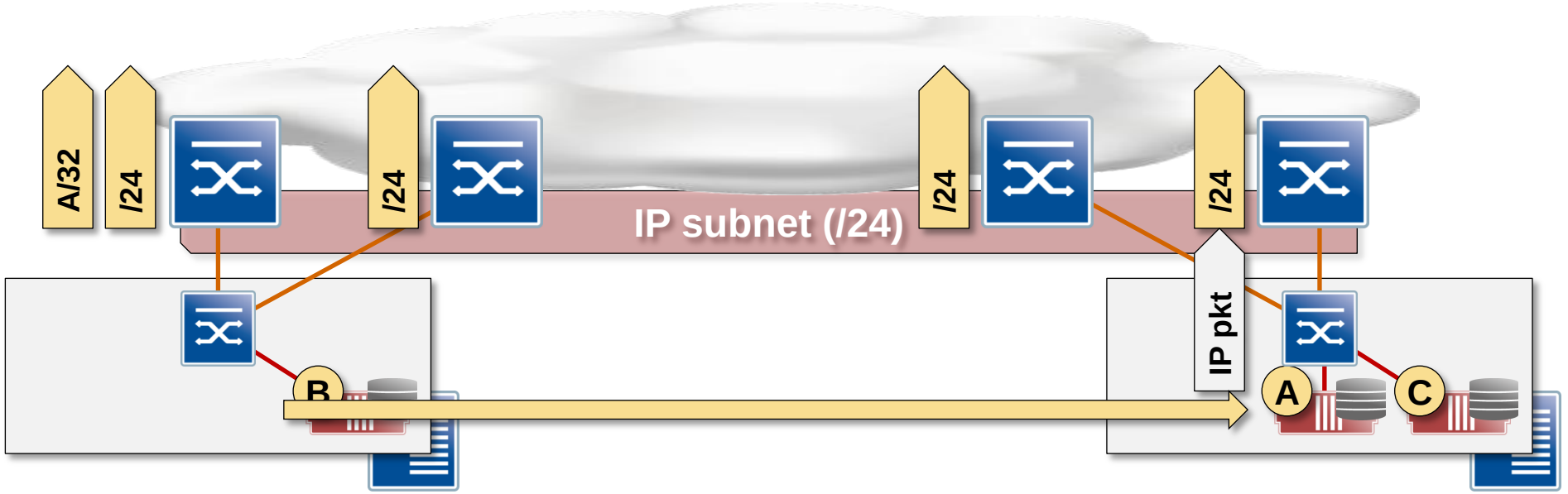


Host Routing and VM Mobility



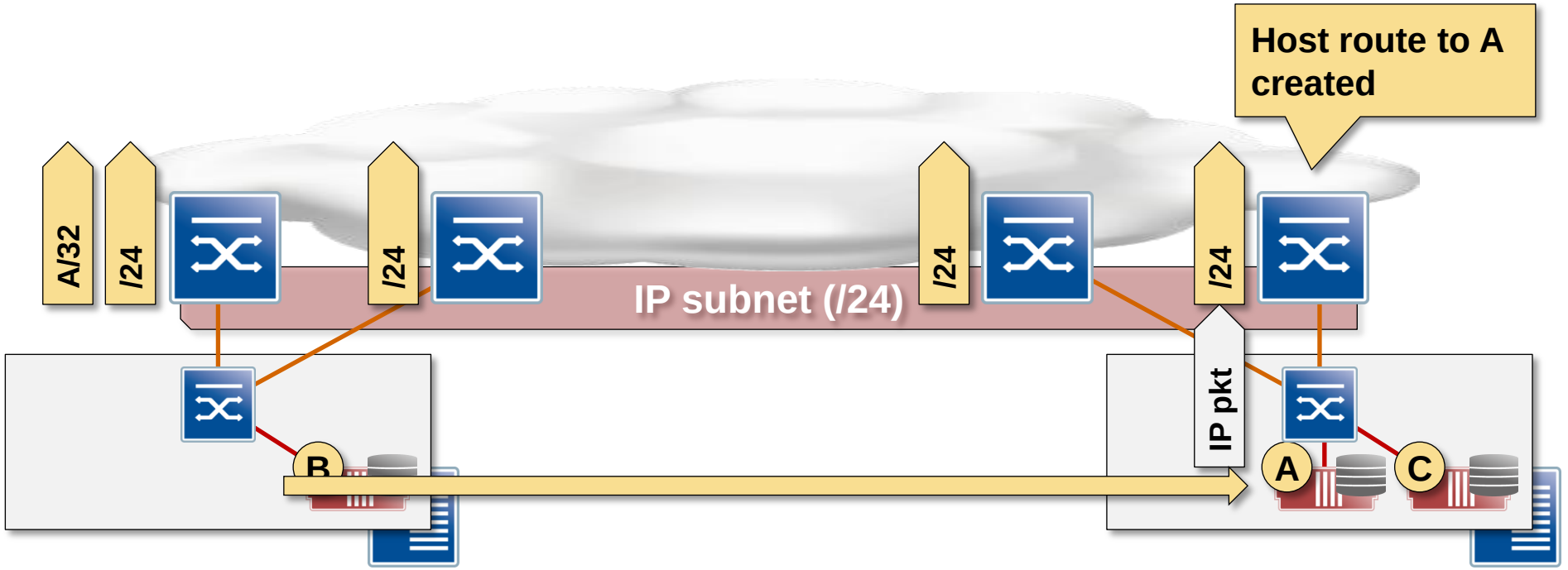
- VM A is moved to another hypervisor

Host Routing and VM Mobility



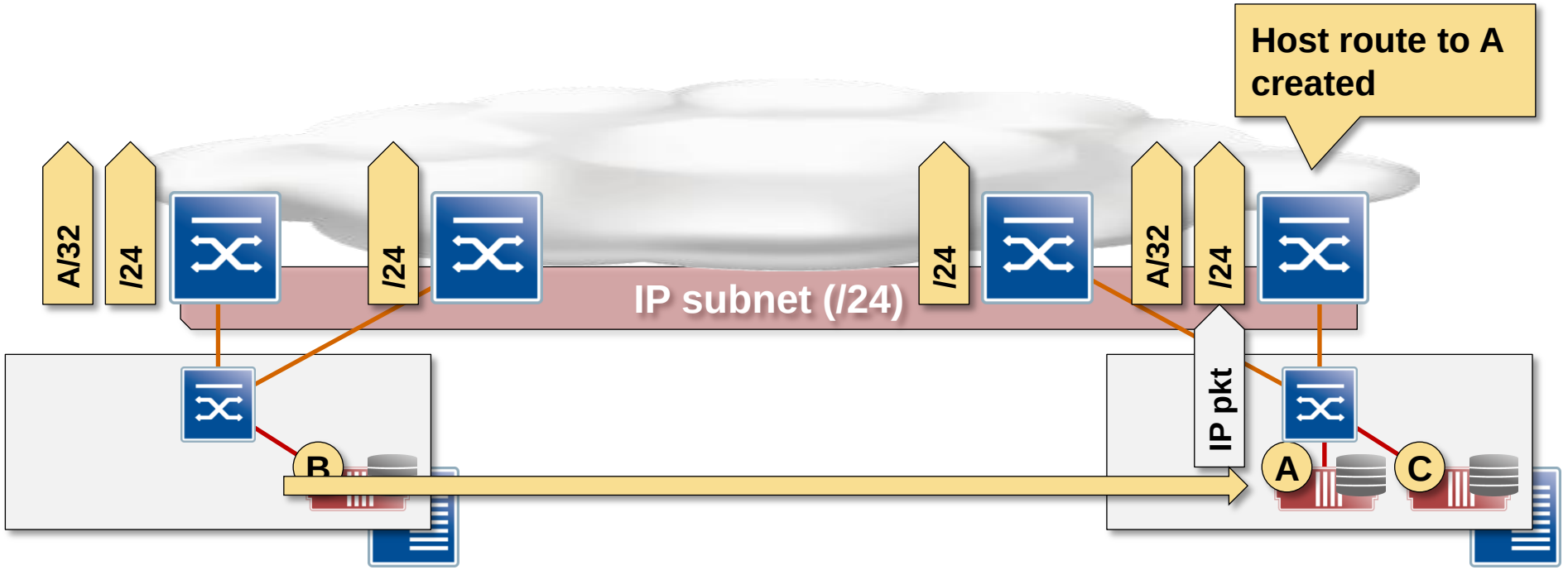
- VM A is moved to another hypervisor
- VM A sends a data packet

Host Routing and VM Mobility



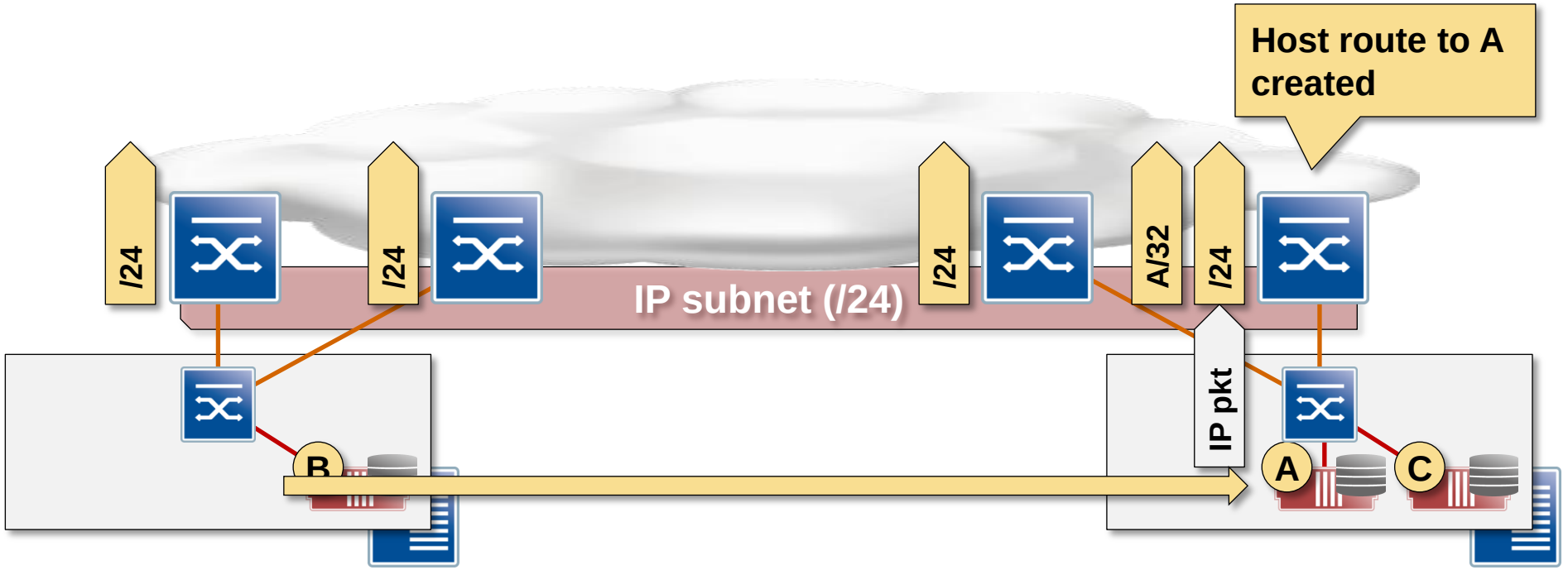
- VM A is moved to another hypervisor
- VM A sends a data packet
- New ToR switch creates a host route

Host Routing and VM Mobility



- VM A is moved to another hypervisor
- VM A sends a data packet
- New ToR switch creates a host route
- Host route A/32 is redistributed into routing protocol(s)
  → temporary suboptimal routing

Host Routing and VM Mobility



- VM A is moved to another hypervisor
- VM A sends a data packet
- New ToR switch creates a host route
- Host route A/32 is redistributed into routing protocol(s)
  → temporary suboptimal routing
- Old ToR switch ages out the host route and revokes it from routing protocol(s)

Host Routing and Security

- Spoofed IP packets could result in DoS or traffic hijack attacks
- Layer-2 security is mandatory for stable host routing

Options

- Source MAC and IP address checks in hypervisors
- Dynamic ARP inspection and IP Source Guard on ToR switches

How Far Did We Get?

Fabric routing: optimal server-to-network routing

Host routing: optimal network-to-server routing

How Far Did We Get?

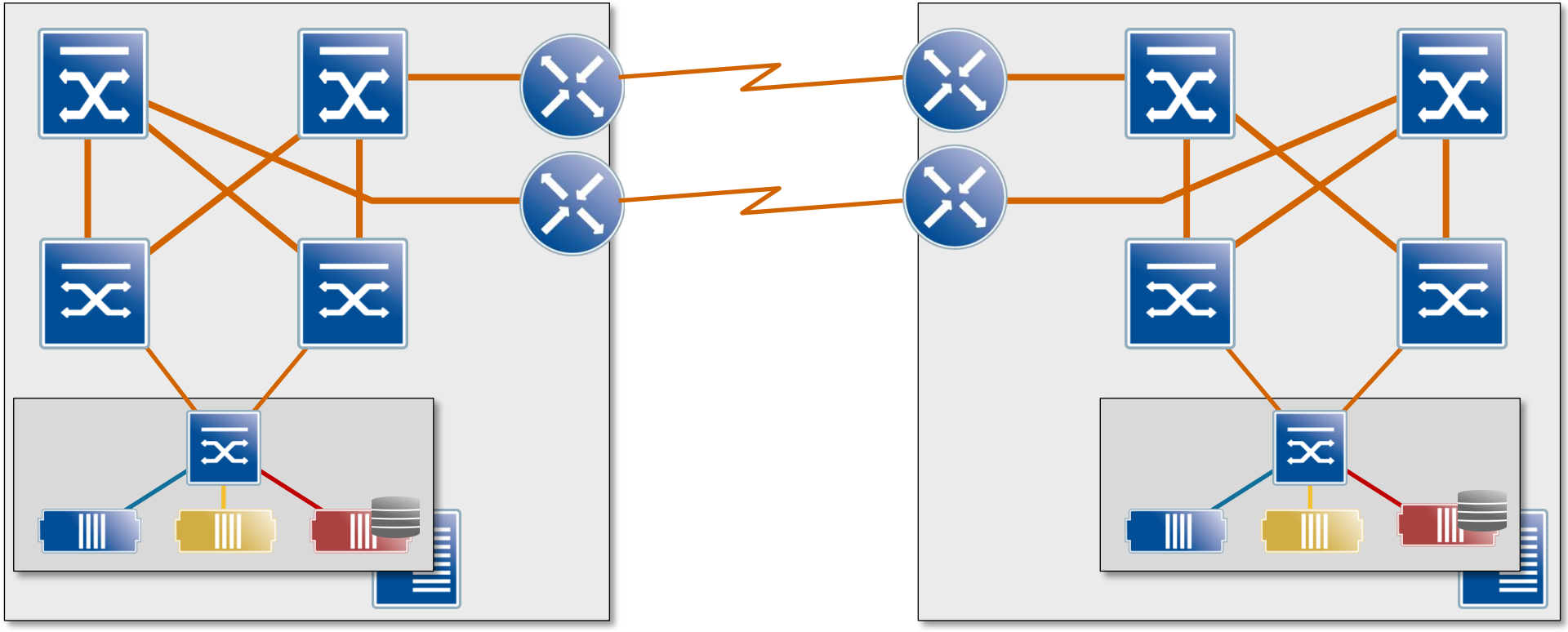
Fabric routing: optimal server-to-network routing

Host routing: optimal network-to-server routing

Can we use them to build robust Data Center Interconnects?

Data Center Interconnects

Data Center Interconnect Scenarios



L3 interconnect: pure IP routing

L2 interconnect: VLANs stretched between locations

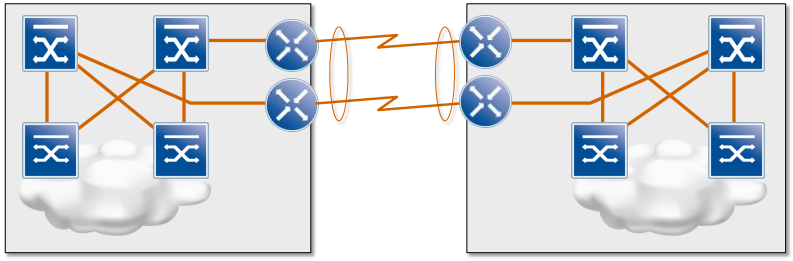
L2 interconnect scenarios:

- Single-subnet applications (iSCSI replications, clusters)
- VM mobility (cold or hot)

Layer-2 DCI – Potential Use Cases

iSCSI replication

- Required by some storage vendors
- No feasible workaround



Stretched clusters

- Don’t use – most clustering solutions provide L3/DNS-based alternatives

Live VM migration

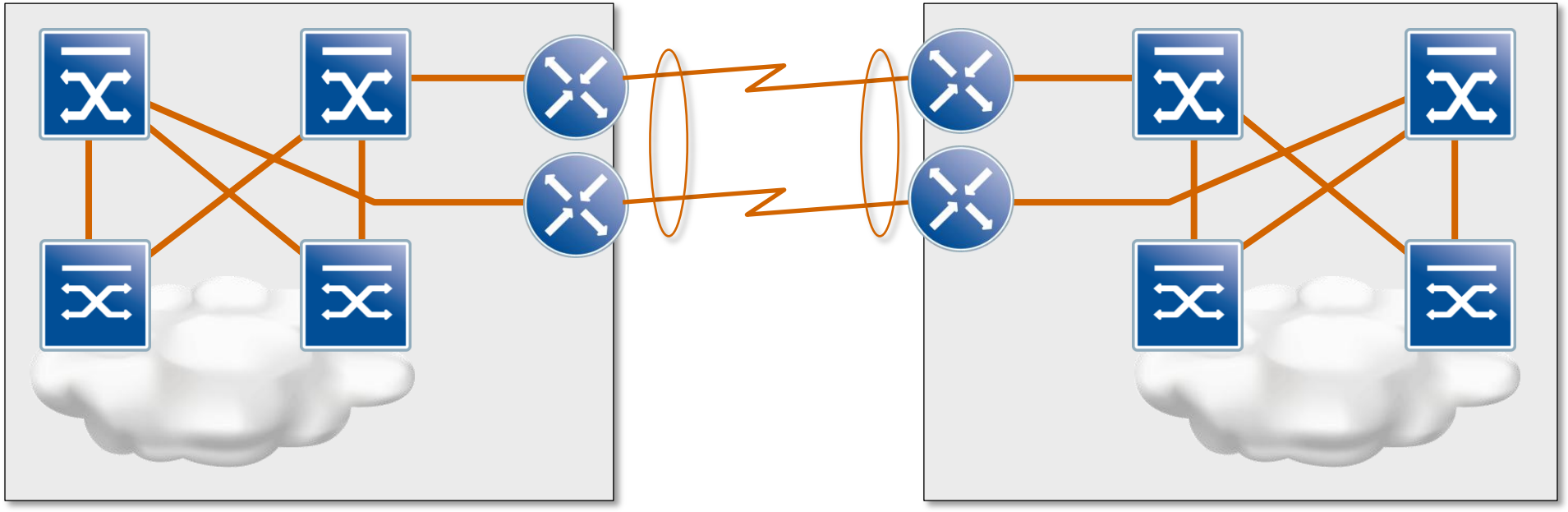
- Mostly impractical – increased latency, bandwidth requirements
- Useful in temporary well controlled migration scenarios

Cold VM migration without IP address change

- Required by some badly written applications
- Sometimes simplifies disaster recovery procedures
- Try to avoid and rely on DNS

Always consider the impact of full DCI link failure

Enterasys L2 DCI – Yesterday

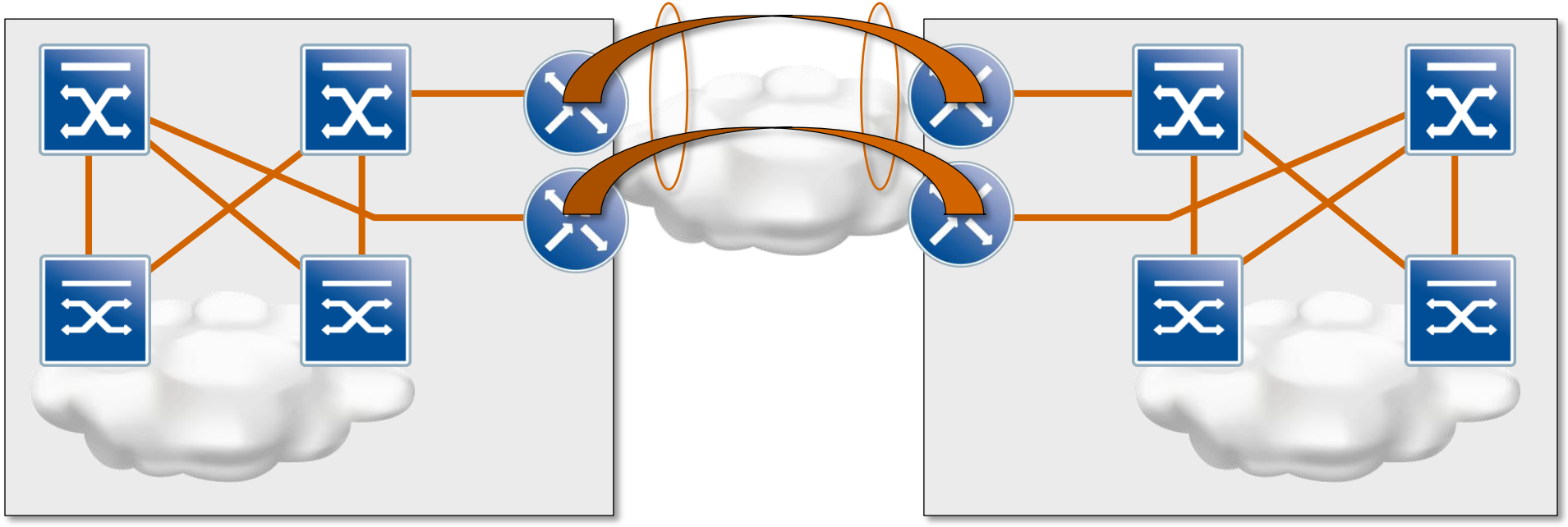


Same solution as most other vendors

- Virtual Switch Bonding (VSB) on WAN edge switches
- MLAG across WAN link
- Multiple MST regions

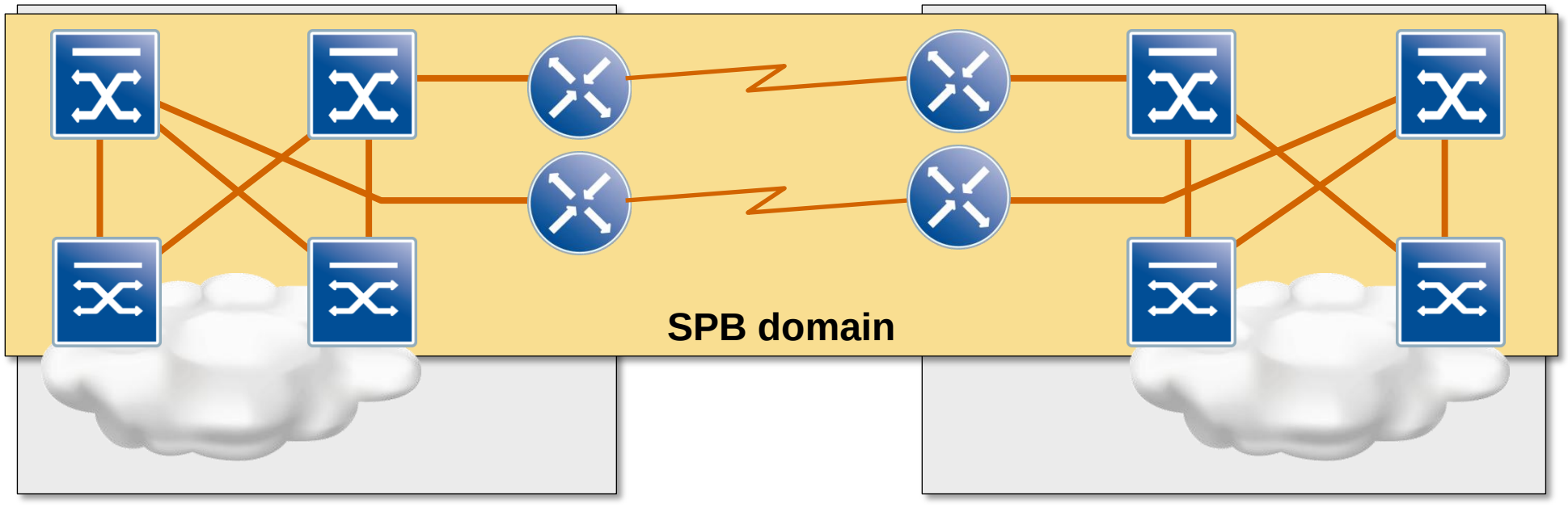
Single failure domain. Don't use for more than 2 sites.

Enterasys L2 DCI – Virtual Private Port Services over GRE Transport



- GRE Layer 2 tunnels between WAN edge switches
- RSTP/MSTP over GRE tunnels (blocks one tunnel or one LAN interface) or VSB + MLAG with LACP
- Somewhat simpler to implement and troubleshoot than VPLS or E-VPN
- Still hard to use at more than two sites

Enterasys L2 DCI with SPB(V)

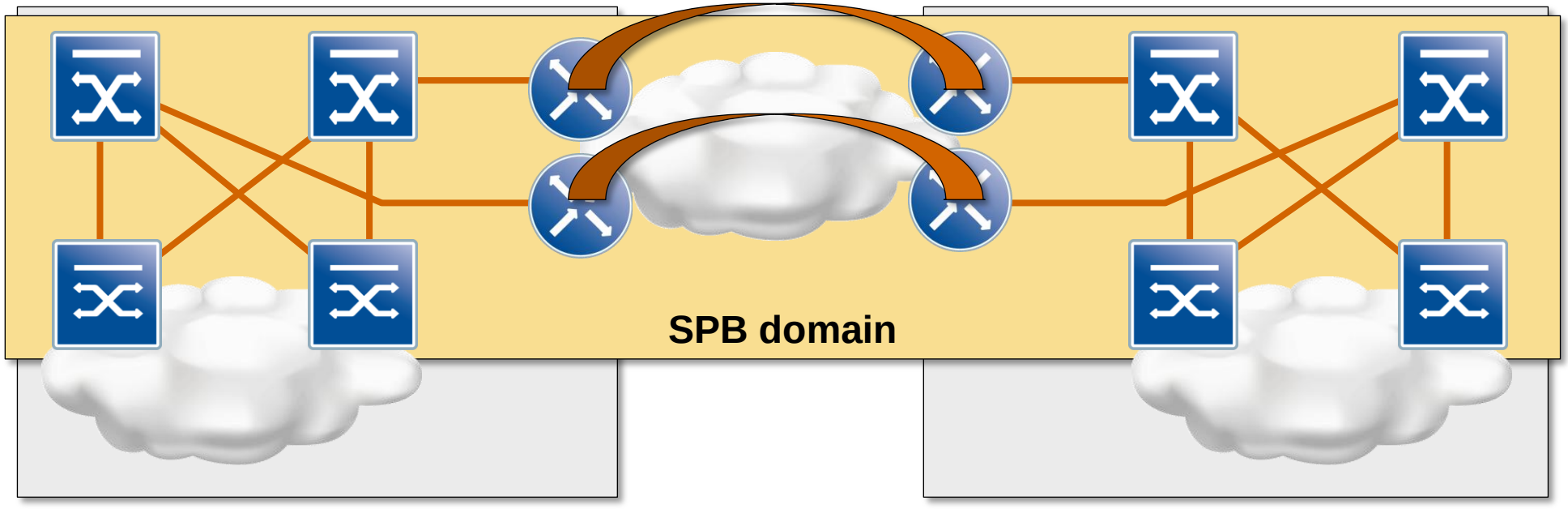


Spanning tree and MLAG/LACP replaced with SPB(V)

- IS-IS routing of layer-2 endpoints
- No spanning tree or MLAG/LACP in the data center fabric
- Simplified configuration
- Significantly reduced probability of forwarding loops
- Multi-site topologies no longer a problem

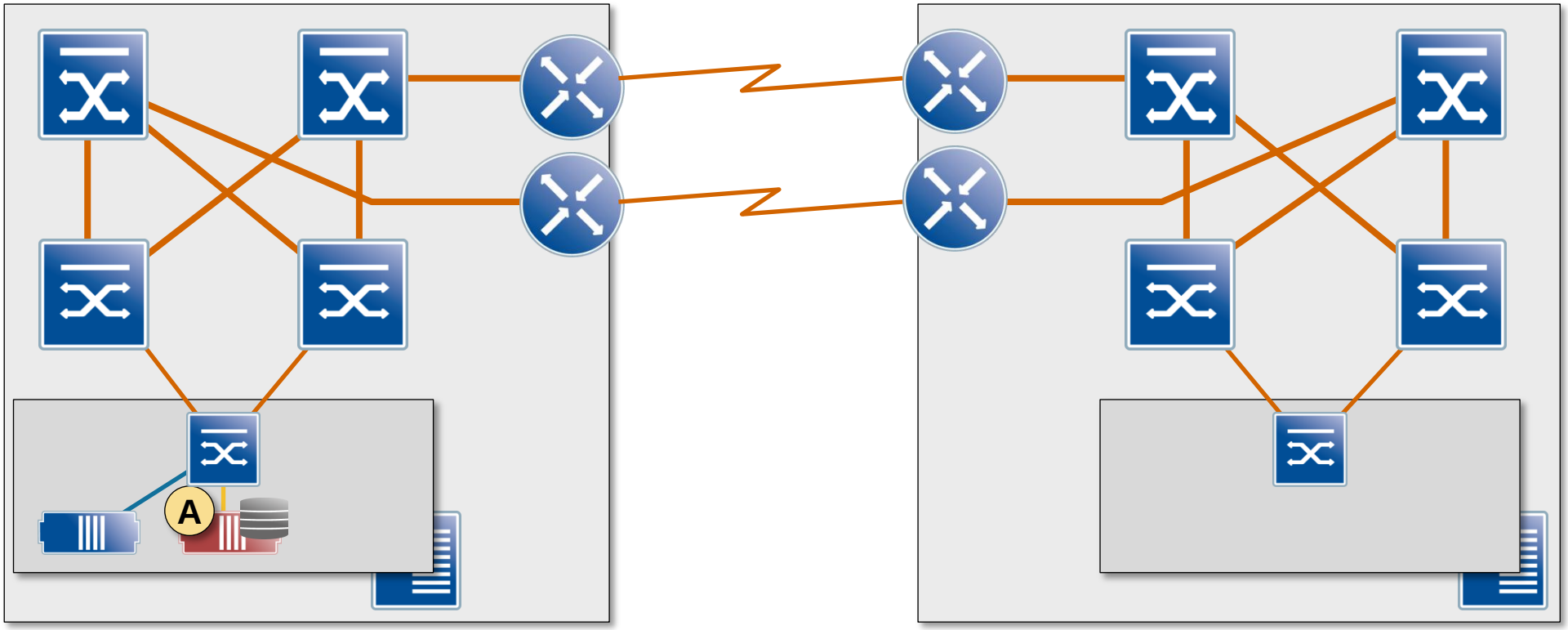
More stable than STP+MLAG/LACP. Still single failure domain.

Enterasys L2 DCI with SPB on Virtual Private Ethernet Services – over GRE

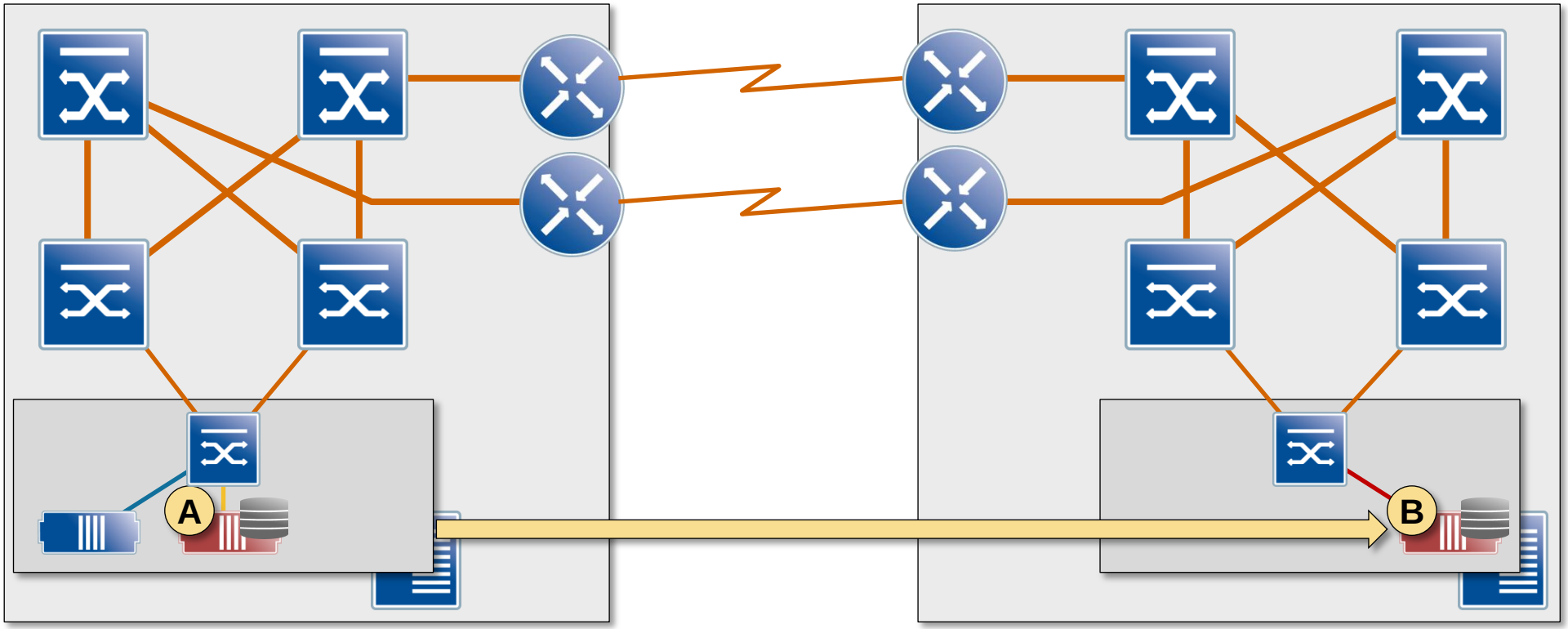


- GRE Layer 2 tunnels act as logical ethernet ports and so P2P links for SPB
- L2 forwarding across L3 transport network
- Seamless integration with SPB(V)
- Multipathing and optimal use of WAN bandwidth without additional technologies like VPLS or E-VPN

Long-Distance Live VM Migration

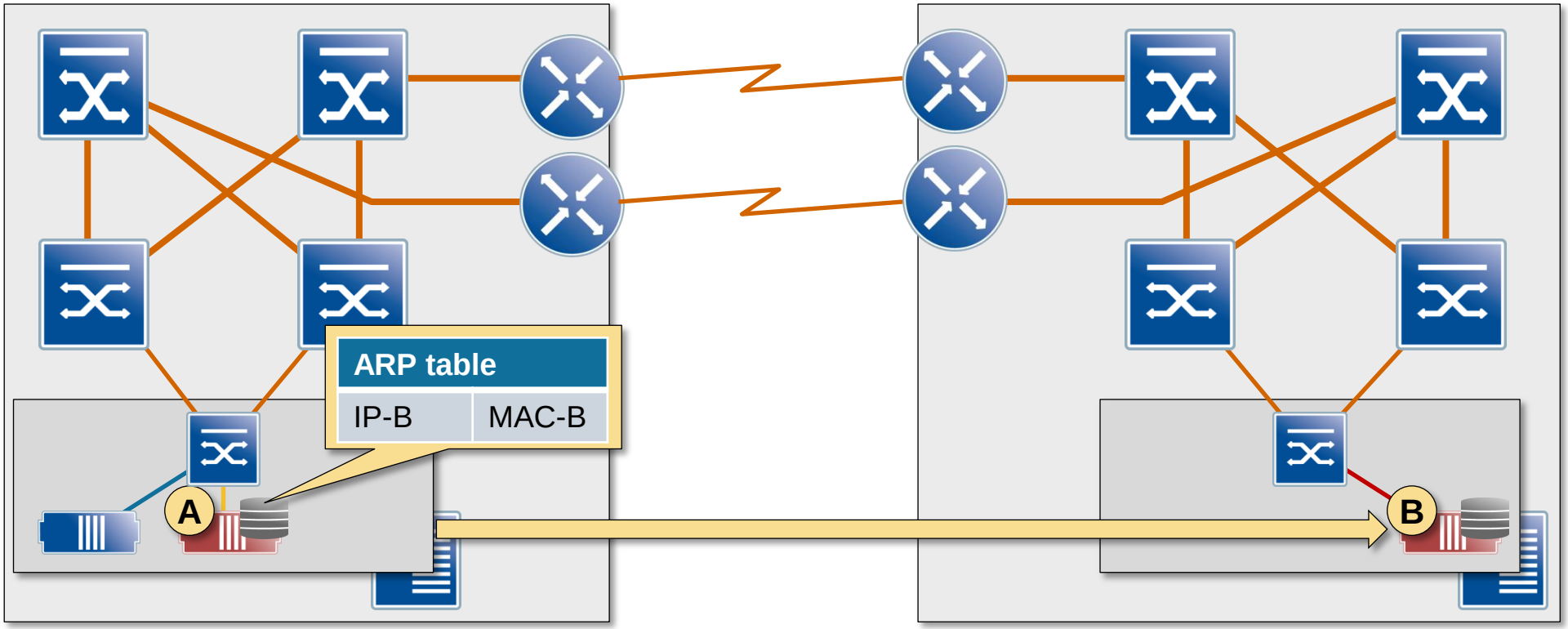


Long-Distance Live VM Migration



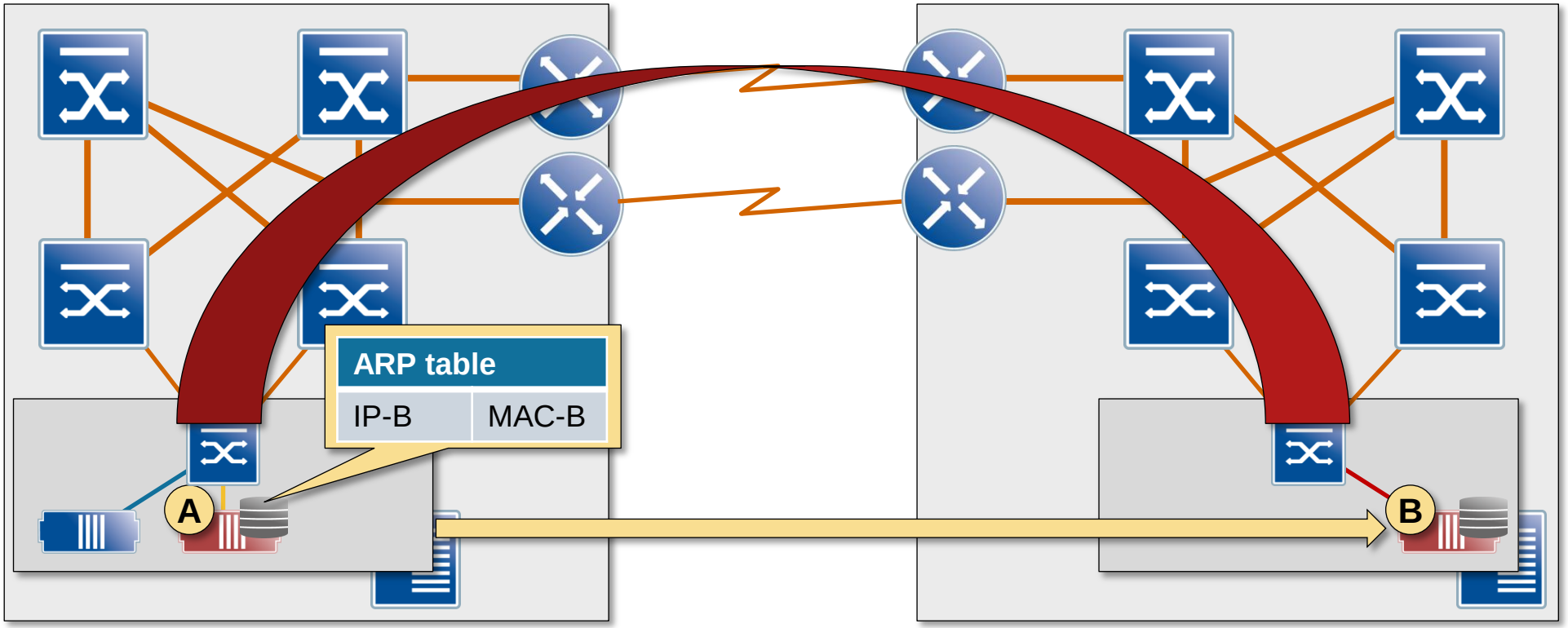
- Live VM moved between data centers has too much state (ARP tables)

Long-Distance Live VM Migration



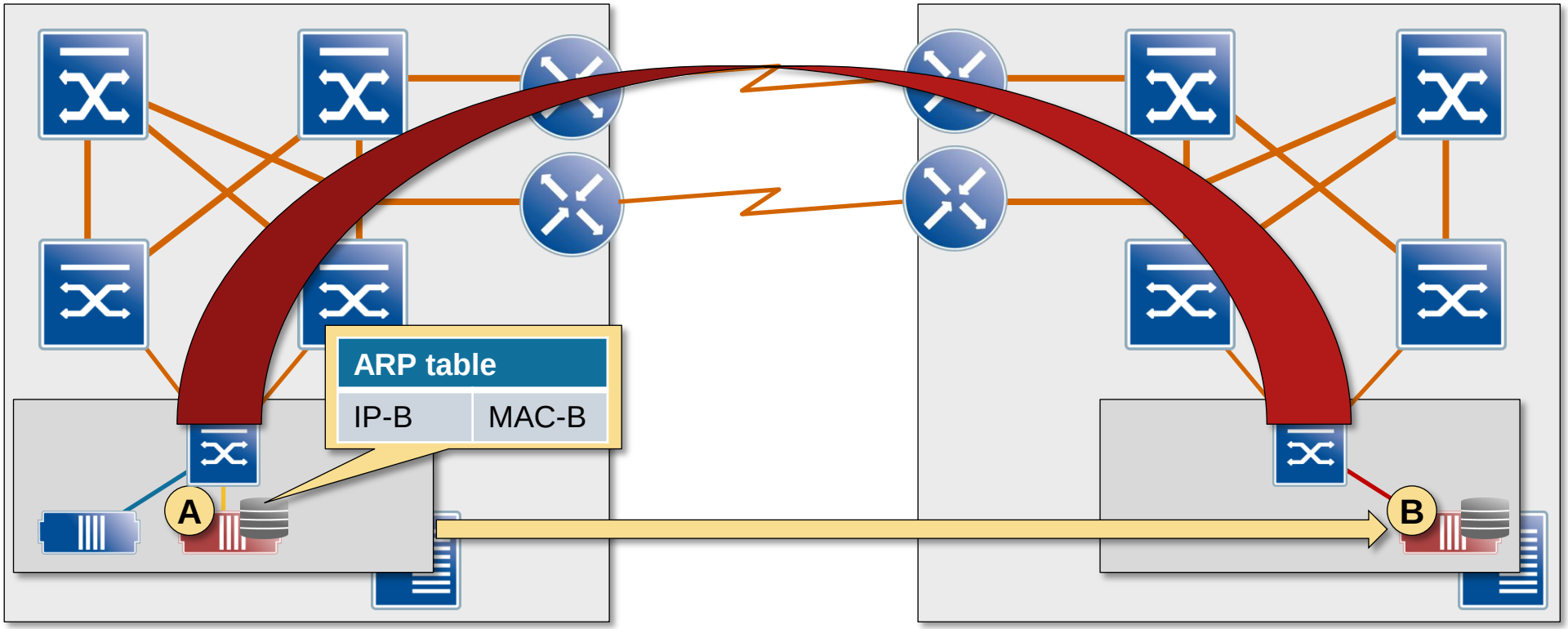
- Live VM moved between data centers has too much state (ARP tables)

Long-Distance Live VM Migration



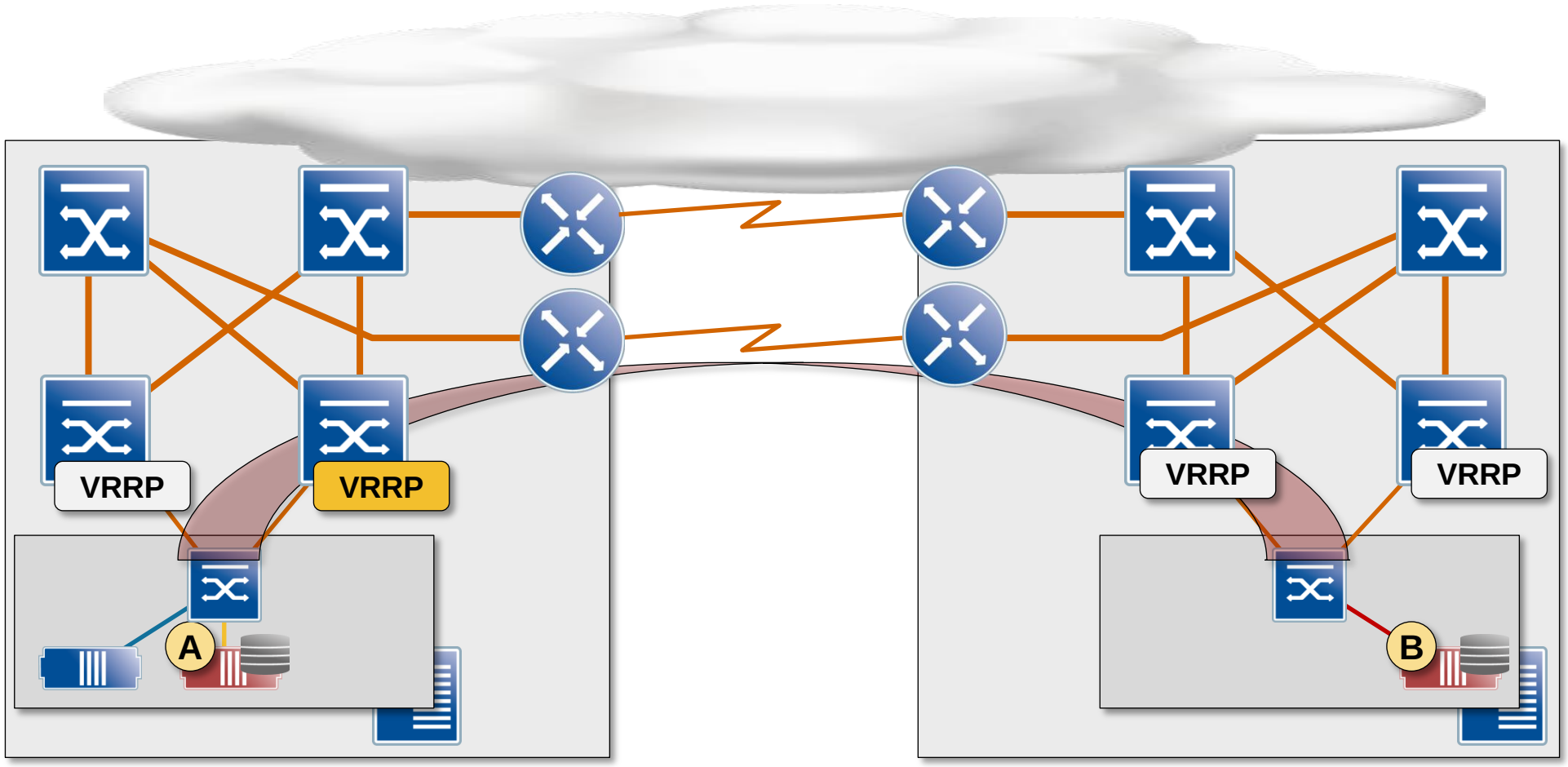
- Live VM moved between data centers has too much state (ARP tables)
- Layer-2 subnet across both data centers is mandatory

Long-Distance Live VM Migration

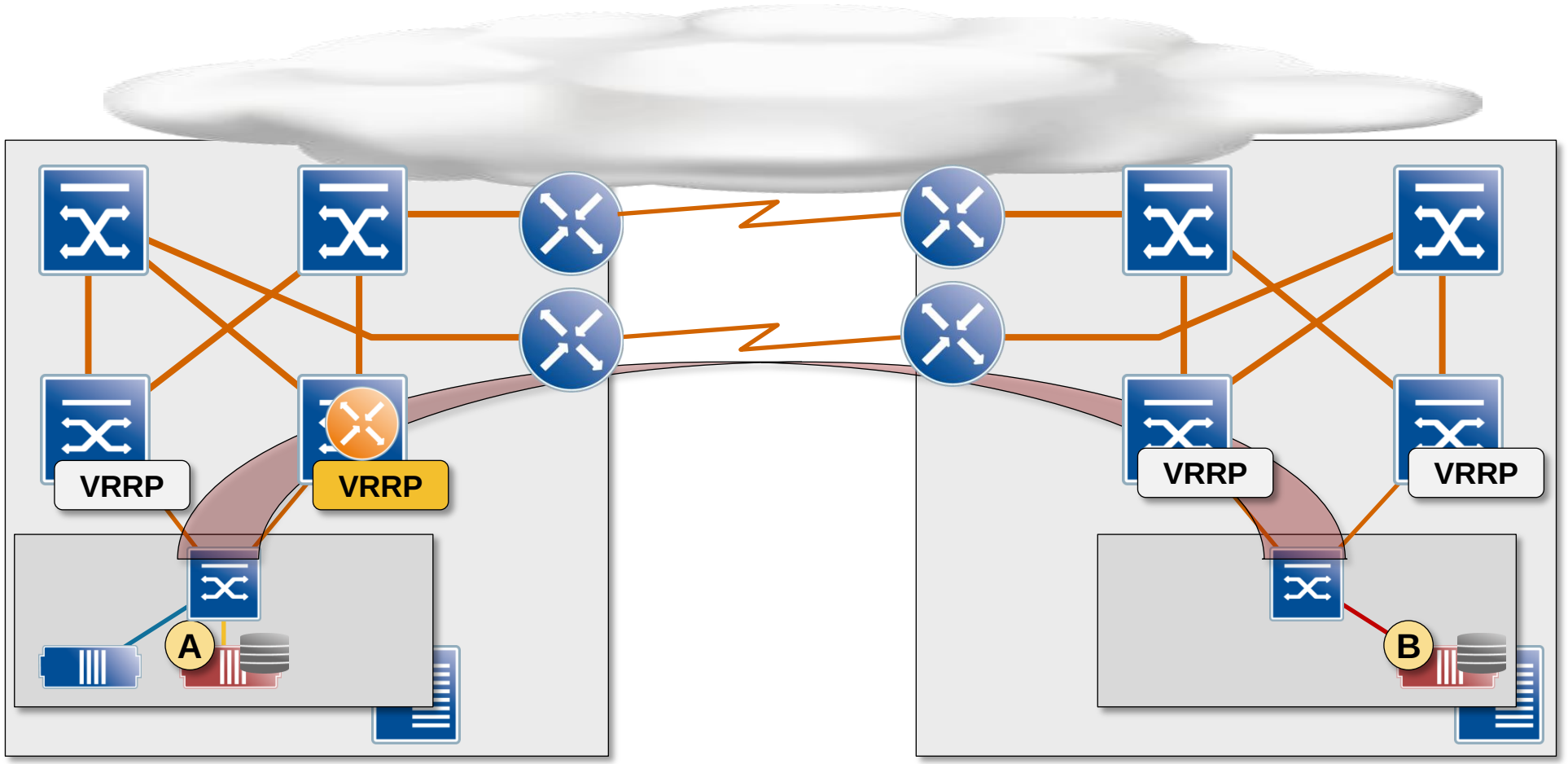


- Live VM moved between data centers has too much state (ARP tables)
- Layer-2 subnet across both data centers is mandatory
- Fabric routing removes inter-subnet traffic trombones
- Host routing (eventually) removes network-to-VM traffic trombones

Inter-Subnet Traffic Without Fabric Routing

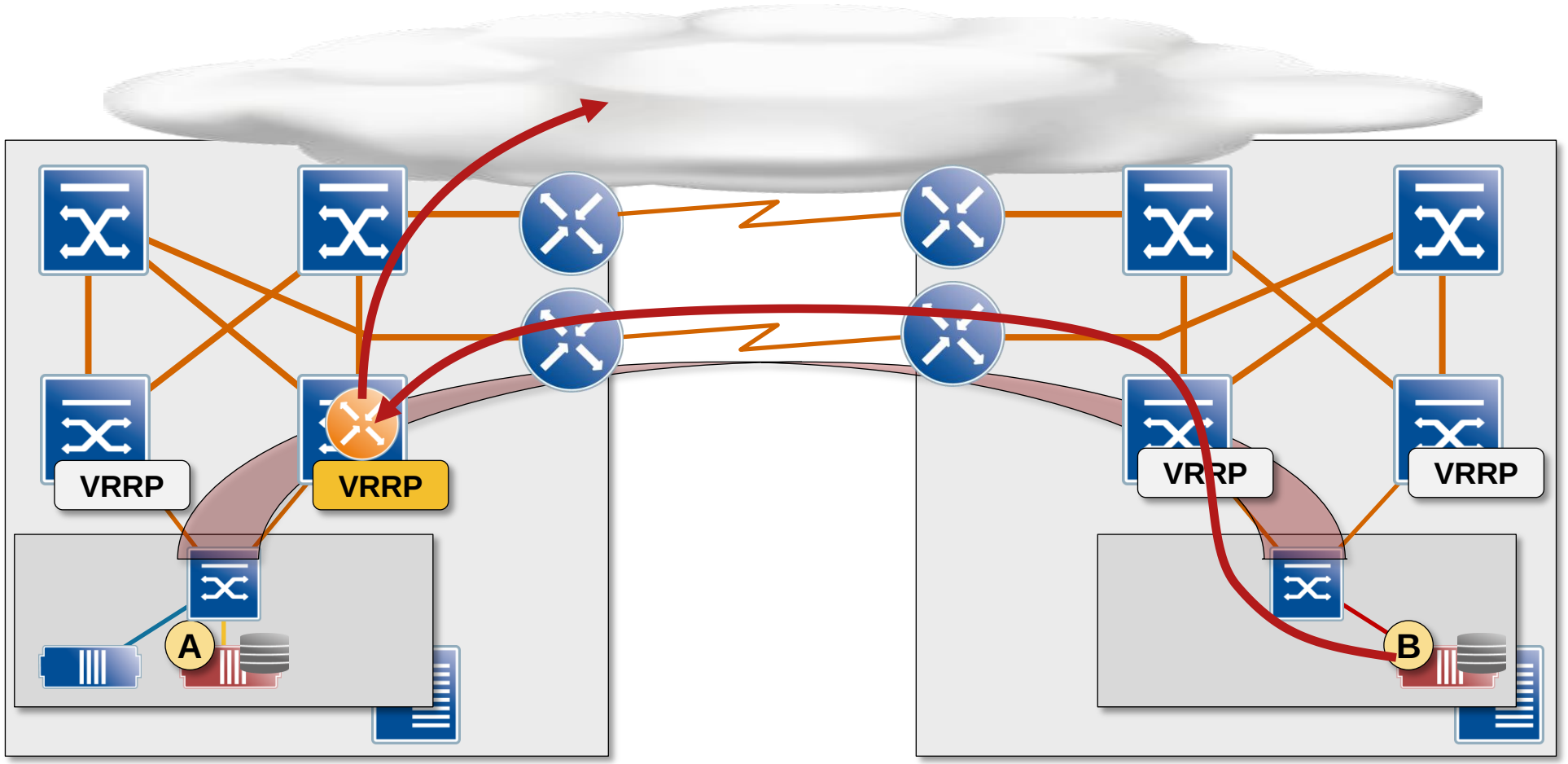


Inter-Subnet Traffic Without Fabric Routing



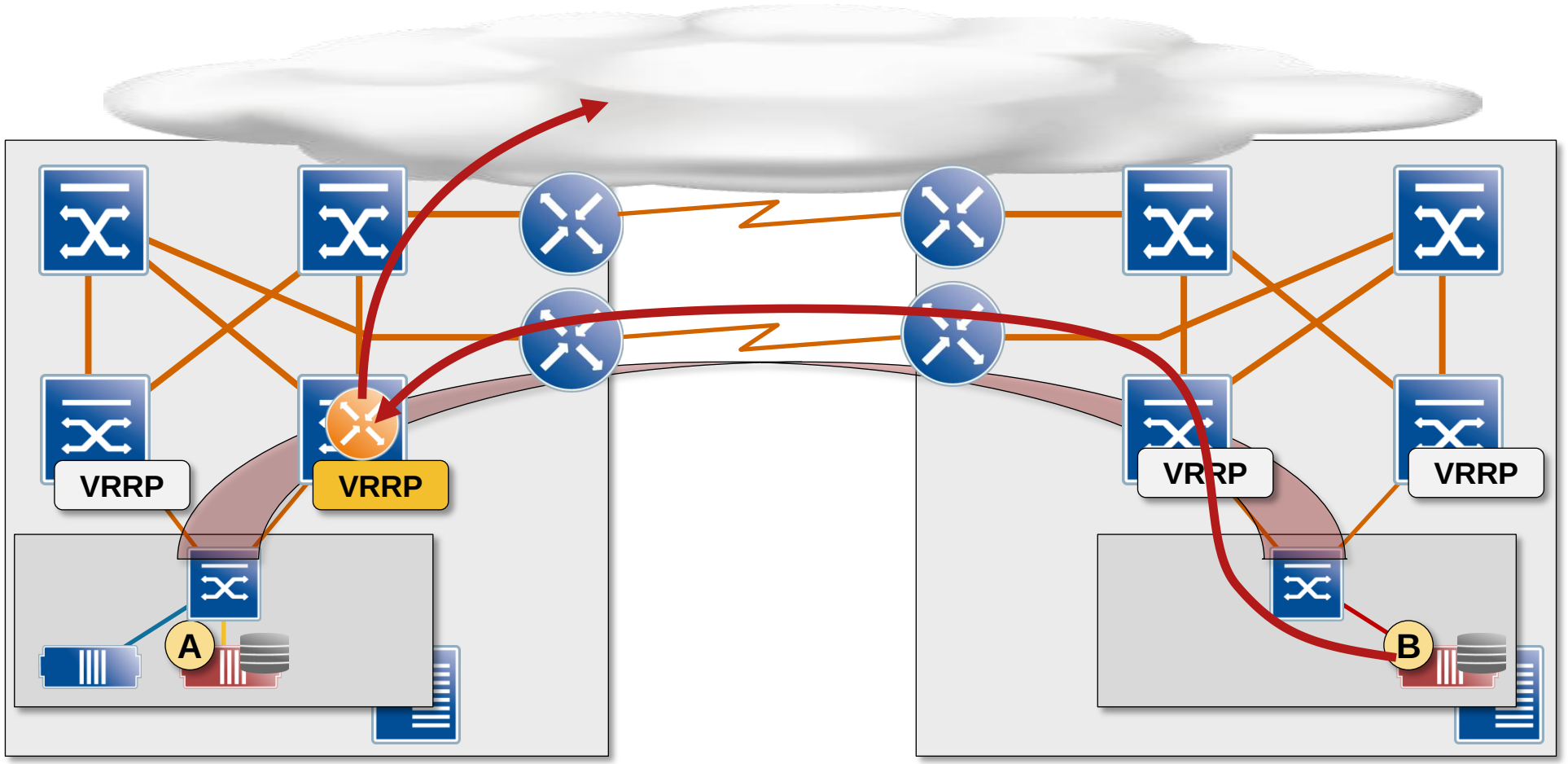
- VM pinned to single exit point (single active VRRP gateway)

Inter-Subnet Traffic Without Fabric Routing



- VM pinned to single exit point (single active VRRP gateway)
- Outbound traffic might be sent across DCI link

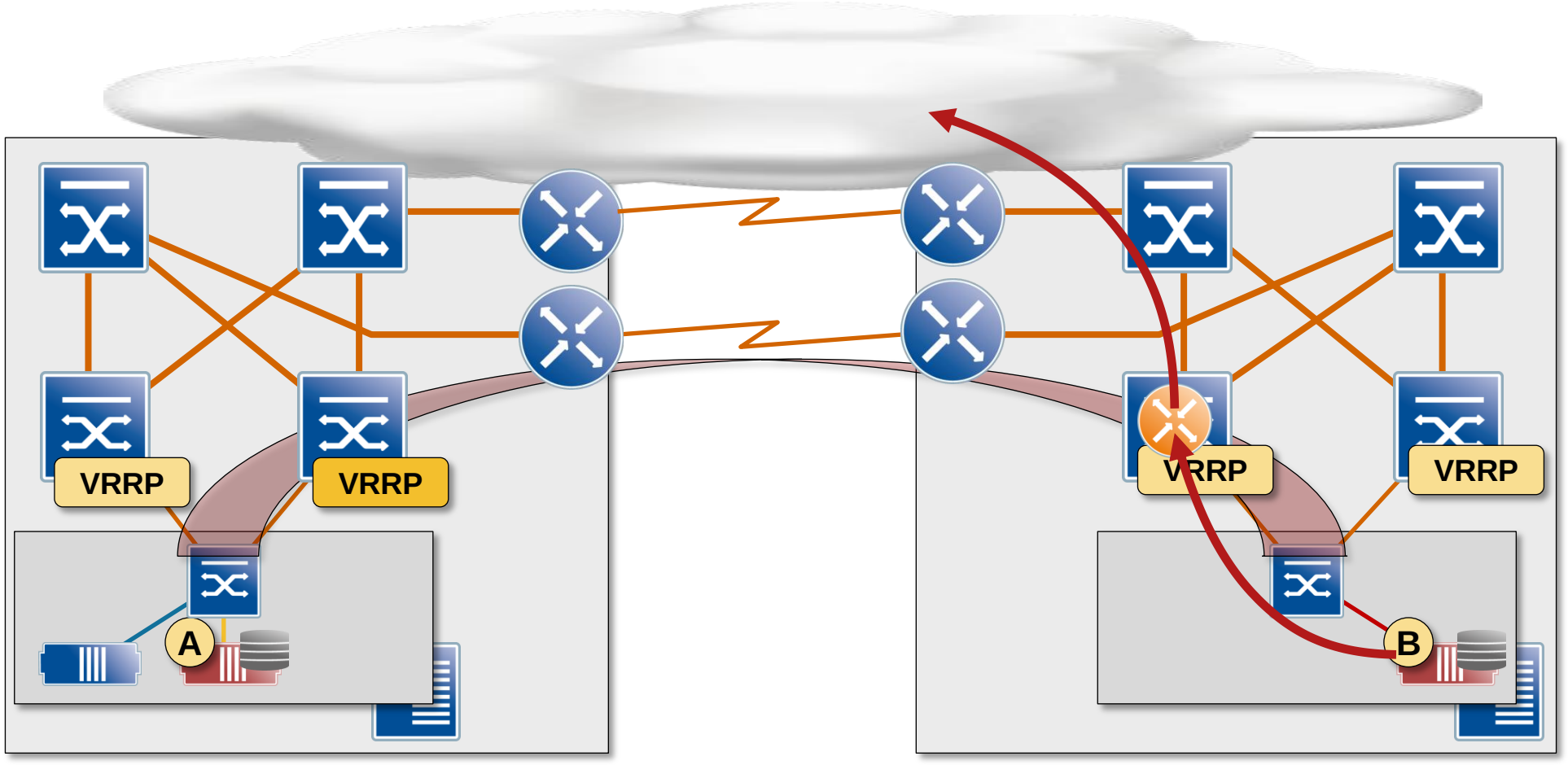
Inter-Subnet Traffic Without Fabric Routing



- VM pinned to single exit point (single active VRRP gateway)
- Outbound traffic might be sent across DCI link

Workaround: First-hop localization (FHRP filters). Don't use!

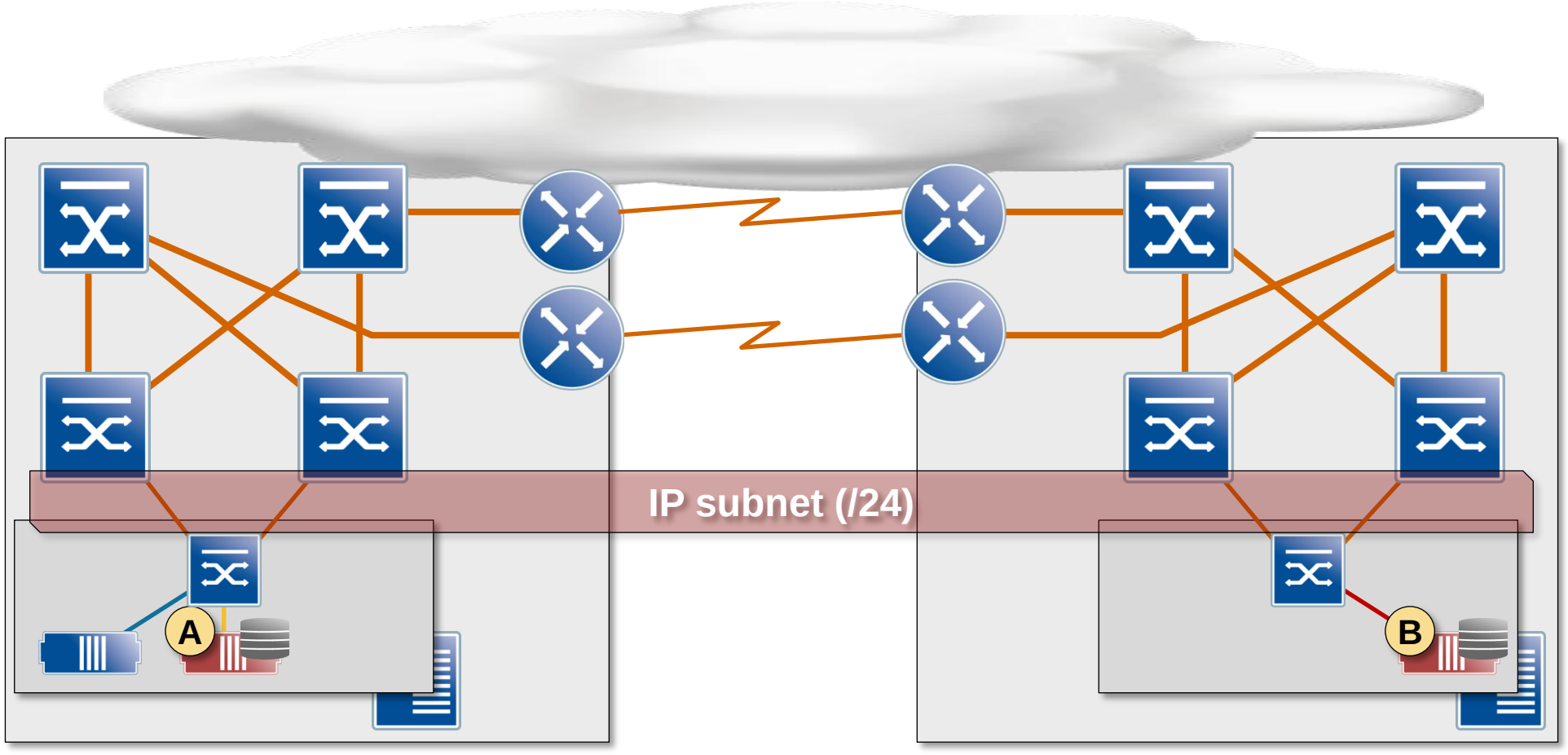
Inter-Subnet Traffic With Fabric Routing



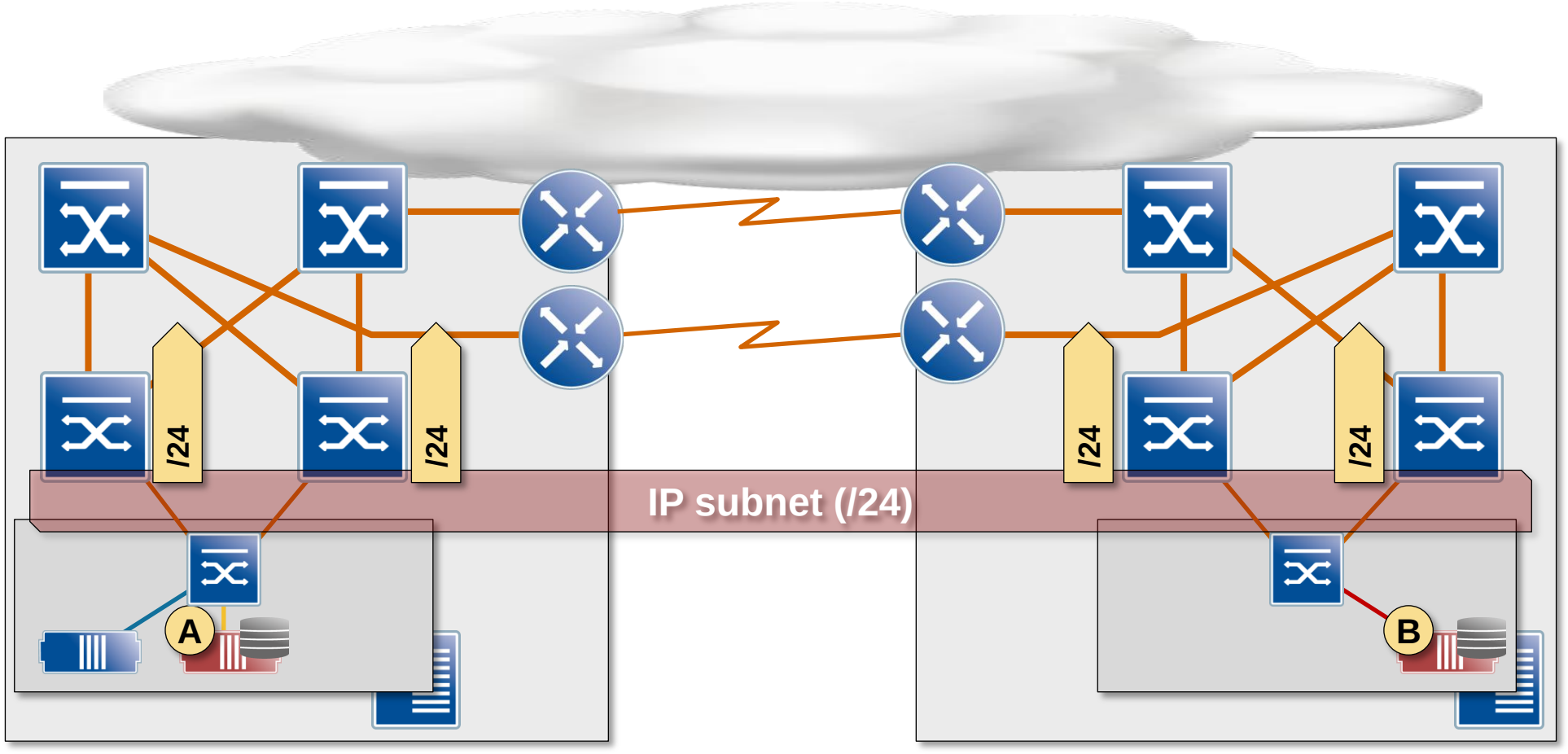
- All ToR switches are active L3 gateways
- Outbound traffic flow is optimal

Warning: stateful appliances in the outbound path break optimal forwarding

Network-to-VM Traffic – Typical Scenario

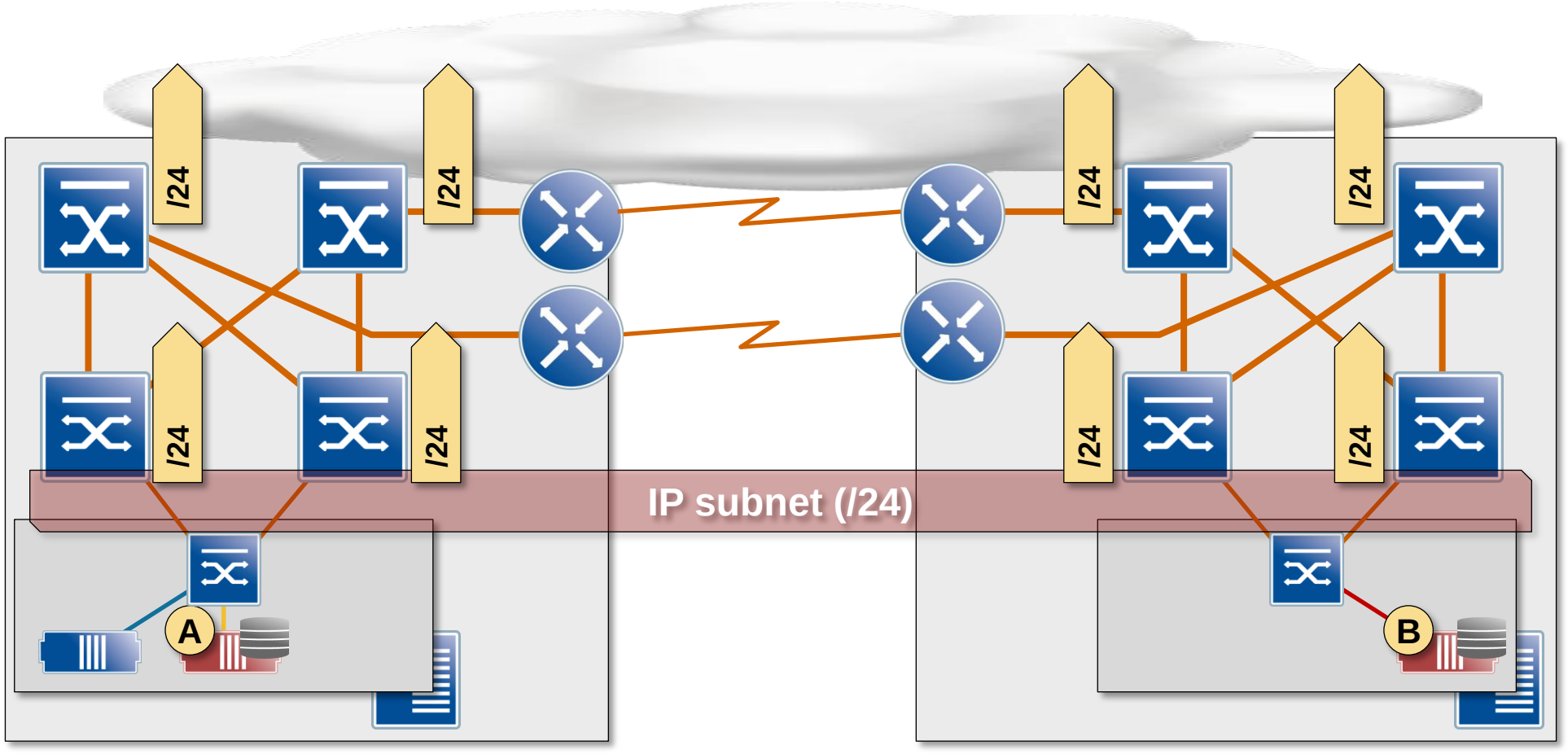


Network-to-VM Traffic – Typical Scenario



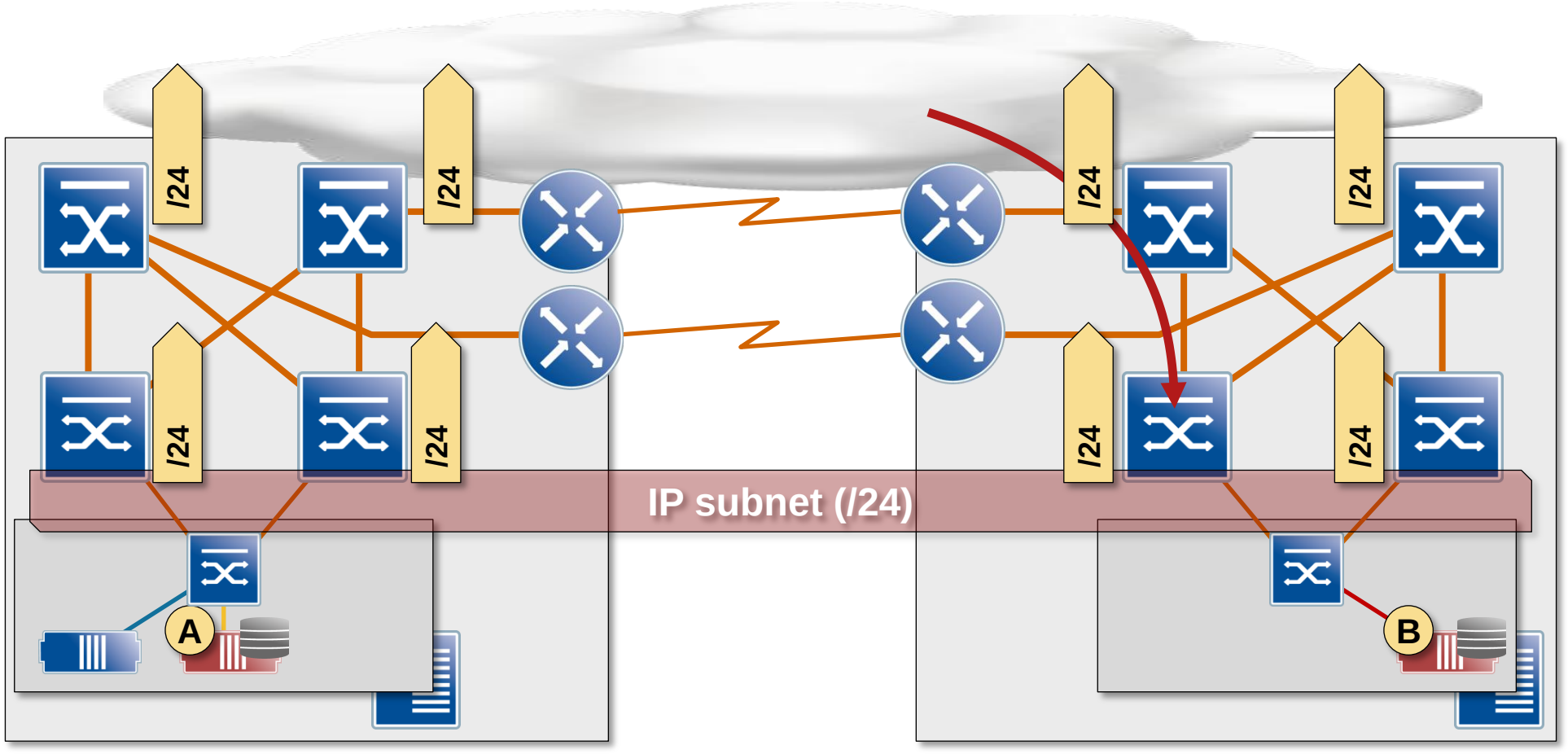
- All ToR switches advertise the subnet prefix

Network-to-VM Traffic – Typical Scenario



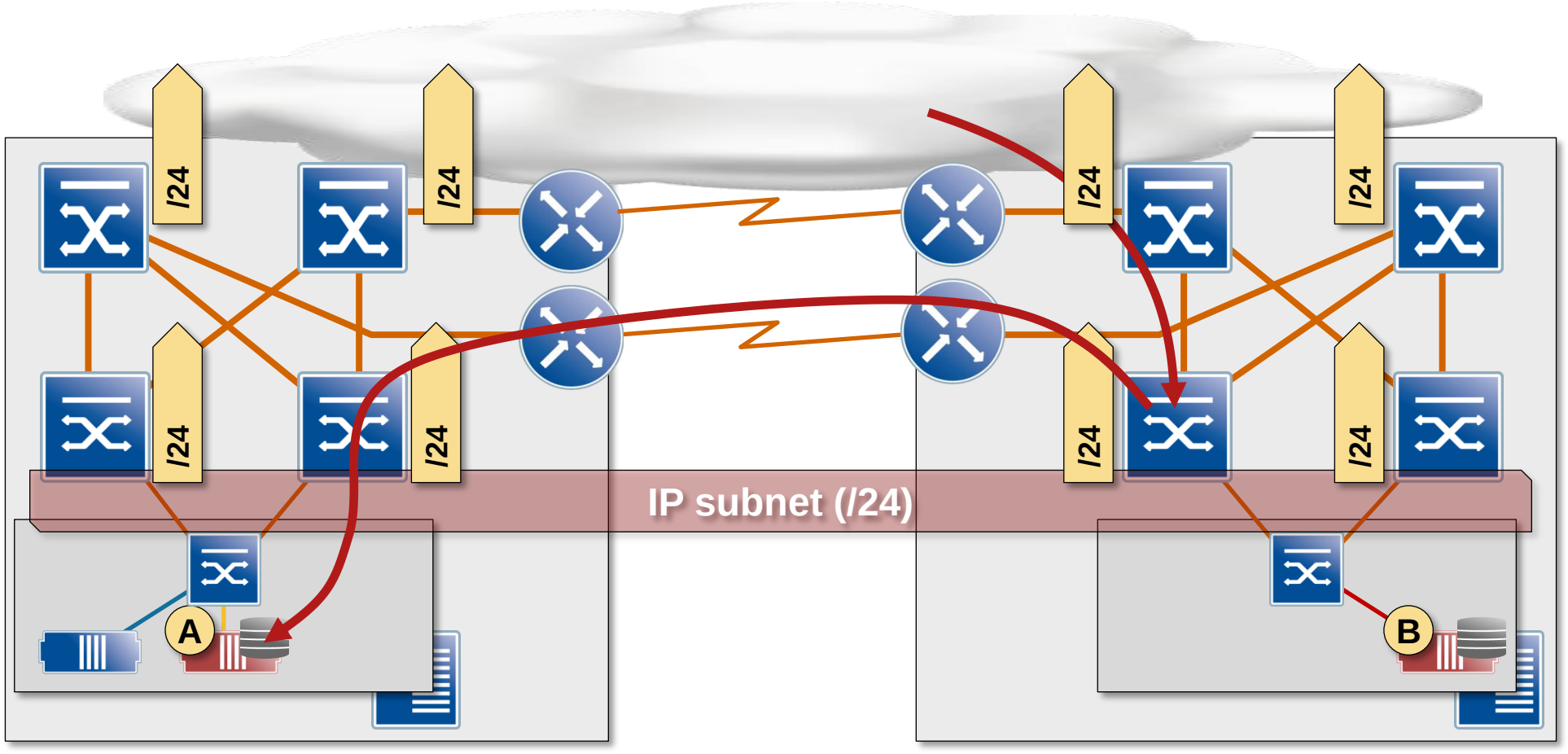
- All ToR switches advertise the subnet prefix
- WAN edge routers advertise the subnet from both data centers

Network-to-VM Traffic – Typical Scenario



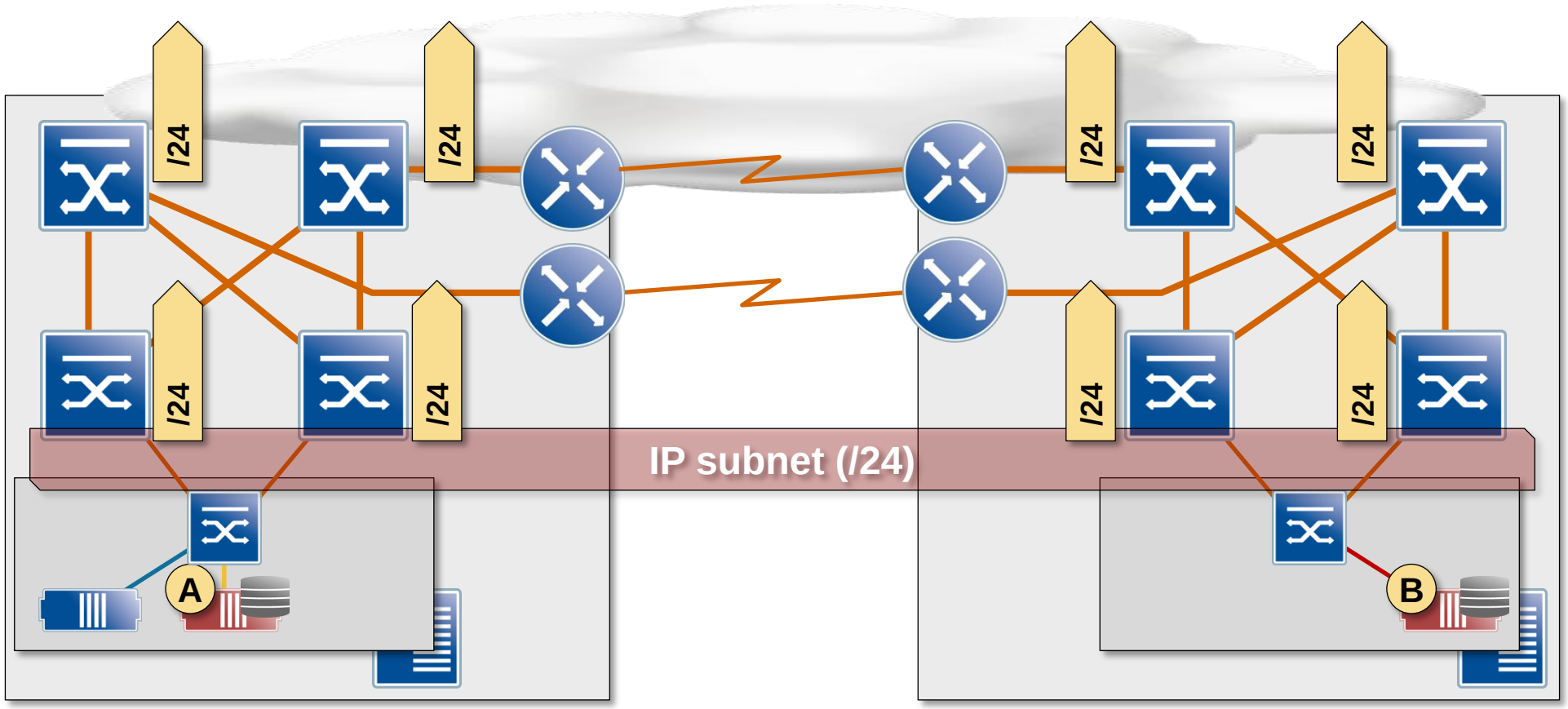
- All ToR switches advertise the subnet prefix
- WAN edge routers advertise the subnet from both data centers
- Half of the inbound traffic arrives to the wrong data center

Network-to-VM Traffic – Typical Scenario

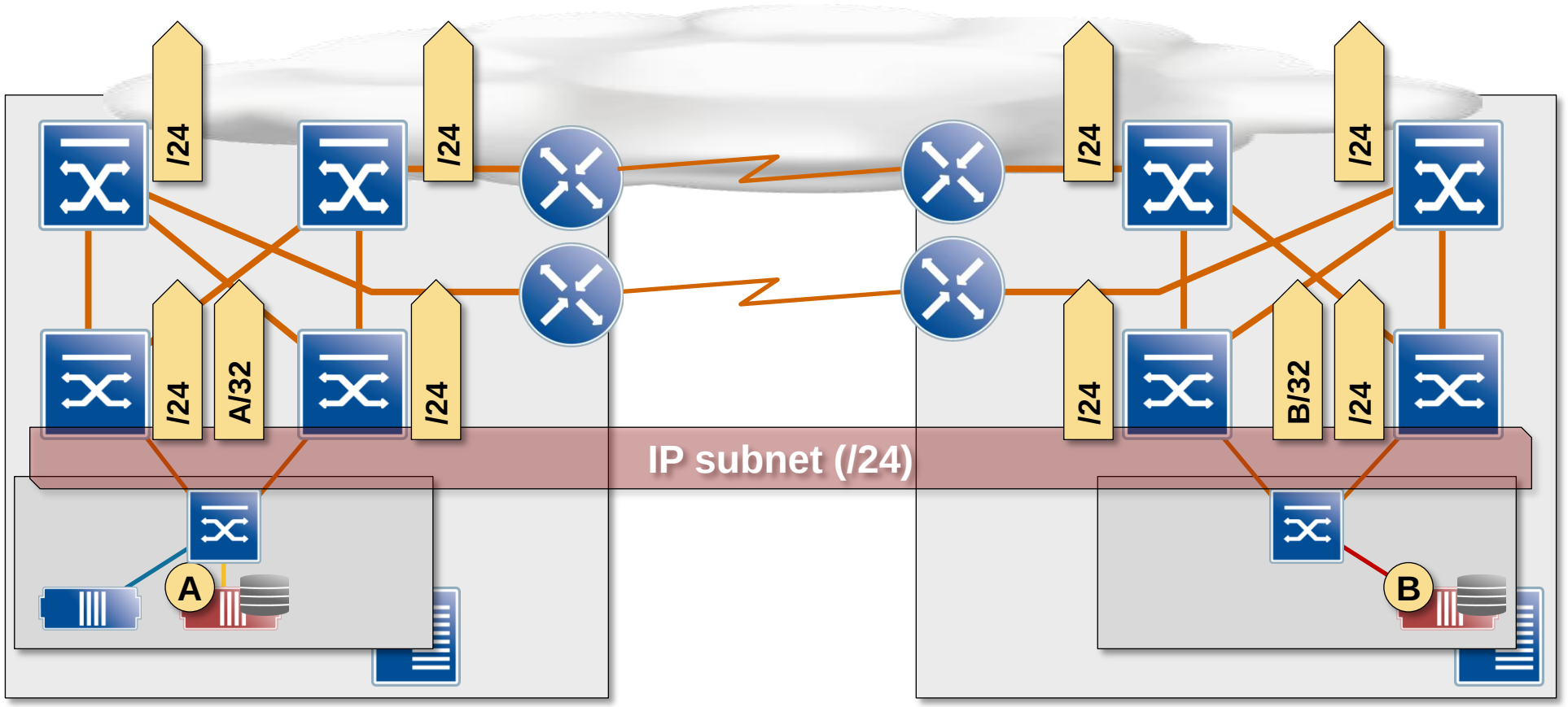


- All ToR switches advertise the subnet prefix
- WAN edge routers advertise the subnet from both data centers
- Half of the inbound traffic arrives to the wrong data center

Southbound Traffic with Host Routing

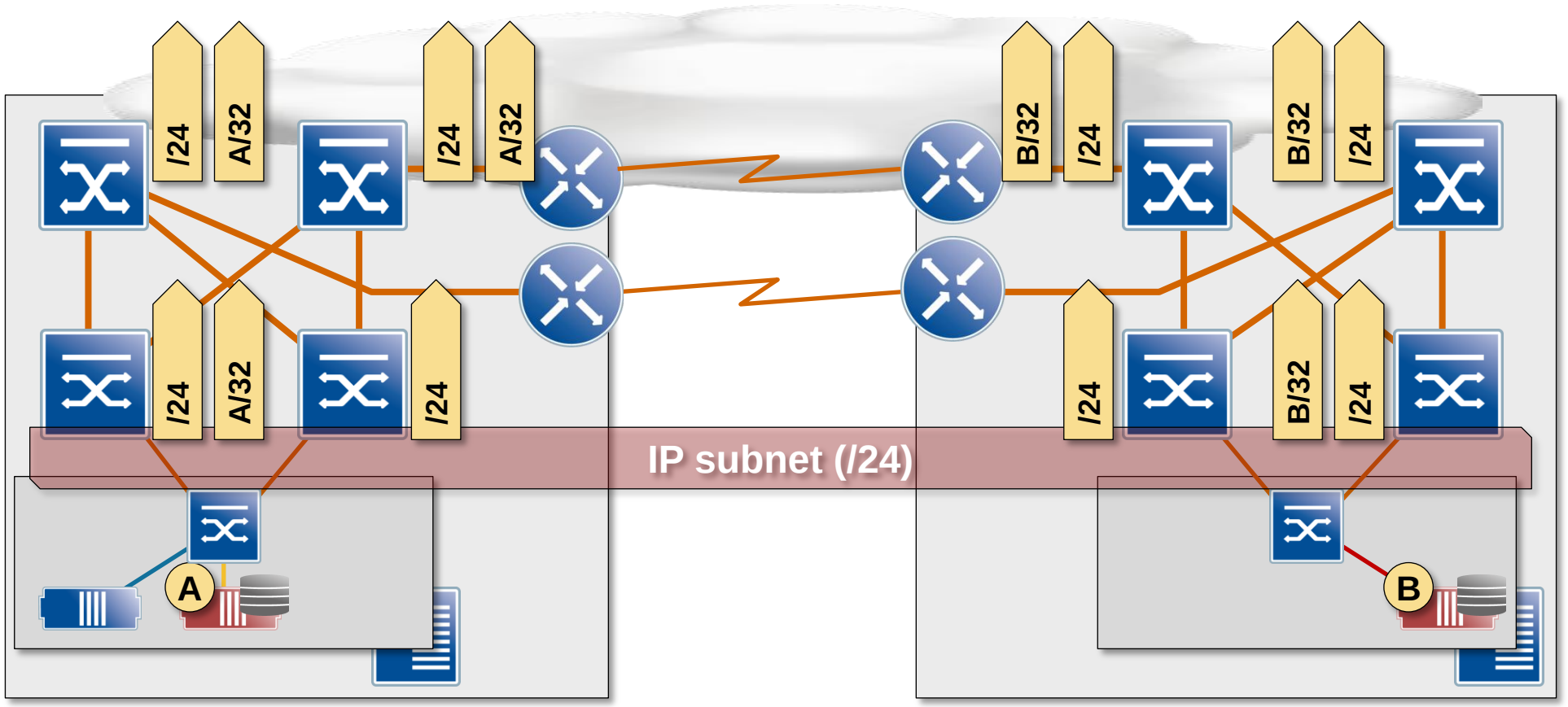


Southbound Traffic with Host Routing



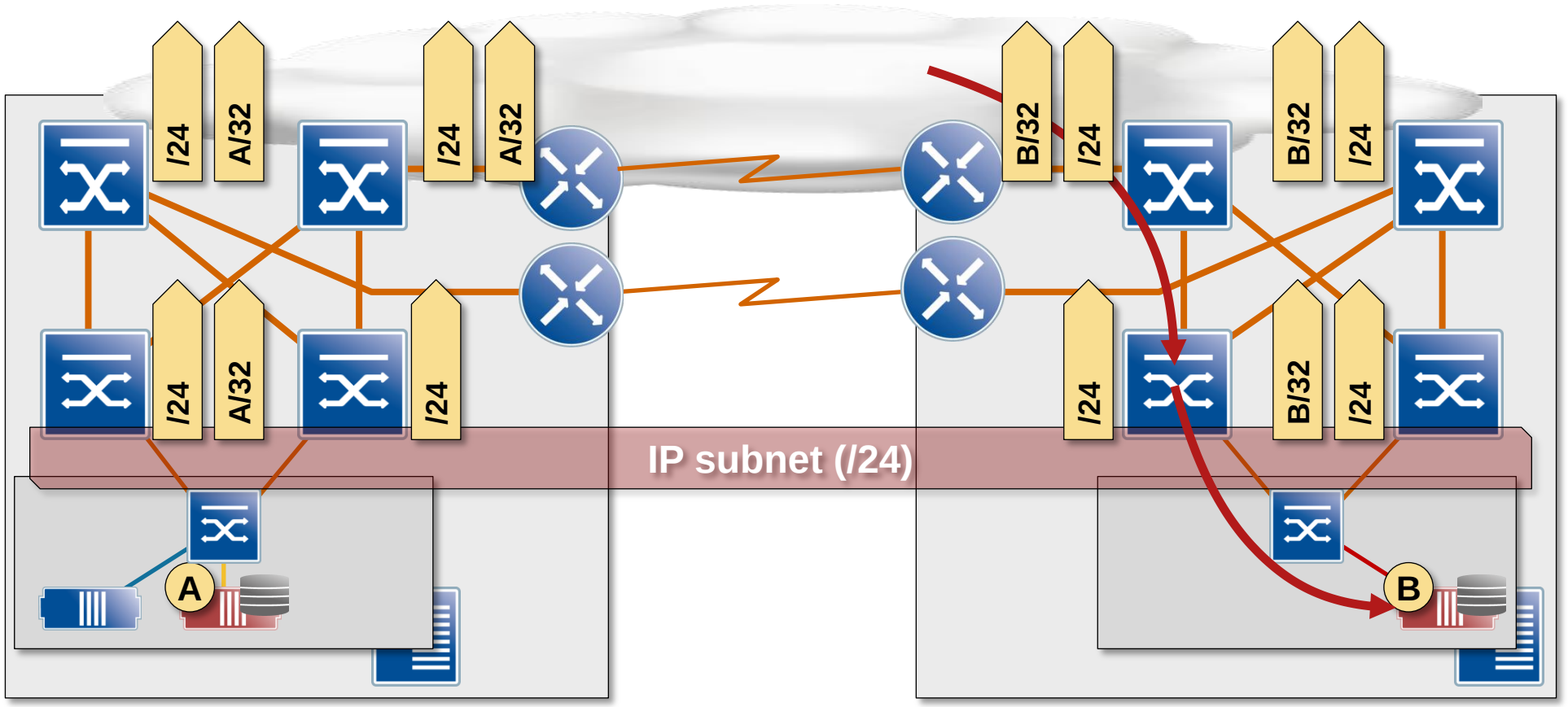
- ToR switches advertise host routes to directly connect VMs

Southbound Traffic with Host Routing



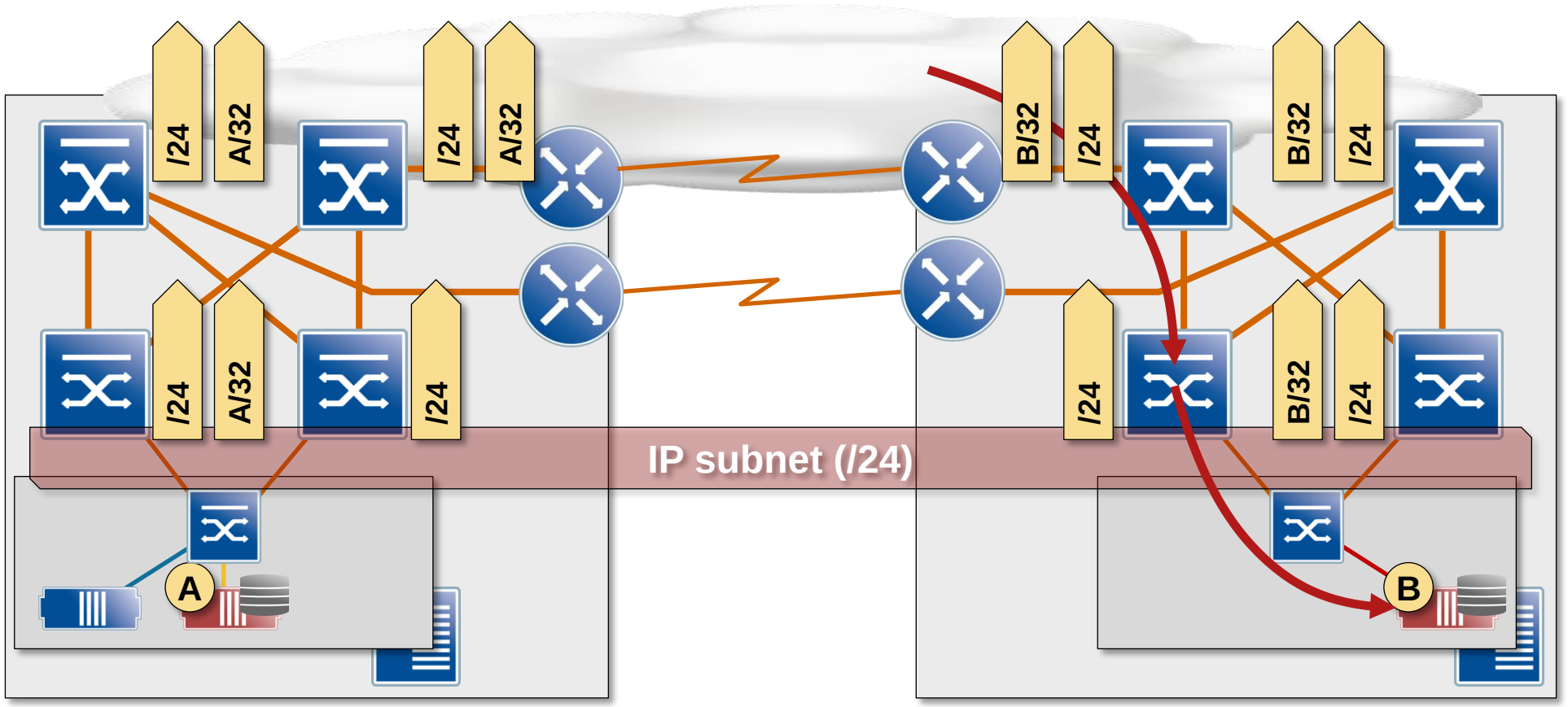
- ToR switches advertise host routes to directly connect VMs
- WAN edge routers advertise individual host routes

Southbound Traffic with Host Routing



- ToR switches advertise host routes to directly connect VMs
- WAN edge routers advertise individual host routes
- Inbound traffic flow is optimal

Southbound Traffic with Host Routing



- ToR switches advertise host routes to directly connect VMs
- WAN edge routers advertise individual host routes
- Inbound traffic flow is optimal

Warning: needs host routes in WAN, won't work with global Internet

How Far Did We Get?

L2 DCI Virtual Private Port Services (today)

L2 DCI SPB(V) via Virtual Private Ethernet Services (autumn)

Fabric routing: optimal server-to-network routing

Host routing: optimal network-to-server routing

How Far Did We Get?

L2 DCI Virtual Private Port Services (today)

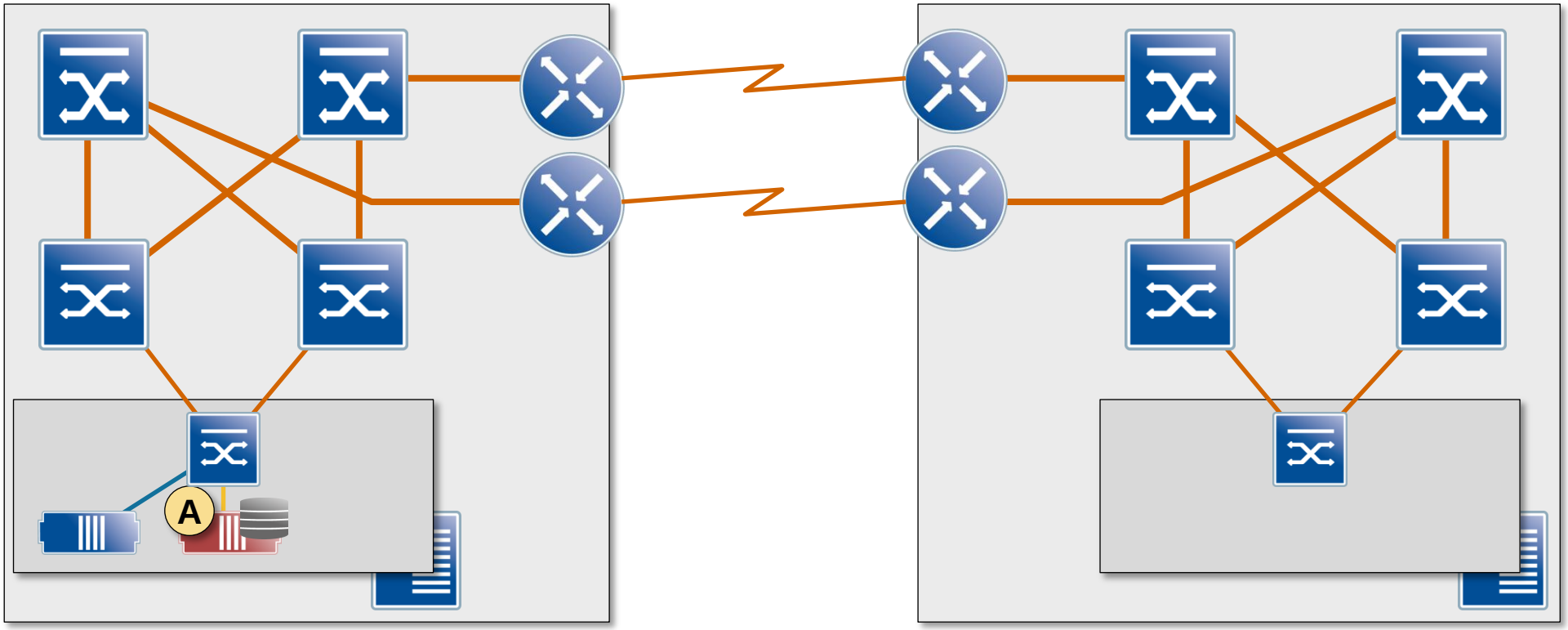
L2 DCI SPB(V) via Virtual Private Ethernet Services (autumn)

Fabric routing: optimal server-to-network routing

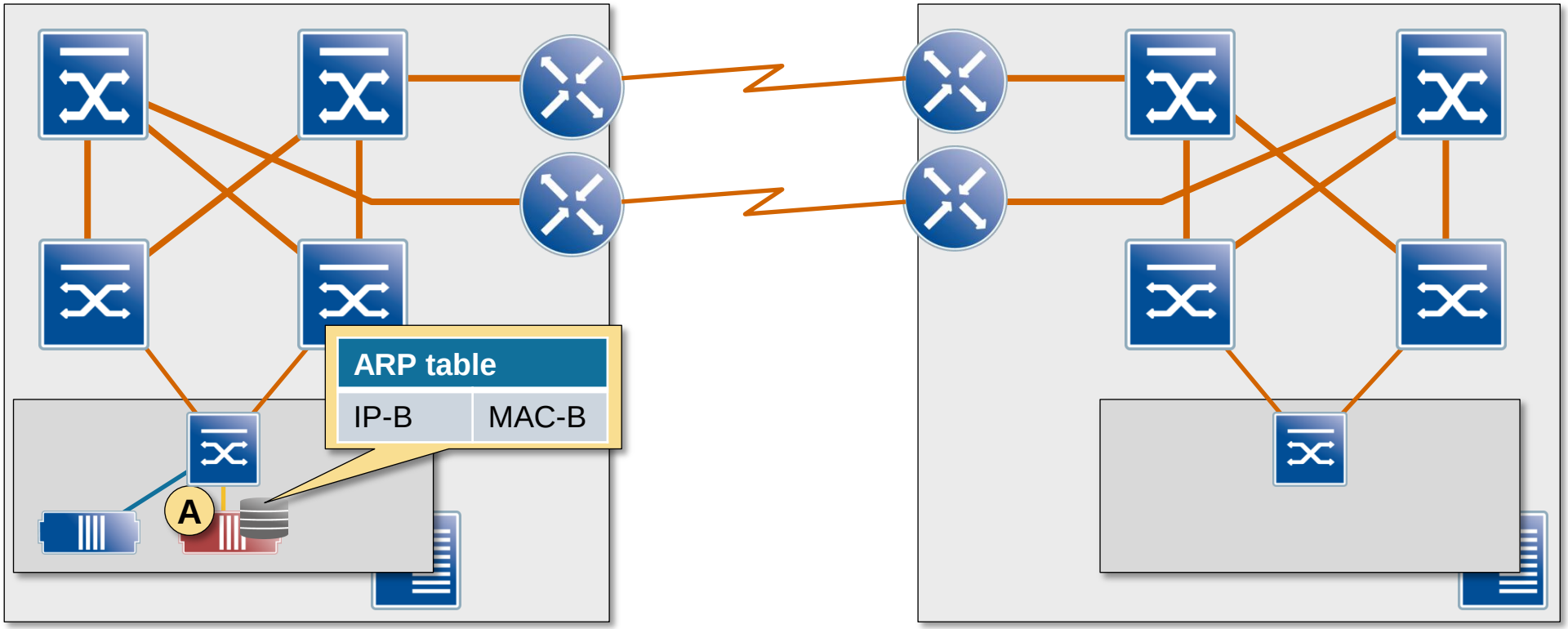
Host routing: optimal network-to-server routing

Can we move VMs over L3 DCI?

Long-Distance Cold VM Migration

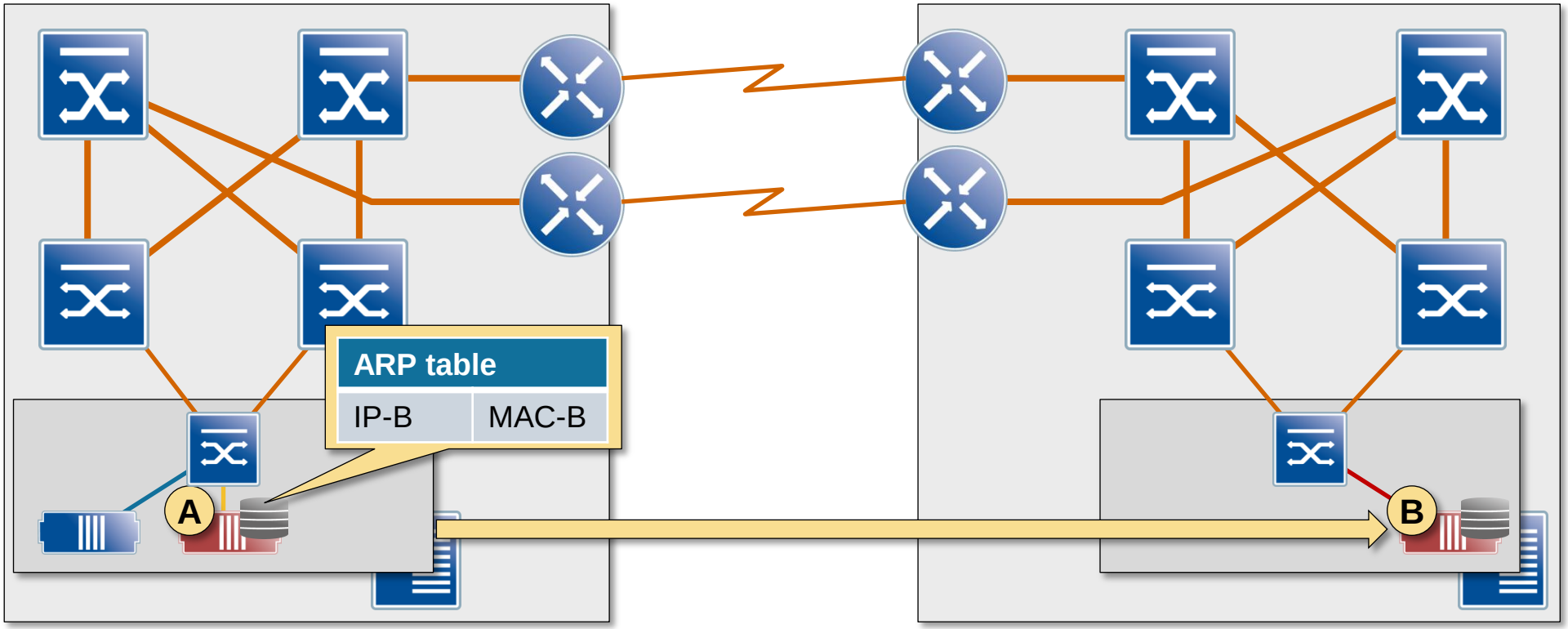


Long-Distance Cold VM Migration



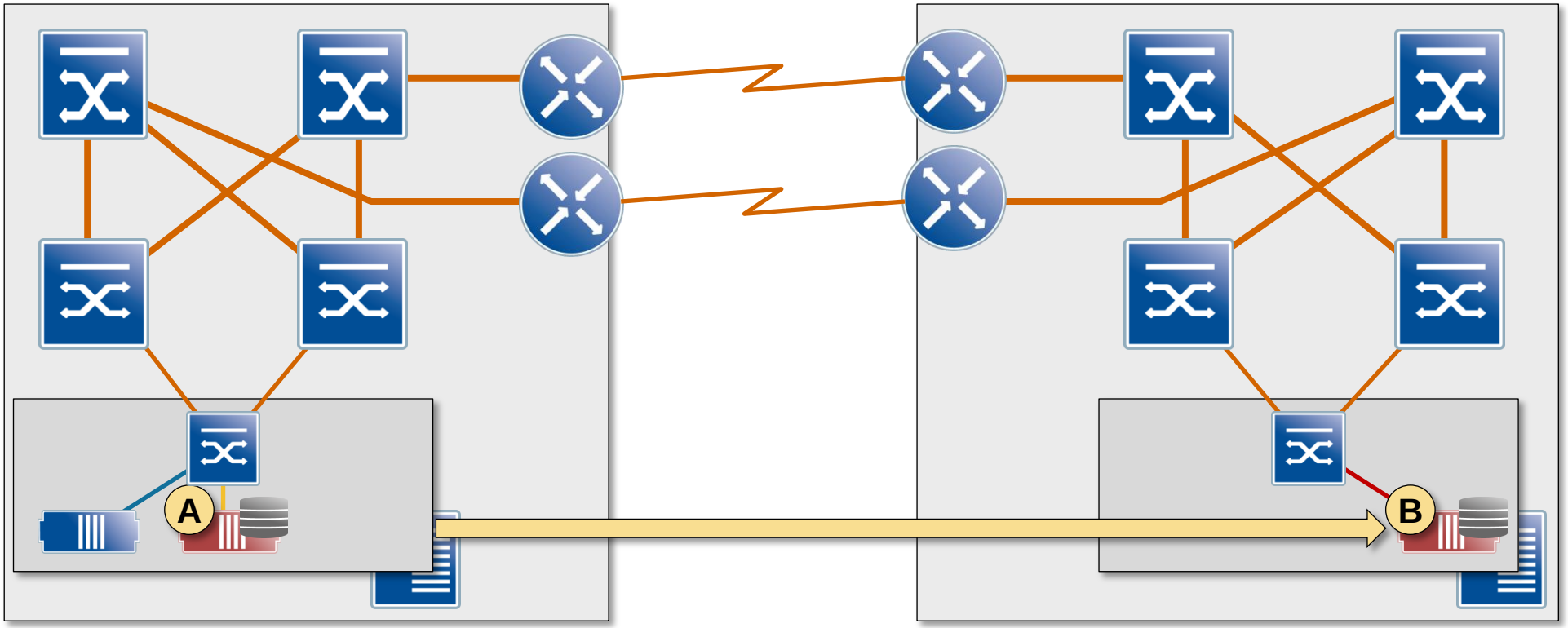
- VM is powered down (optional)

Long-Distance Cold VM Migration



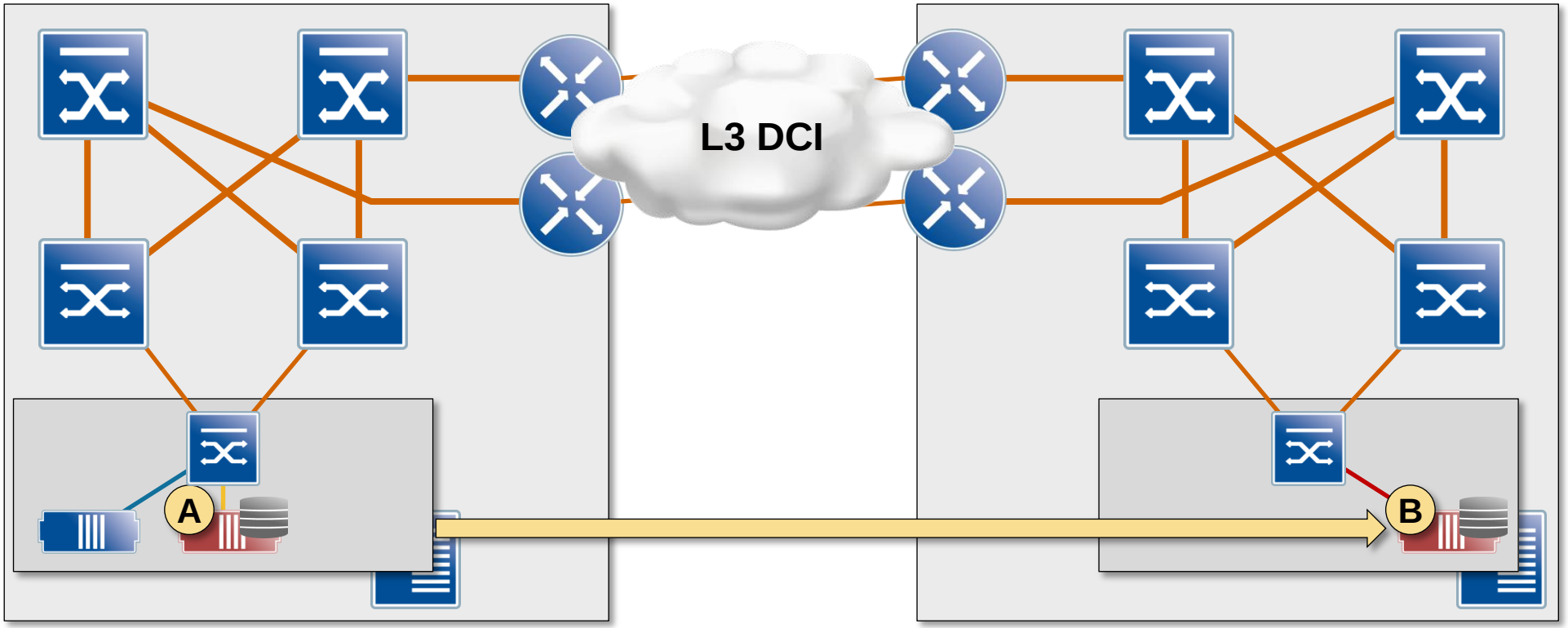
- VM is powered down (optional)
- VM is moved to a cluster in another data center and restarted

Long-Distance Cold VM Migration



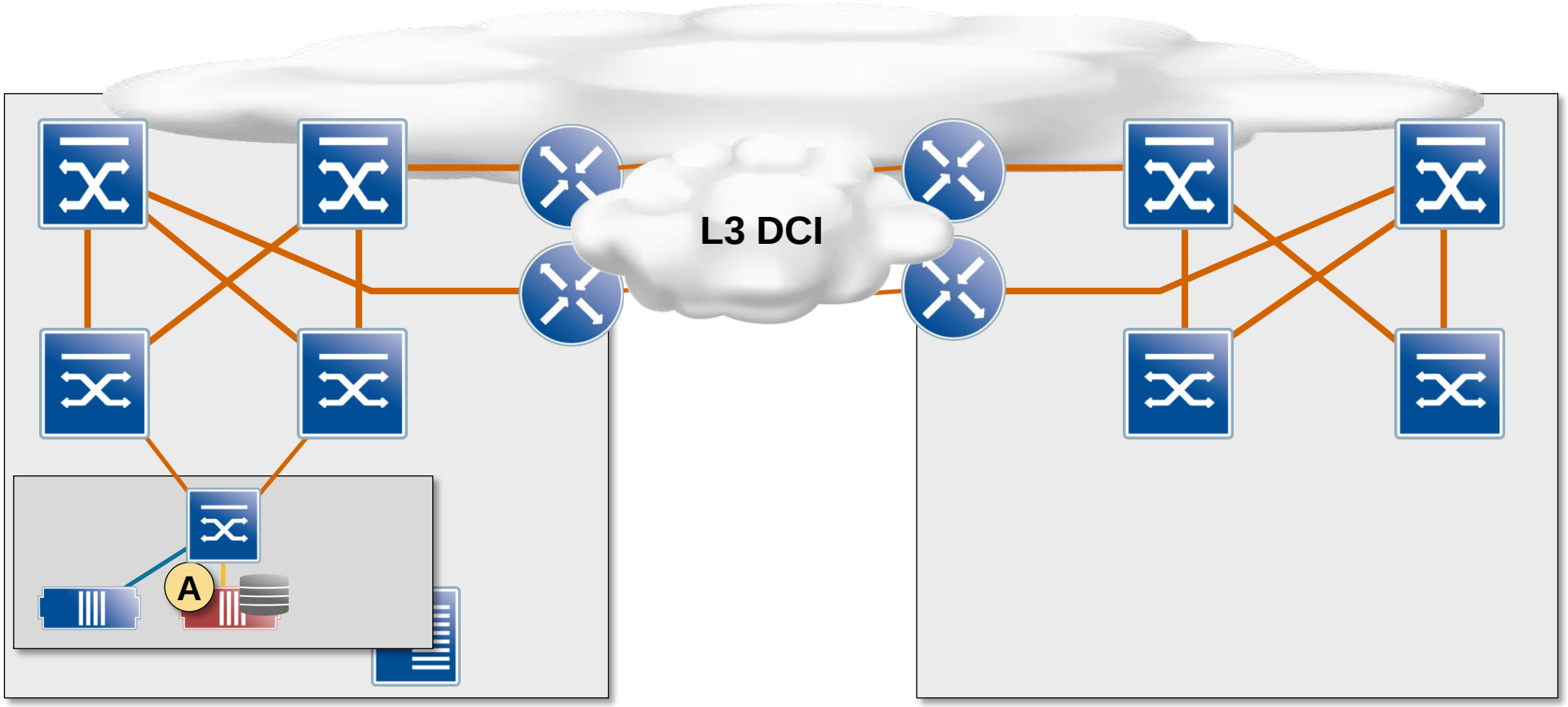
- VM is powered down (optional)
- VM is moved to a cluster in another data center and restarted
- Minimal residual state after ARP cache timeout

Long-Distance Cold VM Migration

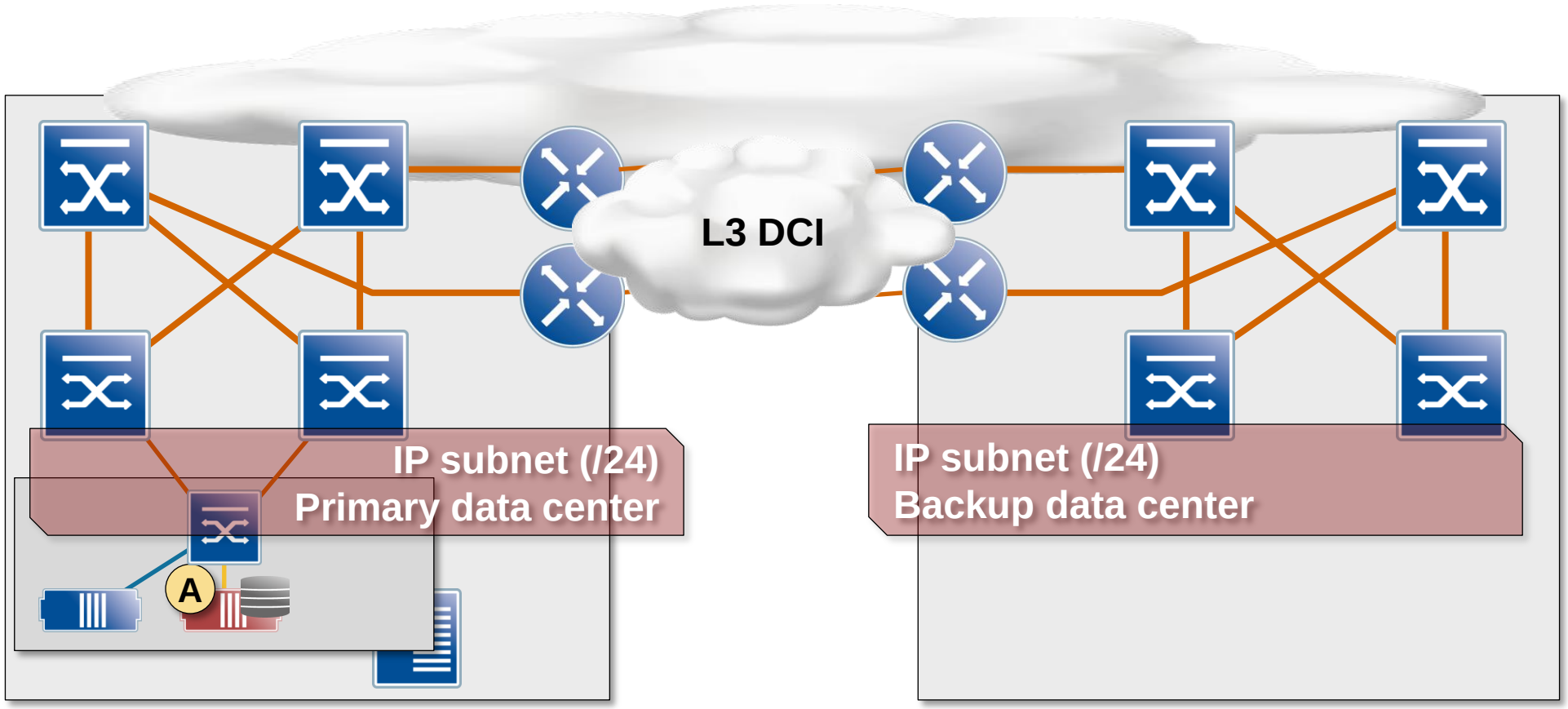


- VM is powered down (optional)
- VM is moved to a cluster in another data center and restarted
- Minimal residual state after ARP cache timeout
- Would fabric/host routing work over L3 DCI?

VM Mobility with L3 DCI – Network Design

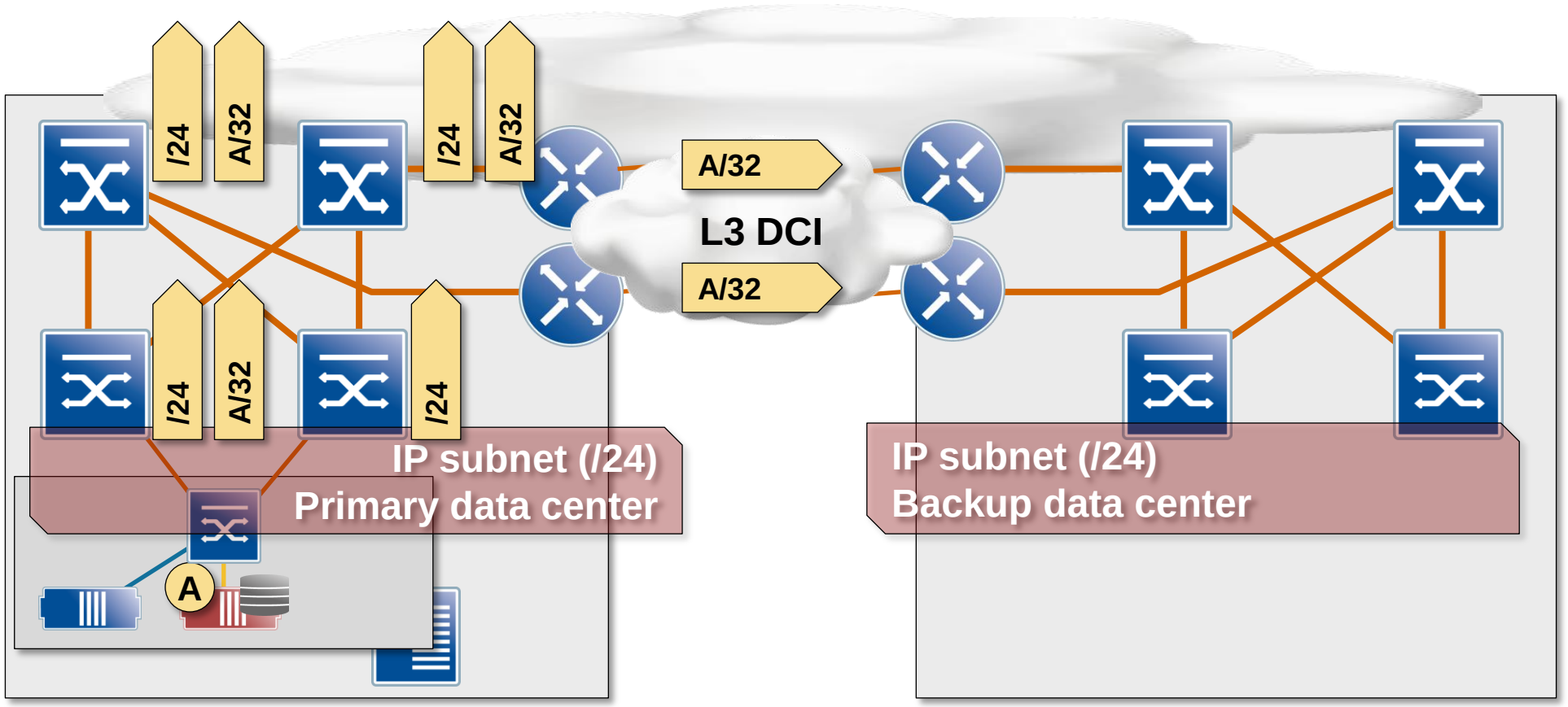


VM Mobility with L3 DCI – Network Design



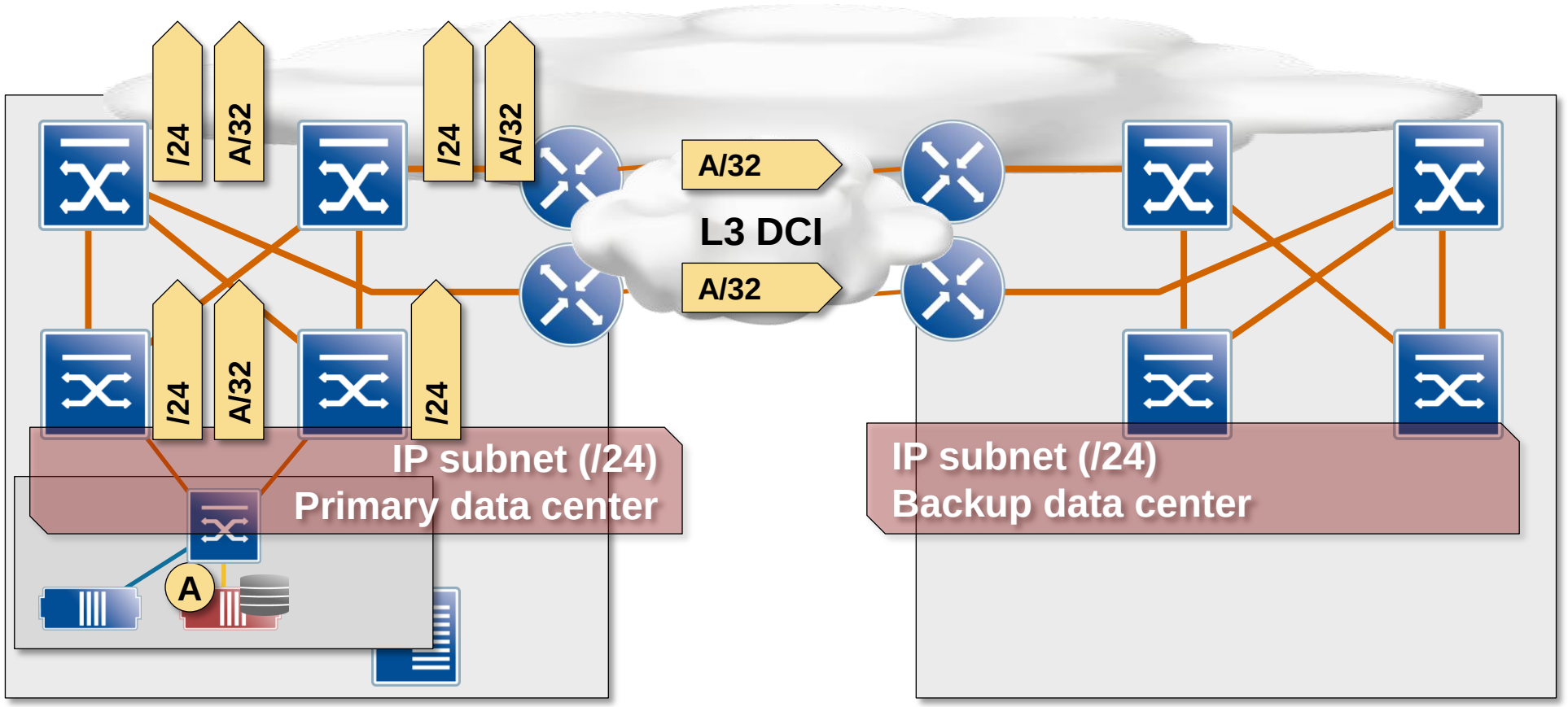
- Same IP subnet configured in both data centers
- All ToR switches run VRRP for the shared subnet

VM Mobility with L3 DCI – Network Design



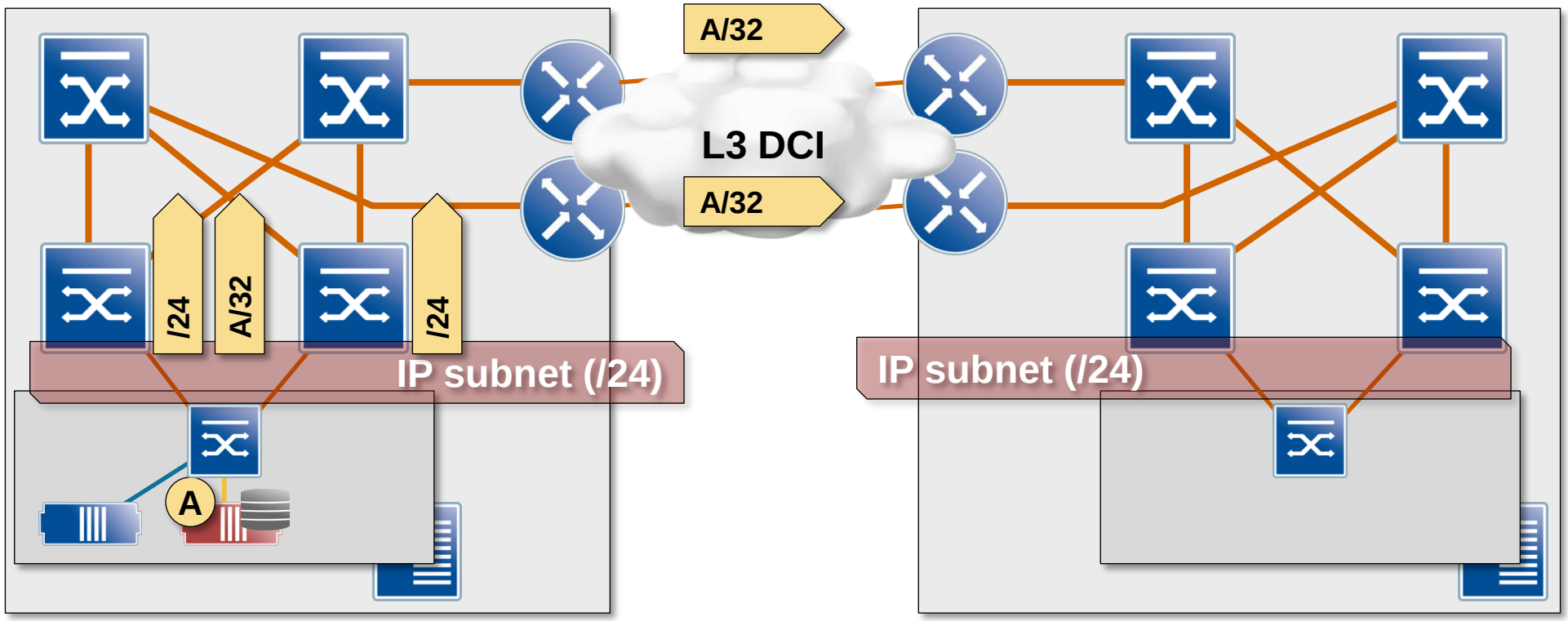
- Same IP subnet configured in both data centers
- All ToR switches run VRRP for the shared subnet
- Primary data center advertises subnet prefix and VM host routes

VM Mobility with L3 DCI – Network Design

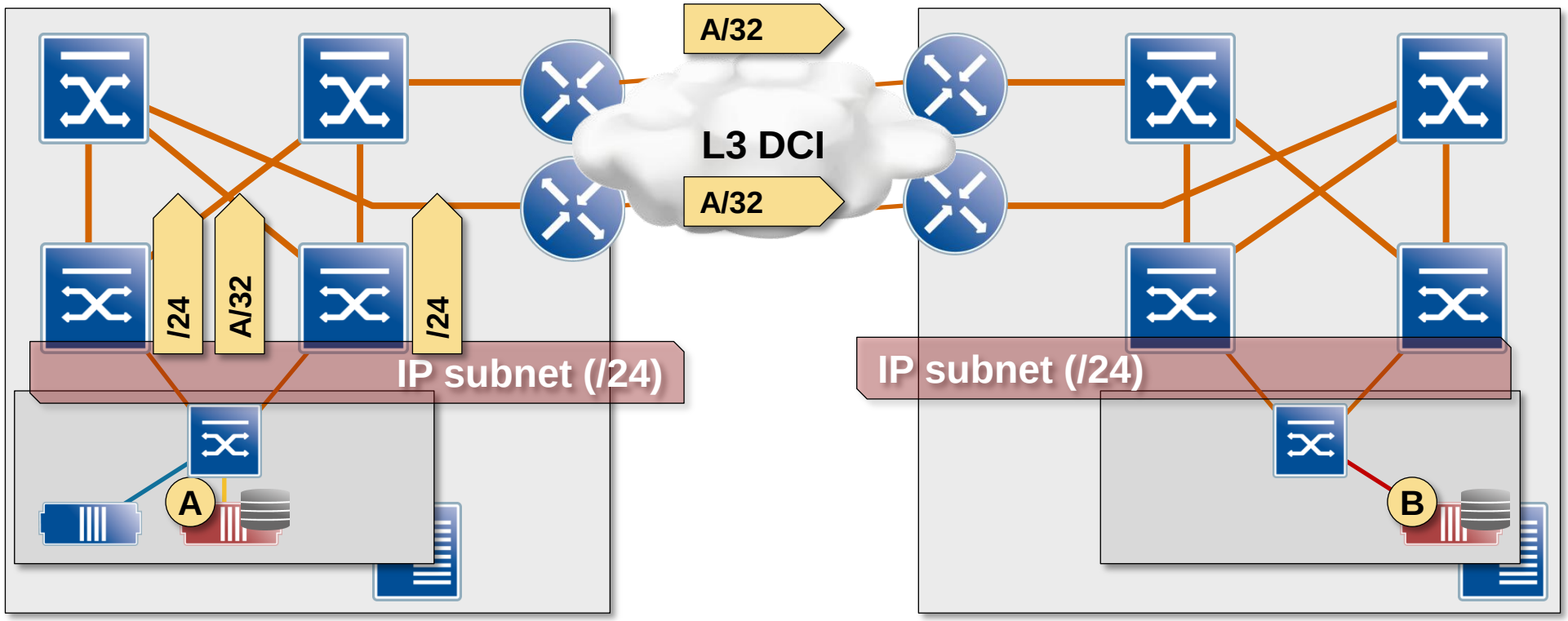


- Same IP subnet configured in both data centers
- All ToR switches run VRRP for the shared subnet
- Primary data center advertises subnet prefix and VM host routes
- Backup data center does not advertise the subnet (or uses higher cost)

VM Restarted After VM Move Event

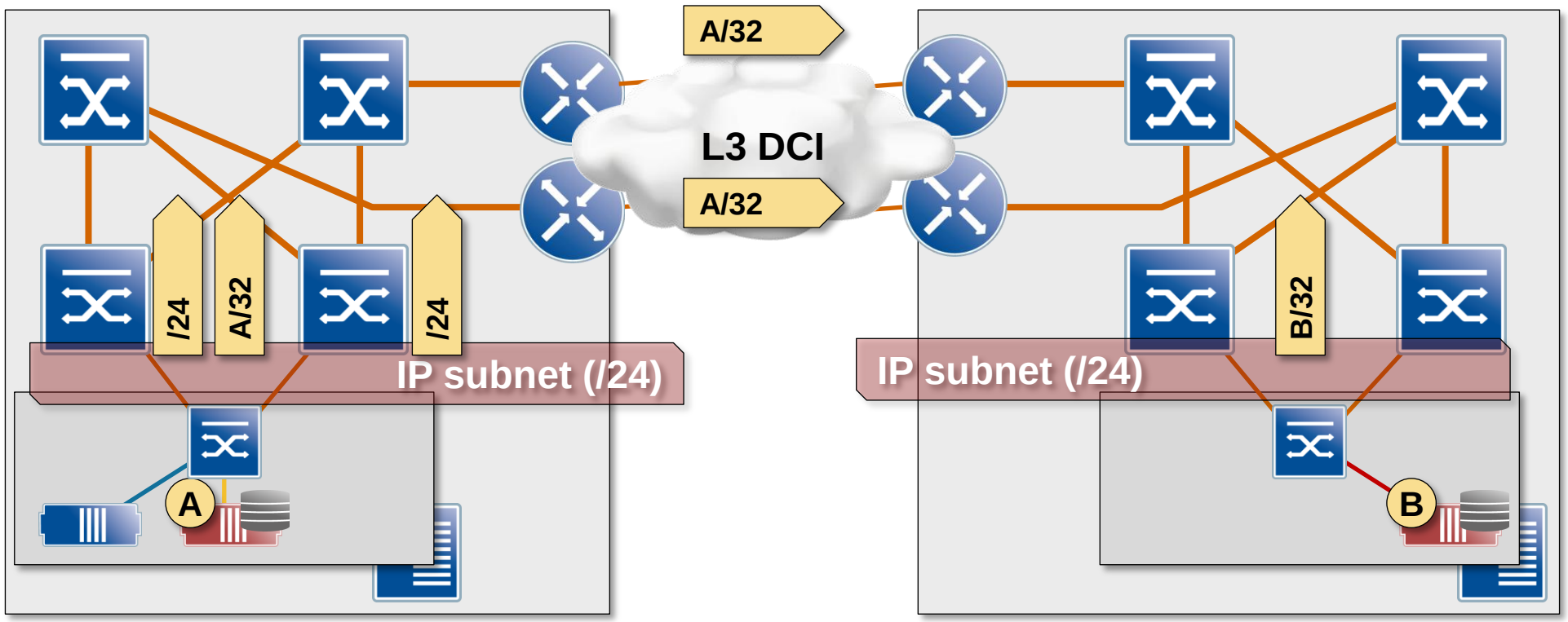


VM Restarted After VM Move Event



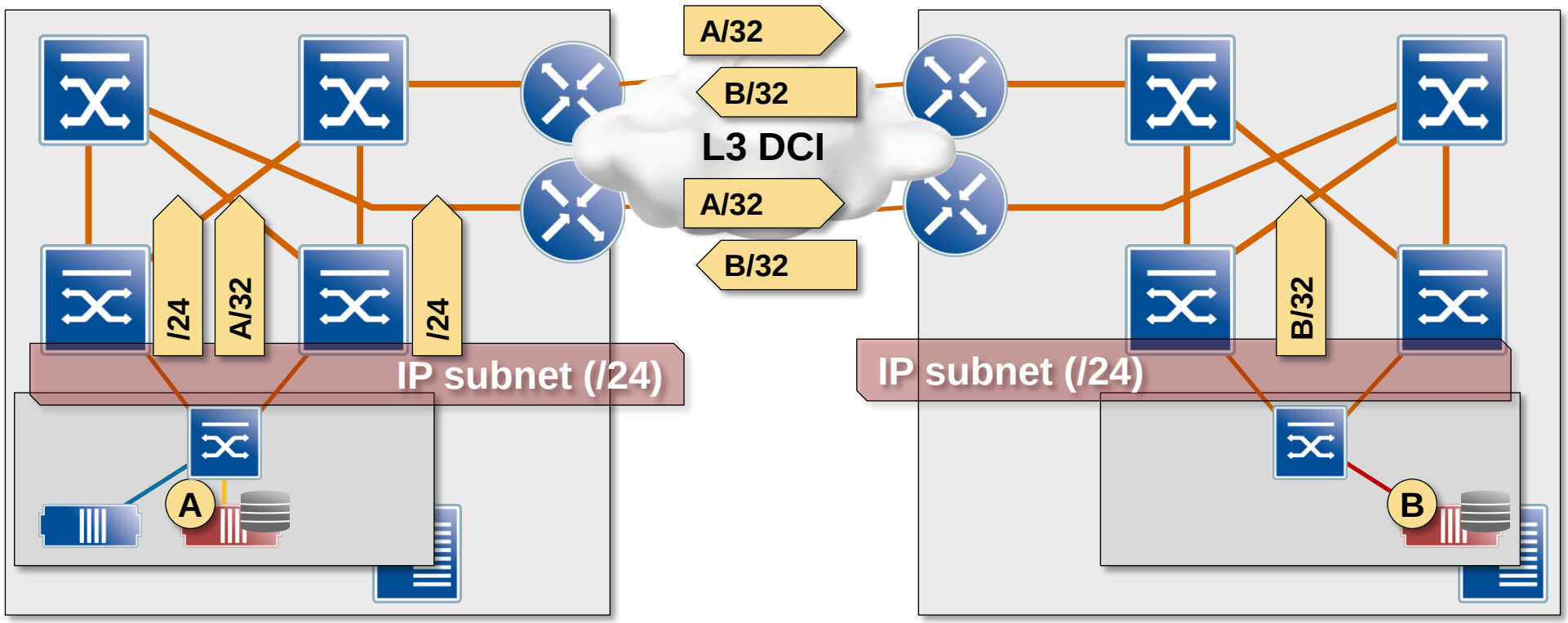
- VM-B is powered up in backup data center

VM Restarted After VM Move Event



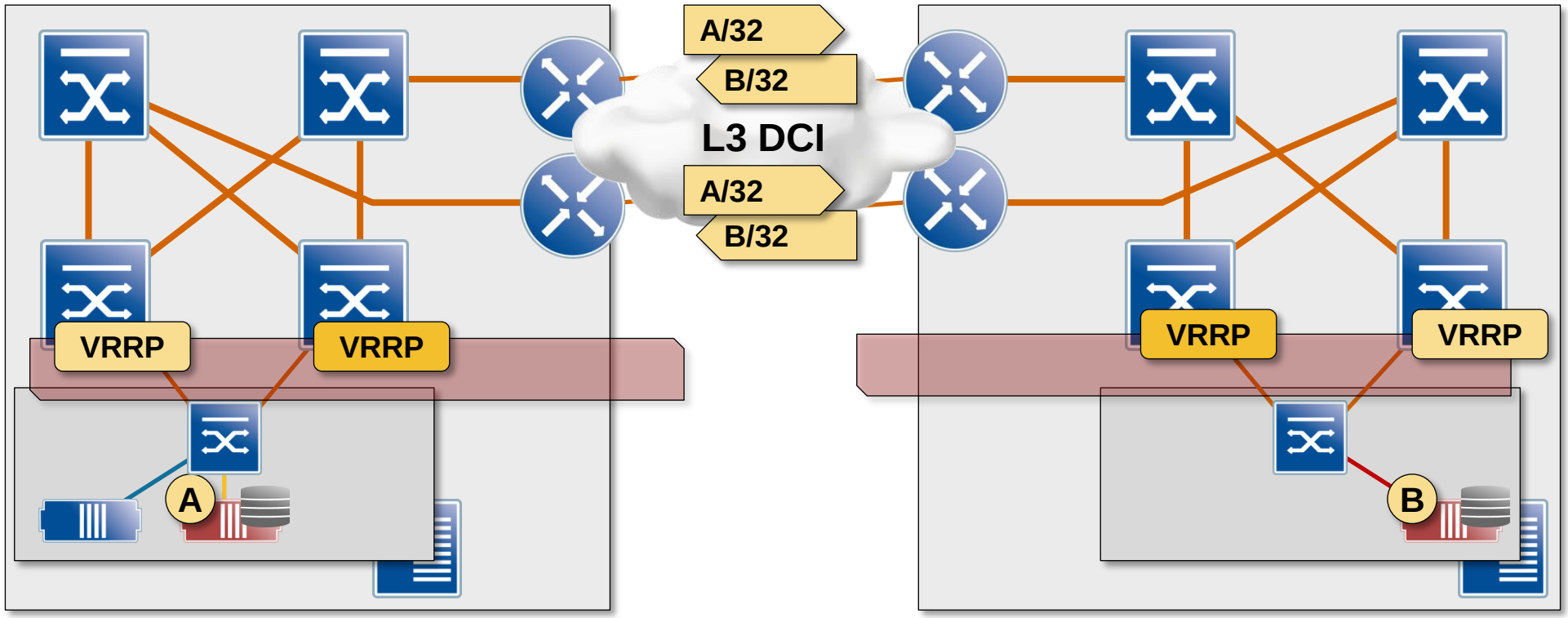
- VM-B is powered up in backup data center
- ToR switch in backup data center creates and advertises a host route

VM Restarted After VM Move Event



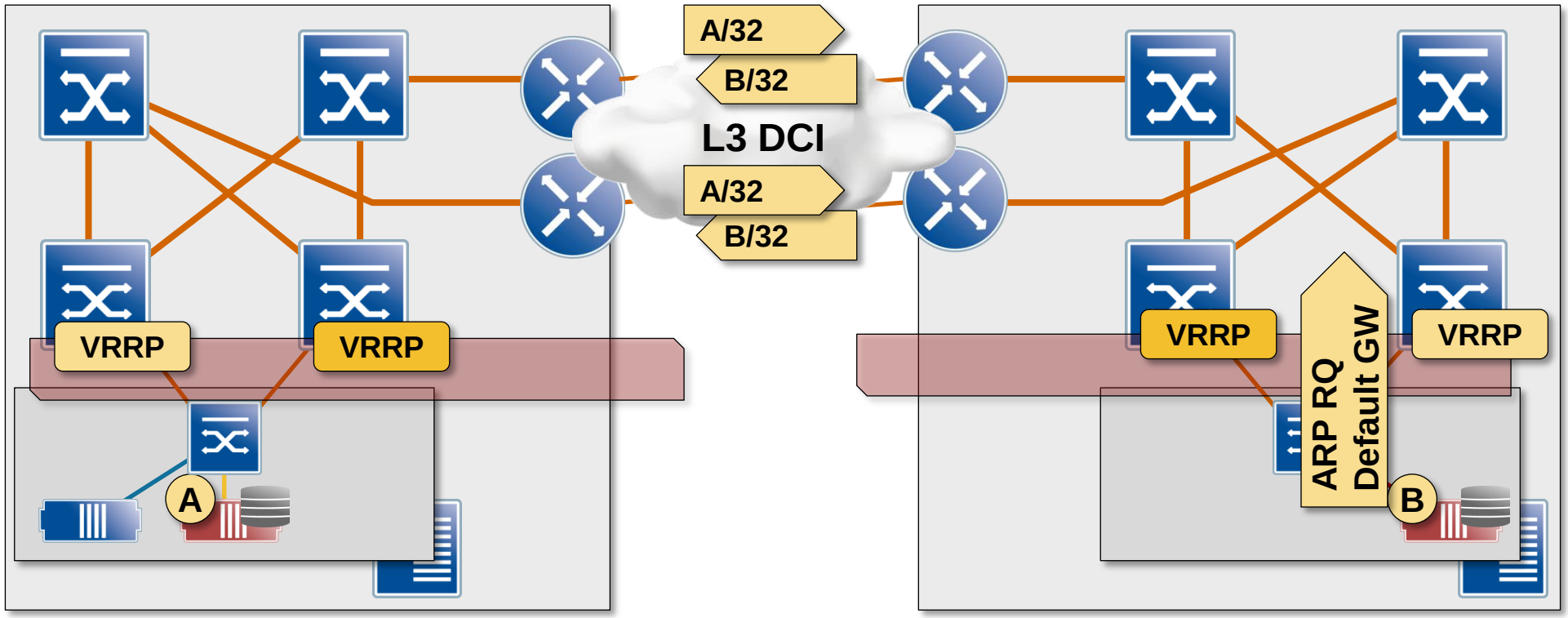
- VM-B is powered up in backup data center
- ToR switch in backup data center creates and advertises a host route
- Host routing across L3 DCI ➔ correct network-to-VM traffic flow

External Connectivity From Moved VM



Migrated VM (VM-B) has the same default gateway as before

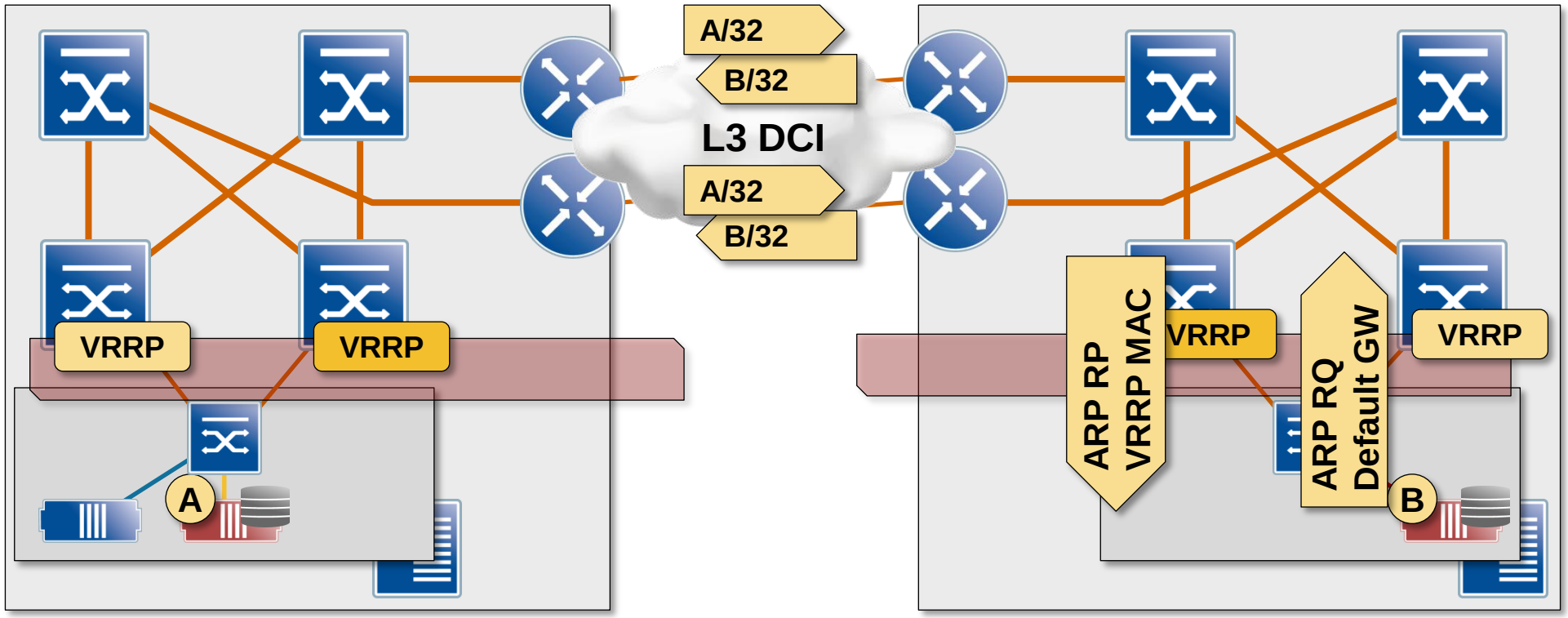
External Connectivity From Moved VM



Migrated VM (VM-B) has the same default gateway as before

- ARP request for default gateway (VRRP IP) is sent by the VM

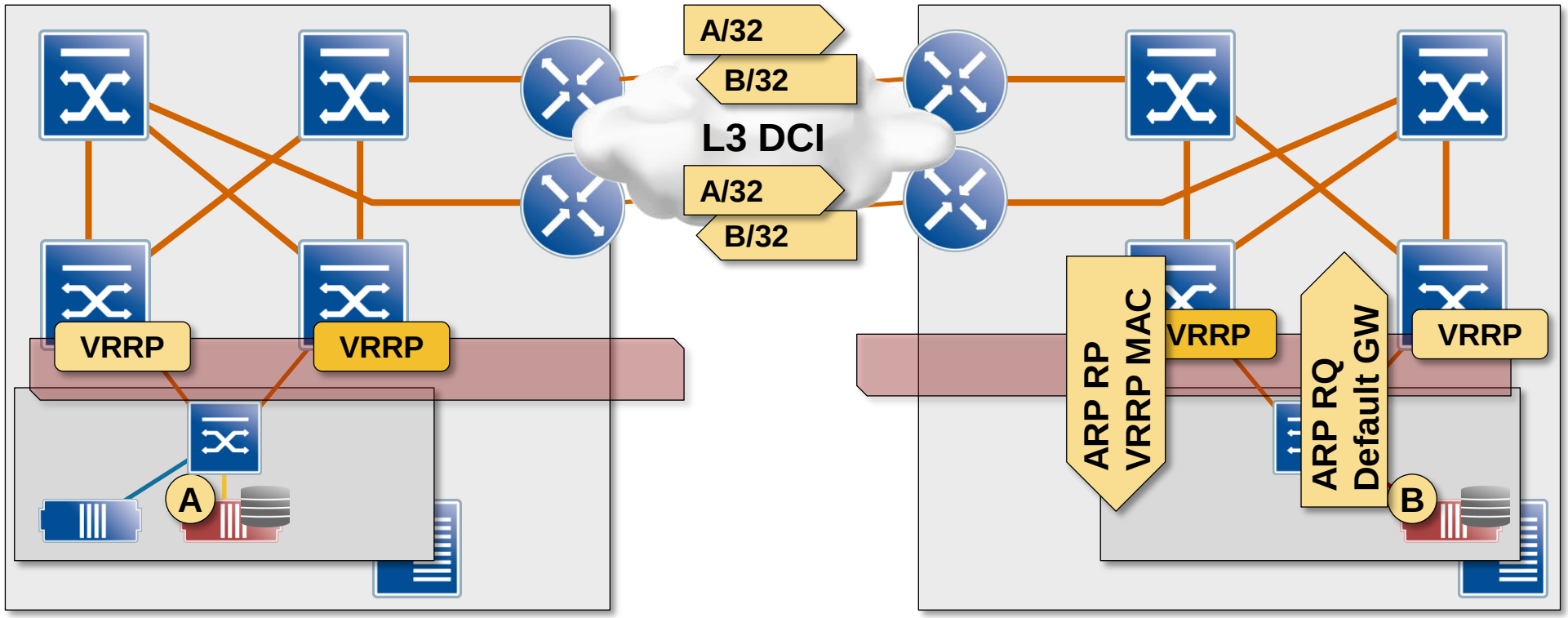
External Connectivity From Moved VM



Migrated VM (VM-B) has the same default gateway as before

- ARP request for default gateway (VRRP IP) is sent by the VM
- One of the ToR switches replies with VRRP MAC address ➔ external connectivity works

External Connectivity From Moved VM

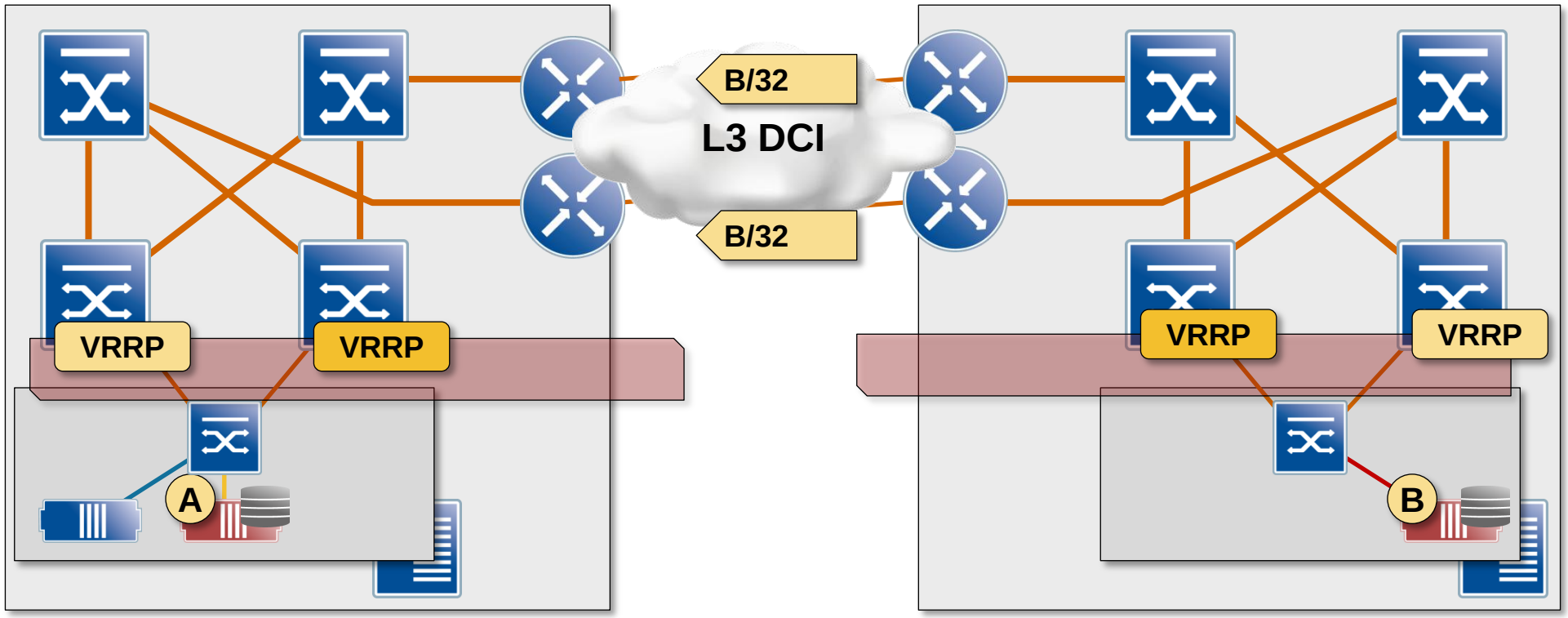


Migrated VM (VM-B) has the same default gateway as before

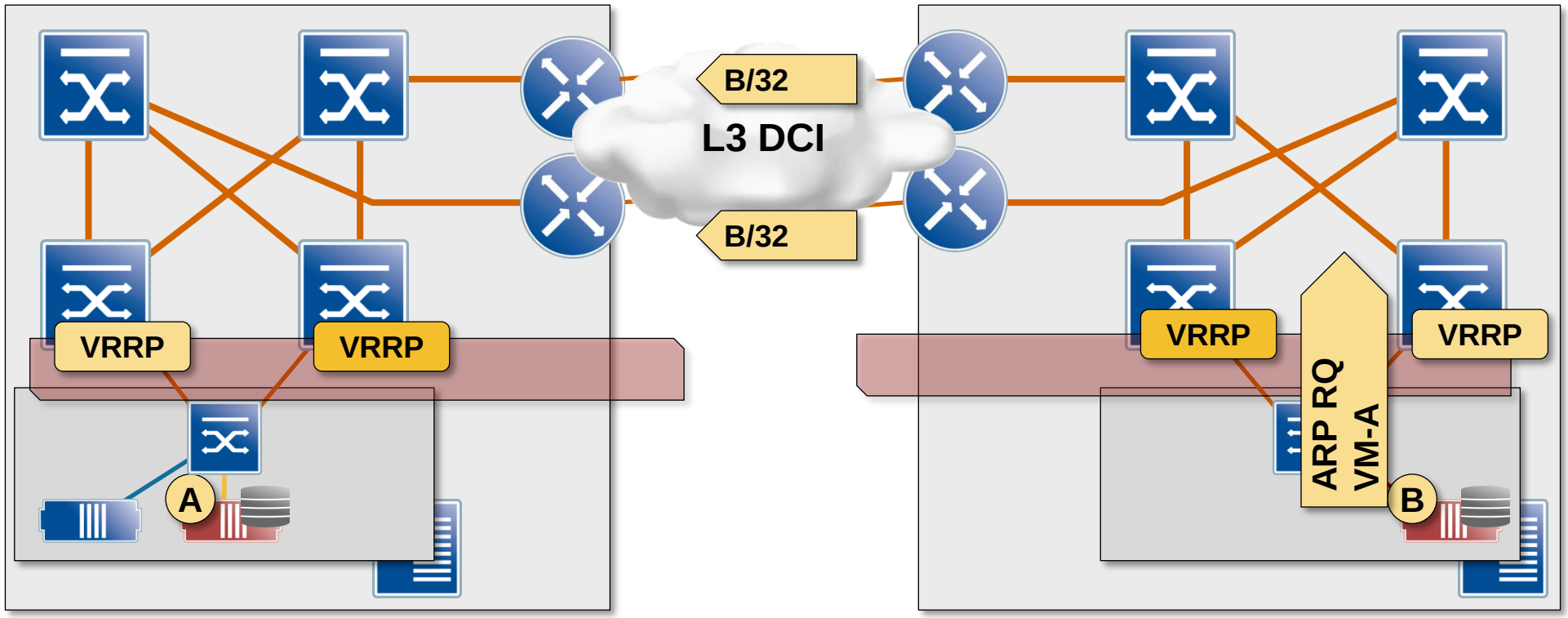
- ARP request for default gateway (VRRP IP) is sent by the VM
- One of the ToR switches replies with VRRP MAC address ➔ external connectivity works

Question: will intra-subnet traffic flow correctly?

Establishing B-to-A Connectivity

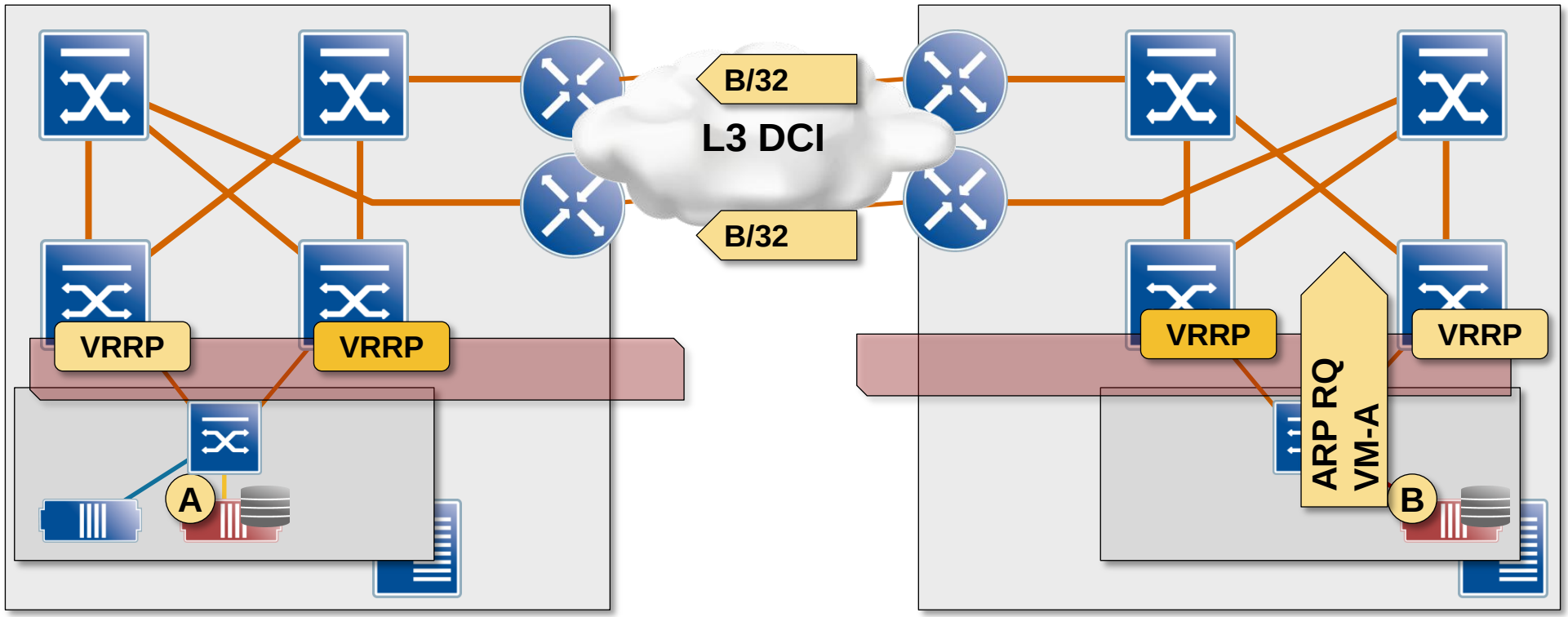


Establishing B-to-A Connectivity



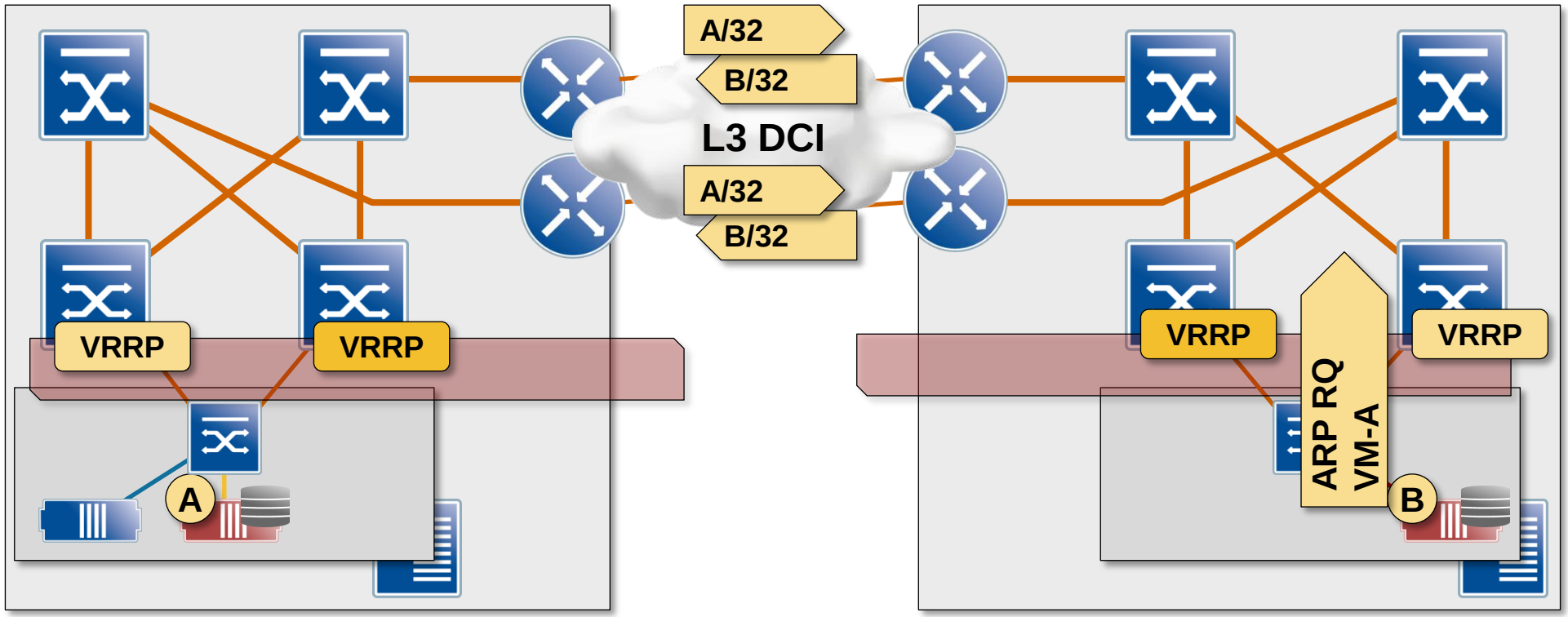
- VM-B sends ARP request for VM-A, no reply from VM-A

Establishing B-to-A Connectivity



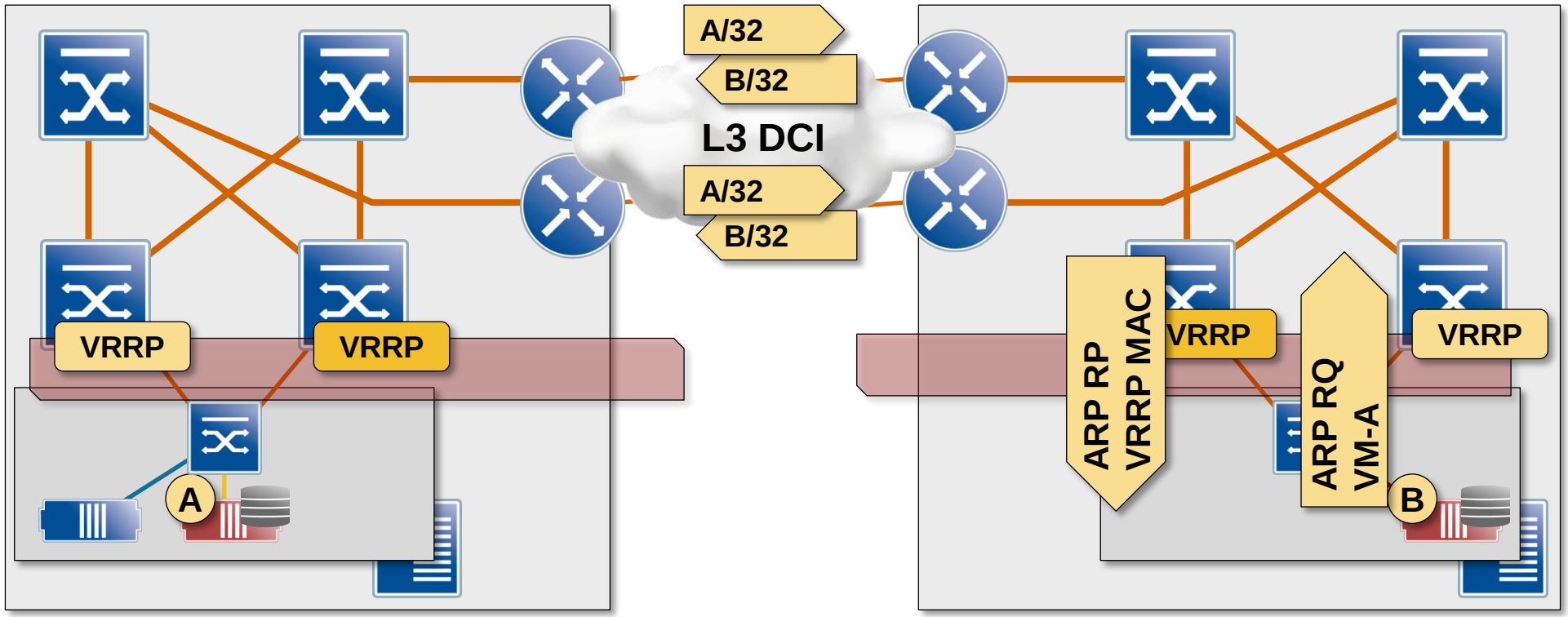
- VM-B sends ARP request for VM-A, no reply from VM-A
- ToR switch receives the ARP request

Establishing B-to-A Connectivity



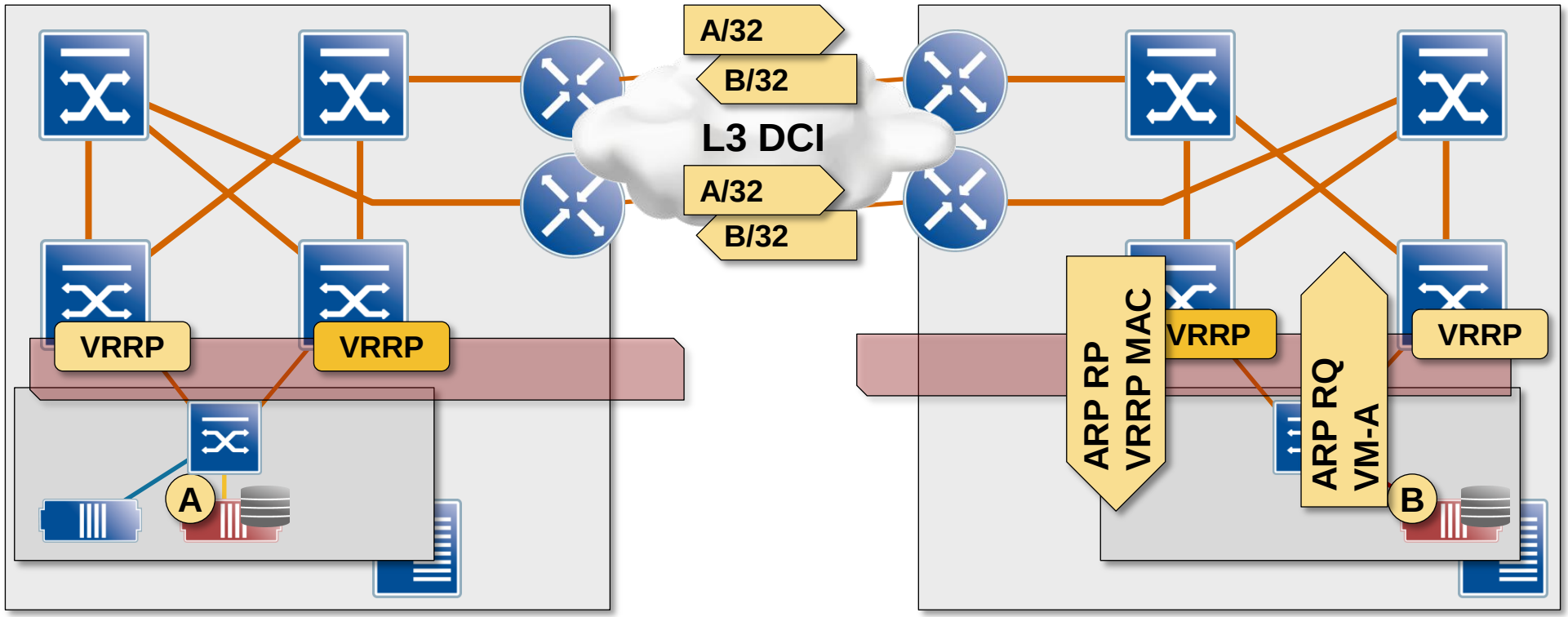
- VM-B sends ARP request for VM-A, no reply from VM-A
- ToR switch receives the ARP request
- Host route to VM-A over a different interface ➔ proxy ARP

Establishing B-to-A Connectivity



- VM-B sends ARP request for VM-A, no reply from VM-A
- ToR switch receives the ARP request
- Host route to VM-A over a different interface ➔ proxy ARP
- ToR switch replies with VRRP MAC address

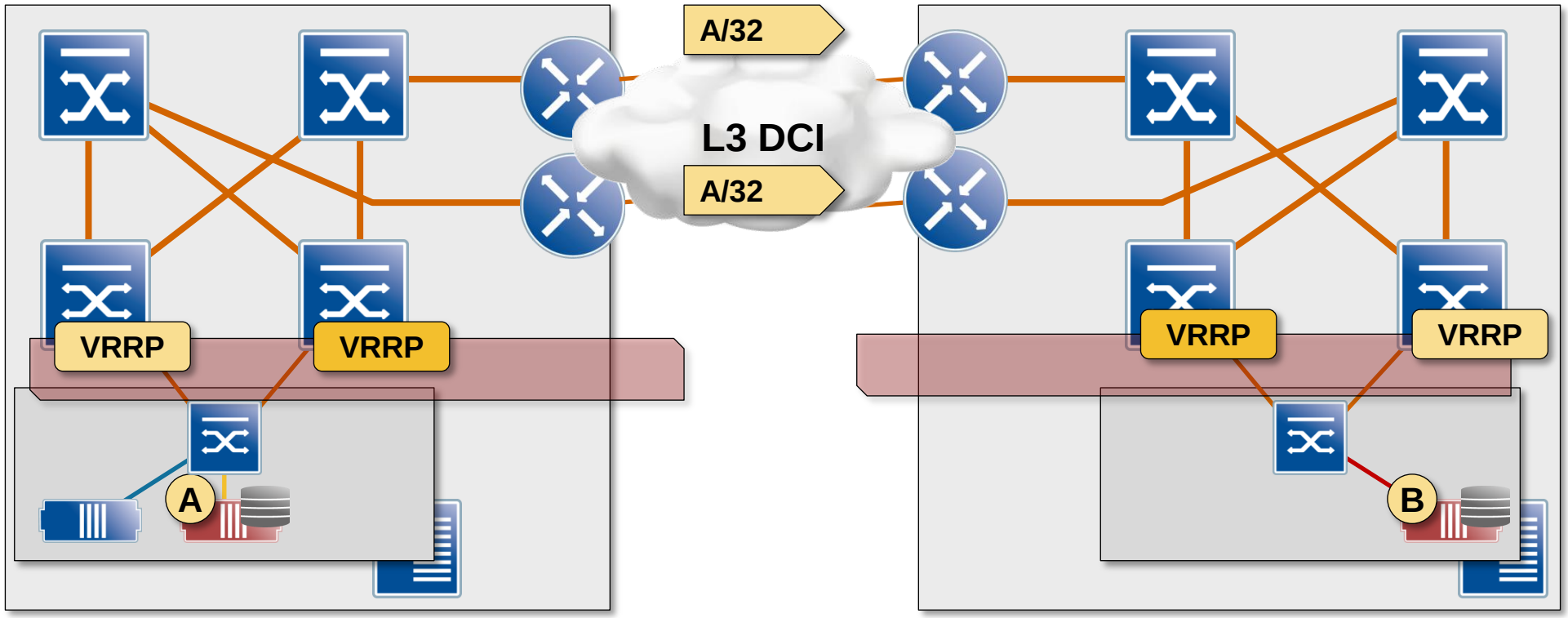
Establishing B-to-A Connectivity



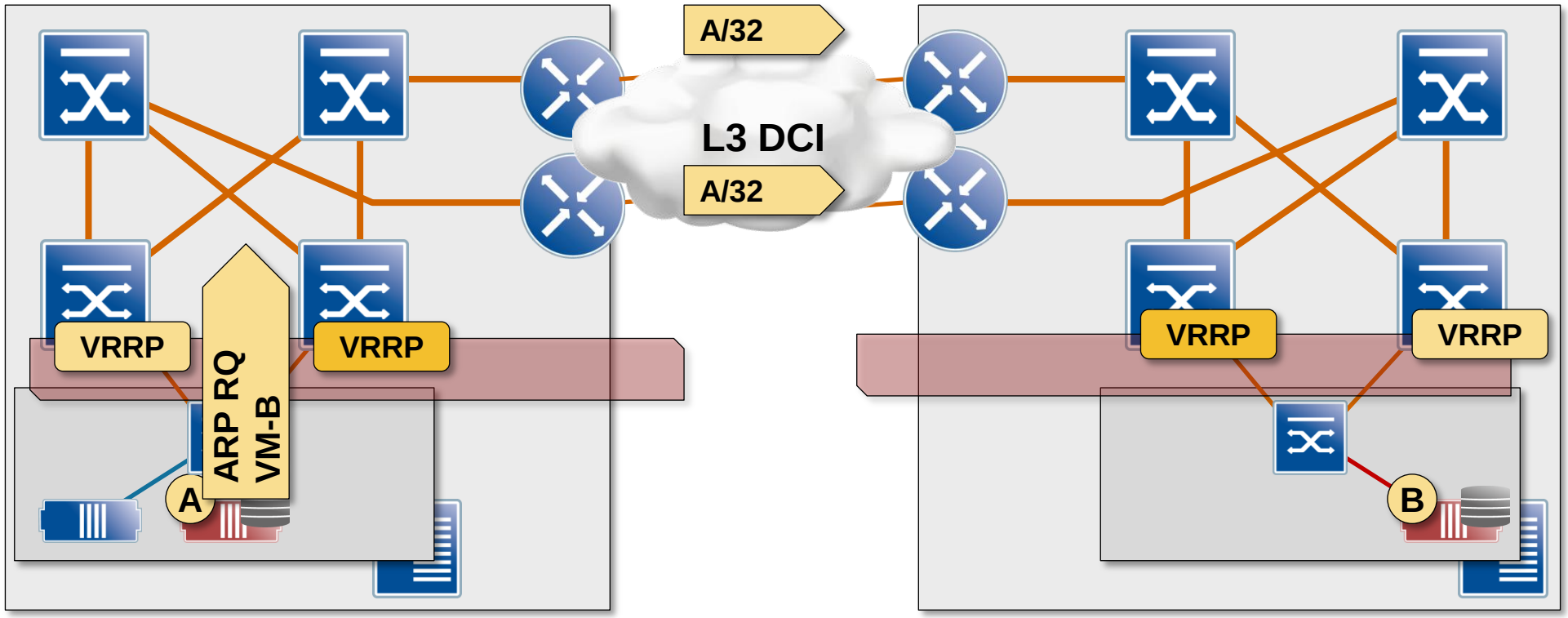
- VM-B sends ARP request for VM-A, no reply from VM-A
- ToR switch receives the ARP request
- Host route to VM-A over a different interface ➔ proxy ARP
- ToR switch replies with VRRP MAC address

Conclusion: B can send traffic to A

Establishing A-to-B Connectivity

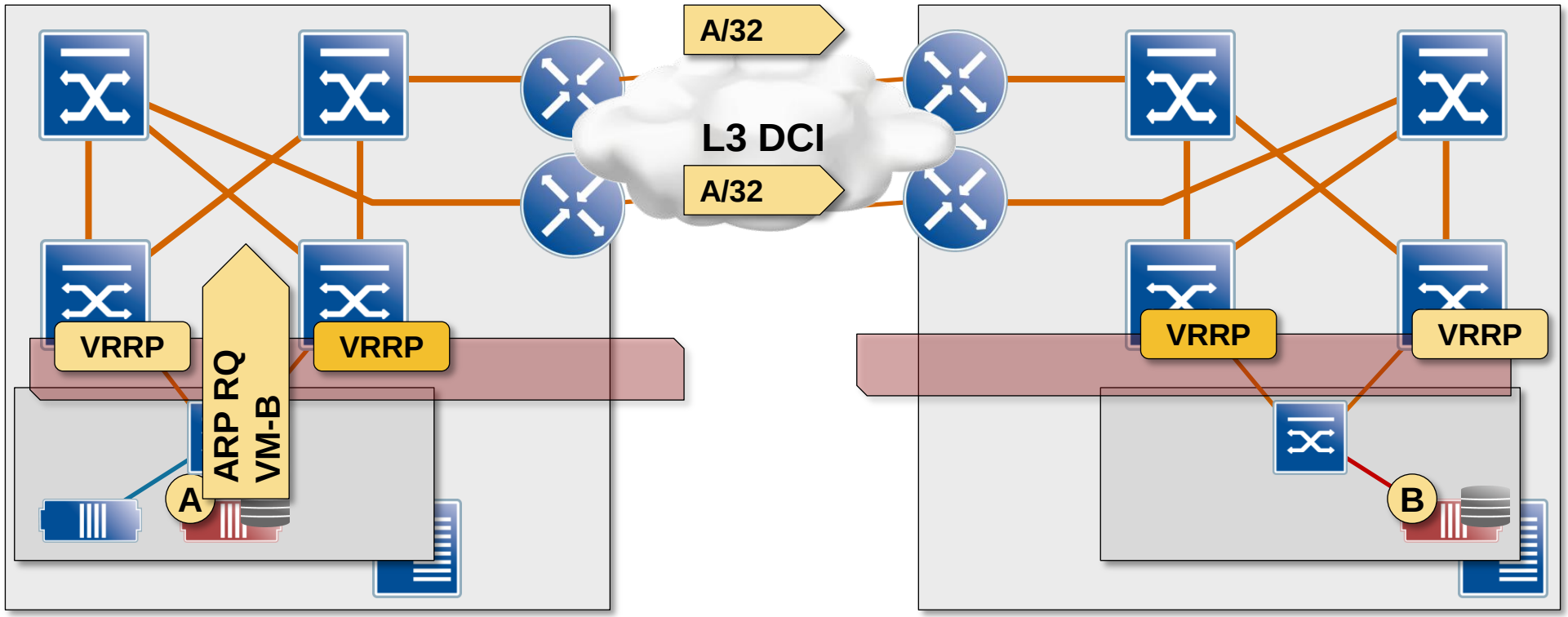


Establishing A-to-B Connectivity



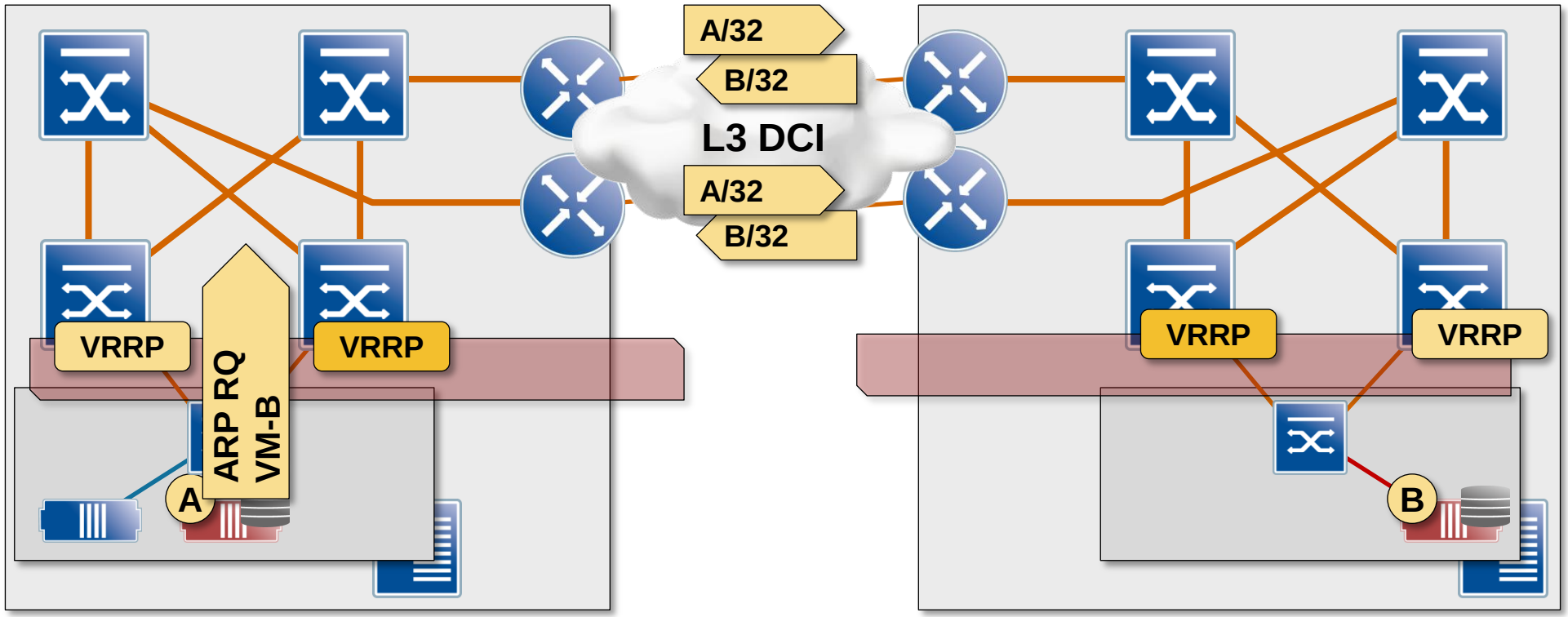
- VM-A tries to reply to VM-B, sends ARP request for VM-B

Establishing A-to-B Connectivity



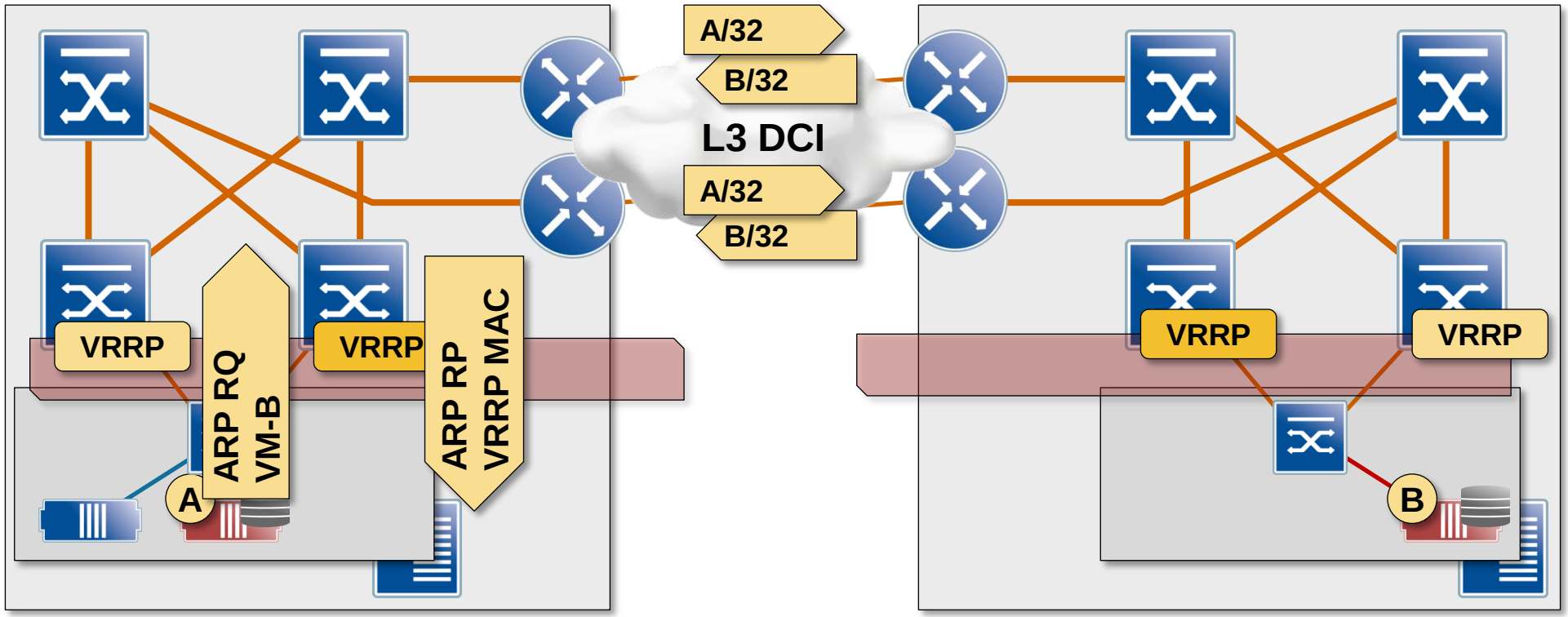
- VM-A tries to reply to VM-B, sends ARP request for VM-B
- ToR switch receives the ARP request

Establishing A-to-B Connectivity



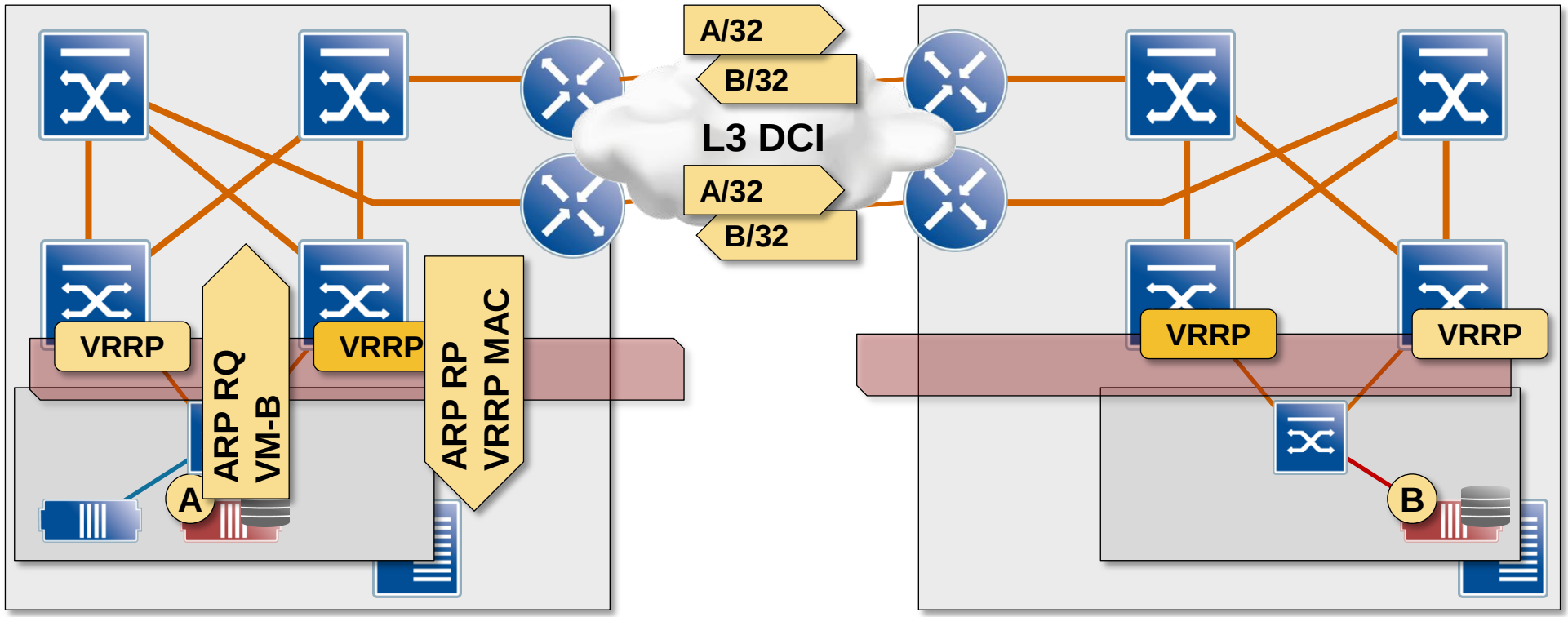
- VM-A tries to reply to VM-B, sends ARP request for VM-B
- ToR switch receives the ARP request
- Host route to VM-B over a different interface ➔ proxy ARP

Establishing A-to-B Connectivity



- VM-A tries to reply to VM-B, sends ARP request for VM-B
- ToR switch receives the ARP request
- Host route to VM-B over a different interface ➔ proxy ARP
- ToR switch replies with VRRP MAC address

Establishing A-to-B Connectivity



- VM-A tries to reply to VM-B, sends ARP request for VM-B
- ToR switch receives the ARP request
- Host route to VM-B over a different interface ➔ proxy ARP
- ToR switch replies with VRRP MAC address

Conclusion: A can send traffic to B (assuming ARP entry for VM-B expired)

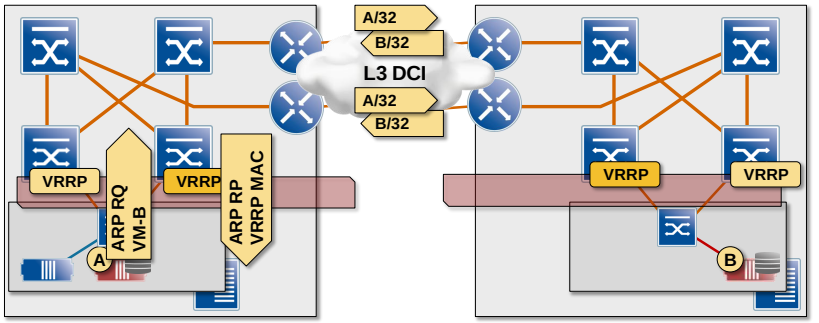
VM Migration Across L3 DCI – Summary

Same subnet in both data centers

- Advertised with higher cost in backup data center
- Same VRRP IP and MAC addresses in both data centers
- Host routing for all VMs in the split subnet

Connectivity to/from migrated VM

- Most ARP entries point to VRRP MAC address (proxy ARP)
- VM can communicate within a subnet and with outside world
- Host routing ensures optimal inbound traffic flow



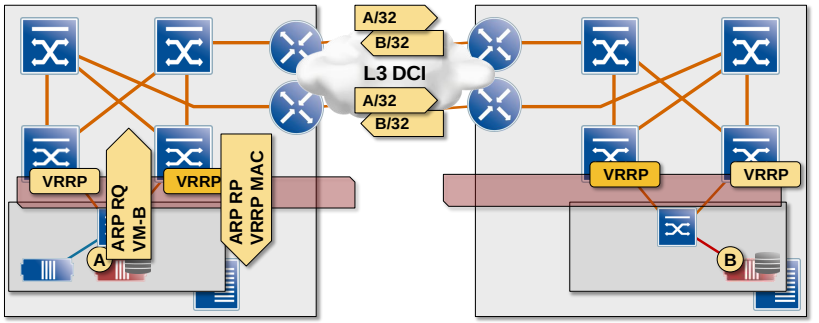
VM Migration Across L3 DCI – Summary

Same subnet in both data centers

- Advertised with higher cost in backup data center
- Same VRRP IP and MAC addresses in both data centers
- Host routing for all VMs in the split subnet

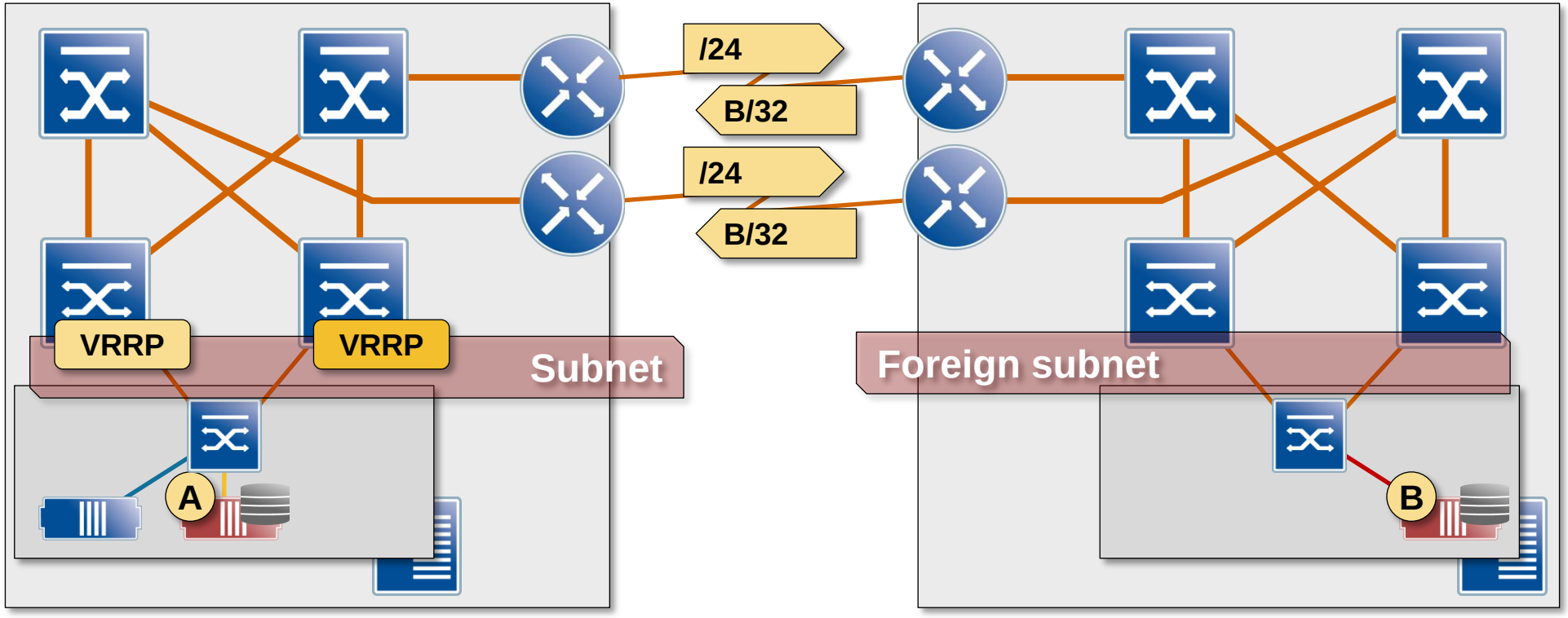
Connectivity to/from migrated VM

- Most ARP entries point to VRRP MAC address (proxy ARP)
- VM can communicate within a subnet and with outside world
- Host routing ensures optimal inbound traffic flow



Split subnet in two data centers is confusing. Can we do better?

VM Mobility into Foreign Subnet



New functionality in Enterasys switches:

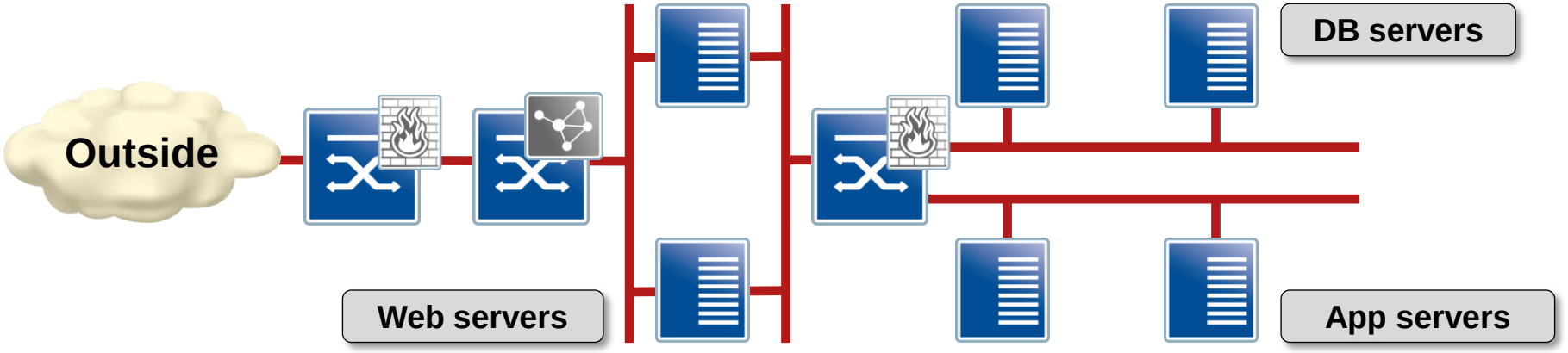
- Host routing for out-of-subnet addresses
- Local Proxy ARP for out-of-subnet requests

Result: VM can be migrated into a foreign subnet

No need to advertise host routes for VMs in primary data center

Interacting With L4-7 Services

Typical Application Architecture

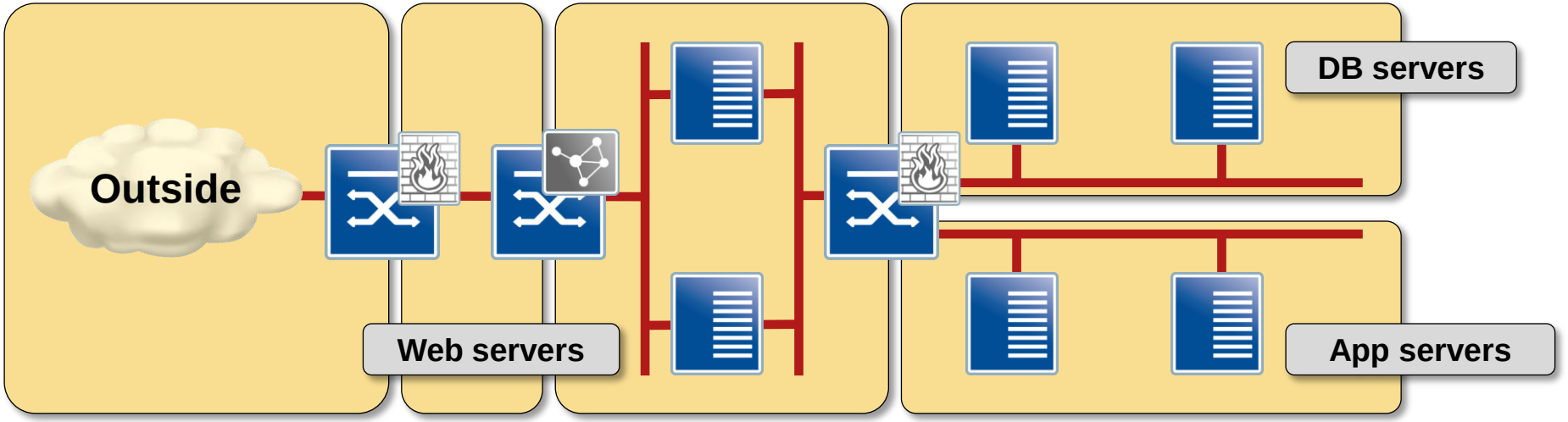


- Application in multiple zones (tiers)
- Each tier = security zone = IP subnet (VLAN)
- Application tiers linked with L4-7 devices: firewalls, NAT and load balancers

Challenges

- Chokepoints
- Multiple routing domains
- Gateway pinning ➔ trombones

Routing Domains with L4-7 Appliances



Every application tier = separate routing domain

- Default gateway for web servers != default gateway for DB servers

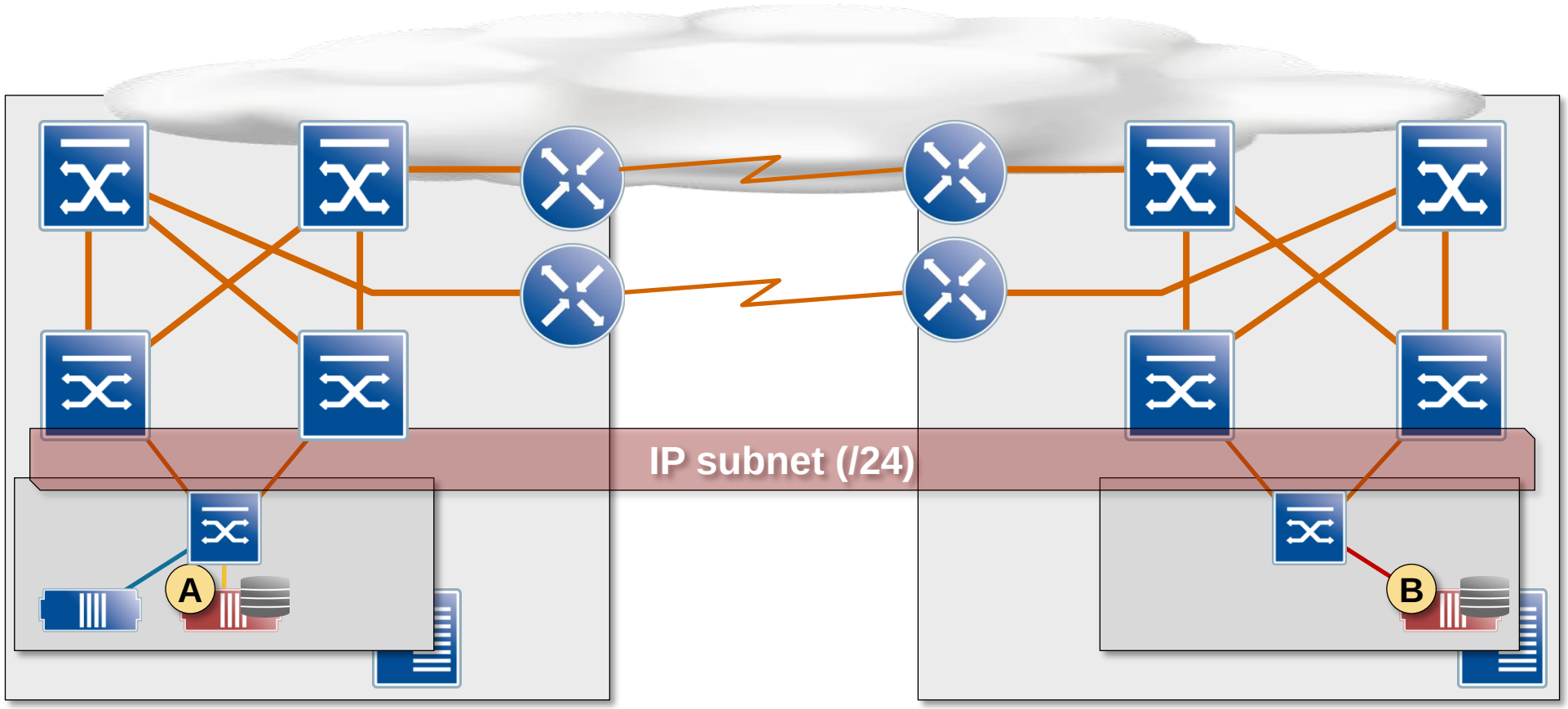
Easy to implement in L2-only world

- Switches don't participate in L3 forwarding

L3 solutions

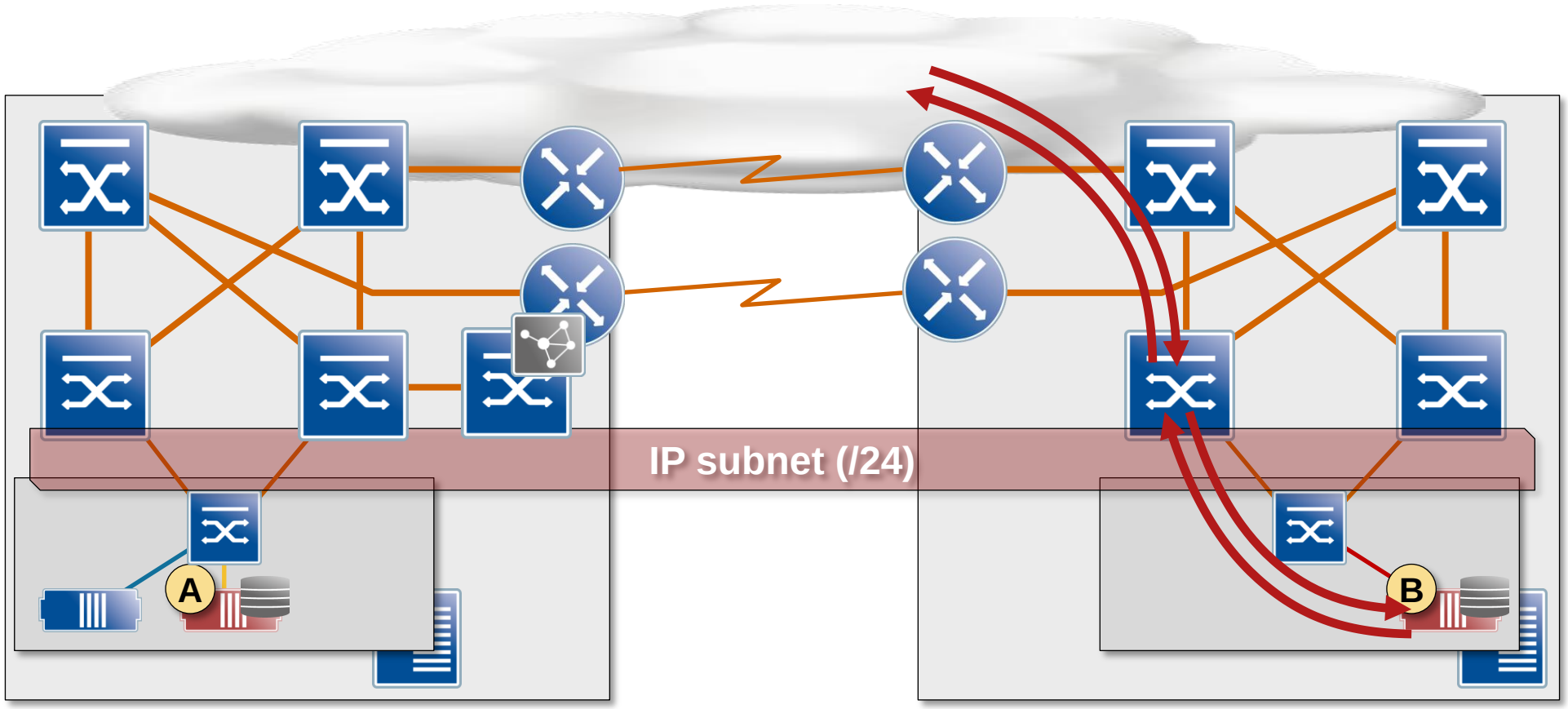
- Virtual Routing and Forwarding tables (L3 routing domains)
- MPLS/VPN in large/scalable networks

Gateway Pinning and Traffic Trombones



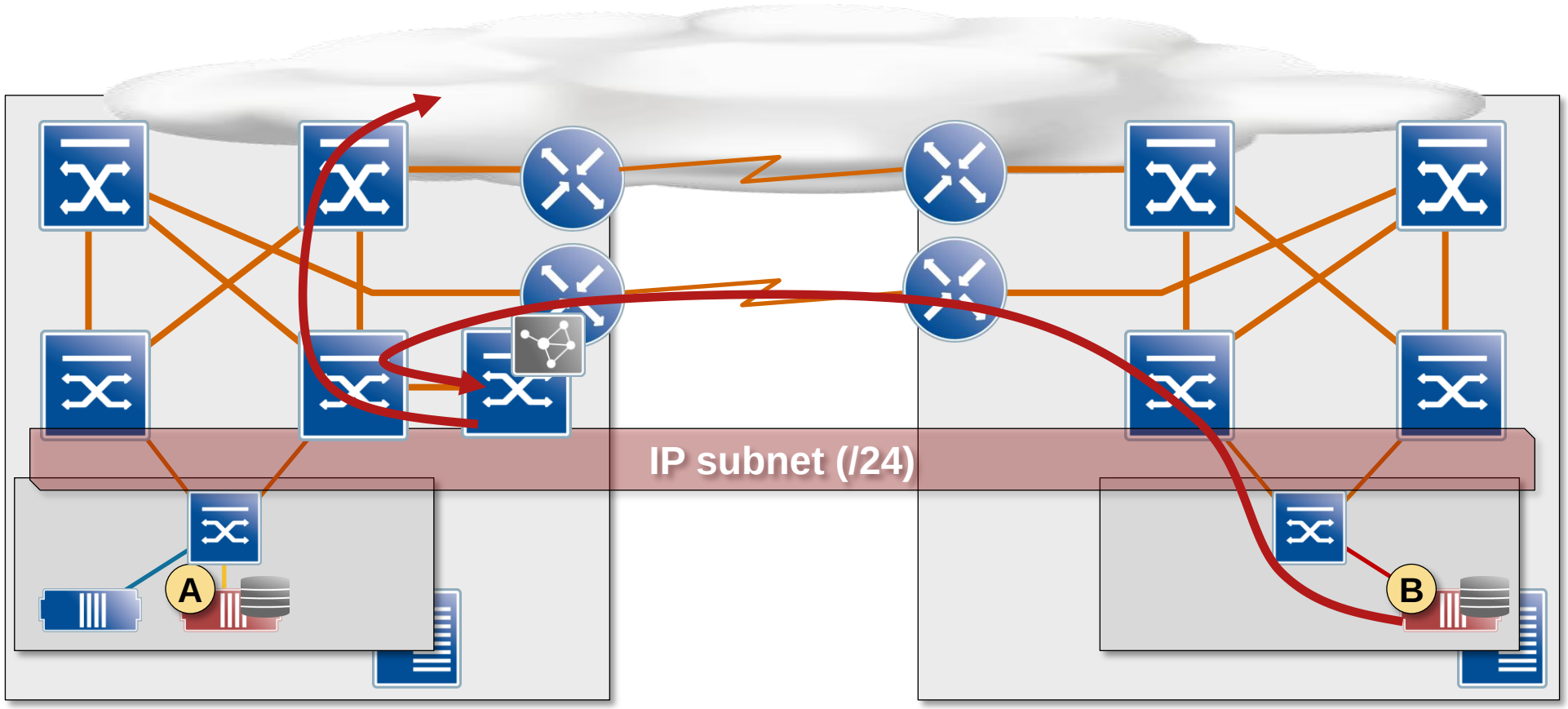
- Without L4-7 appliances in the path: optimal traffic flow

Gateway Pinning and Traffic Trombones



- Without L4-7 appliances in the path: optimal traffic flow
- Adding a firewall to the segment

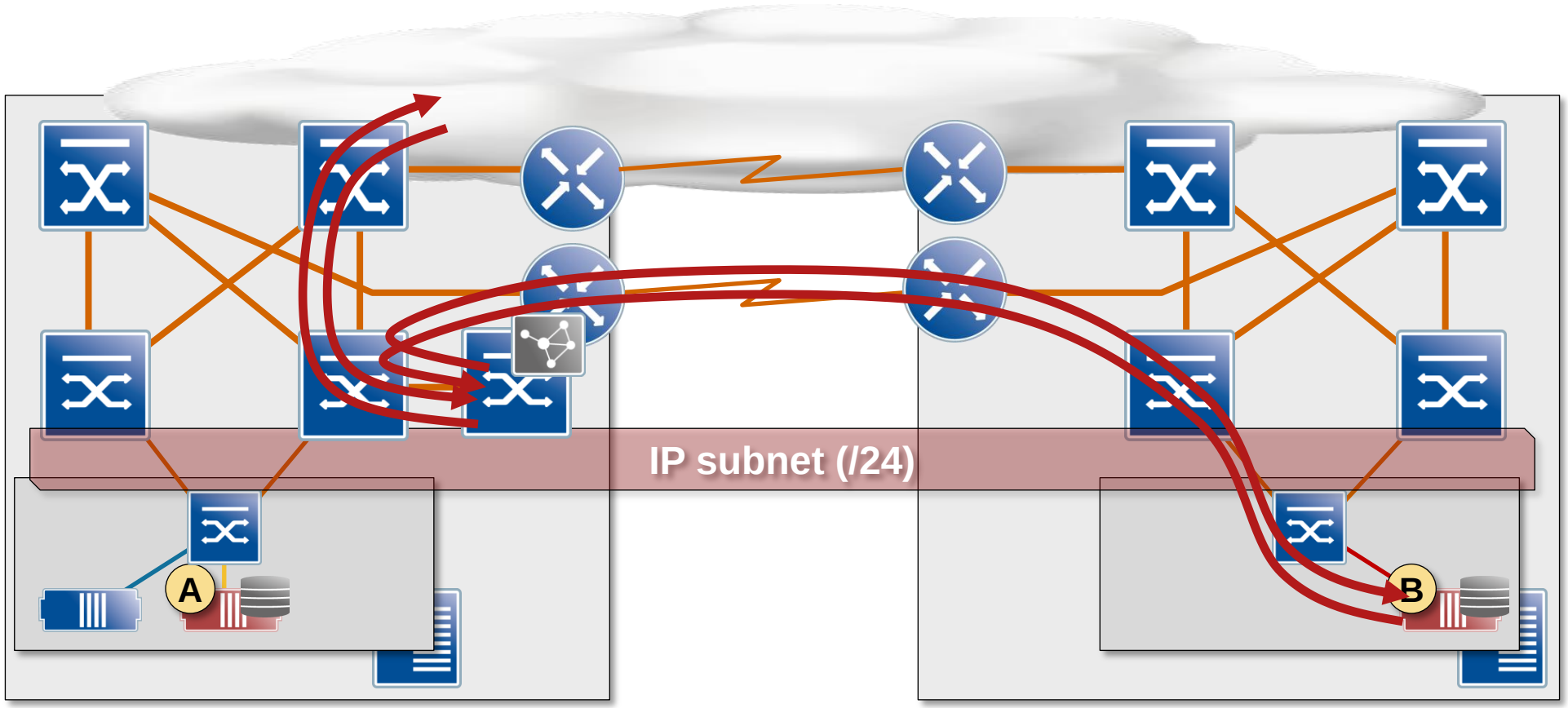
Gateway Pinning and Traffic Trombones



- Without L4-7 appliances in the path: optimal traffic flow
- Adding a firewall to the segment

  ➔ Traffic must flow through L4-7 appliances, resulting in traffic trombones

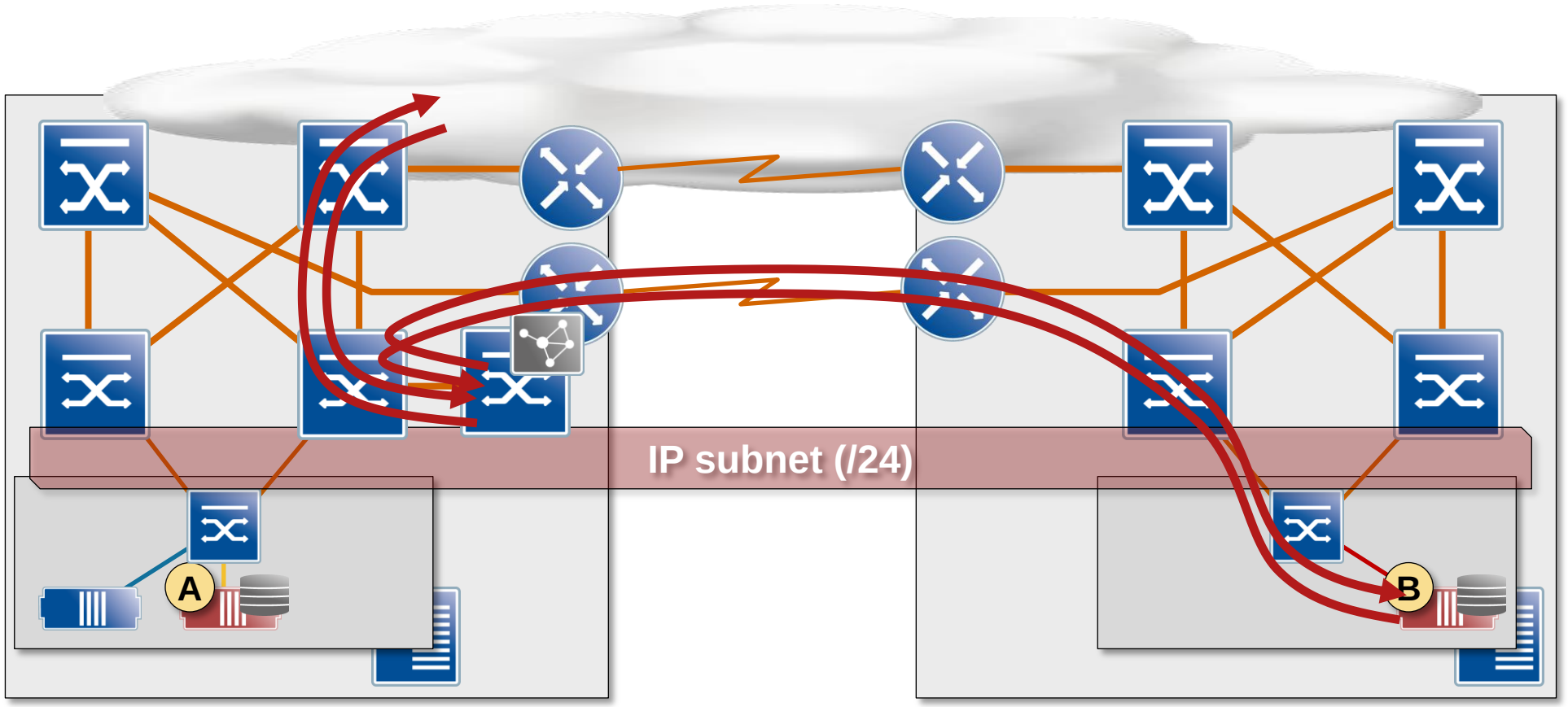
Gateway Pinning and Traffic Trombones



- Without L4-7 appliances in the path: optimal traffic flow
- Adding a firewall to the segment

  ➔ Traffic must flow through L4-7 appliances, resulting in traffic trombones

Gateway Pinning and Traffic Trombones



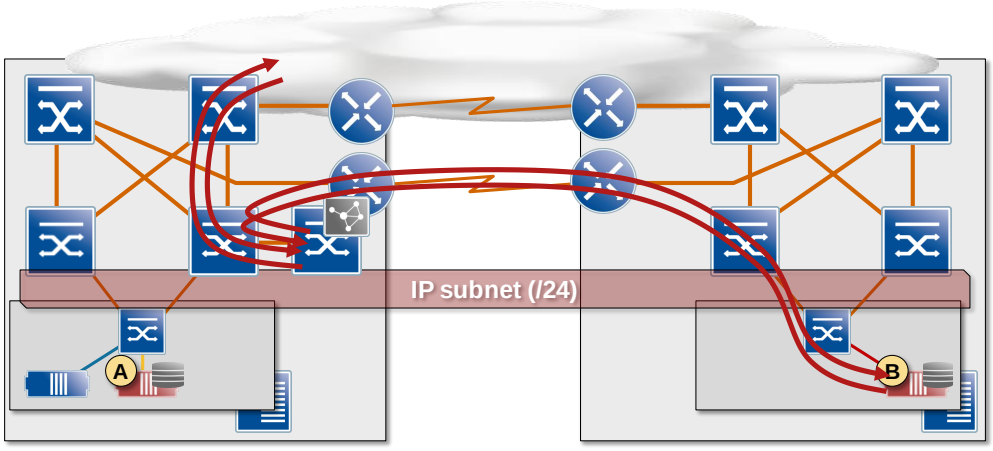
- Without L4-7 appliances in the path: optimal traffic flow
- Adding a firewall to the segment

  ➔ Traffic must flow through L4-7 appliances, resulting in traffic trombones

Transparent firewalls are no different from inter-subnet firewalls

Gateway Pinning: Distributed Firewalls

Replace central firewalls
with distributed reflexive ACLs



Hypervisor-based solutions:

- VMware vShield App,
  Juniper vGW, Cisco VSG
- VMware NVP

ToR-based solutions (Enterasys)

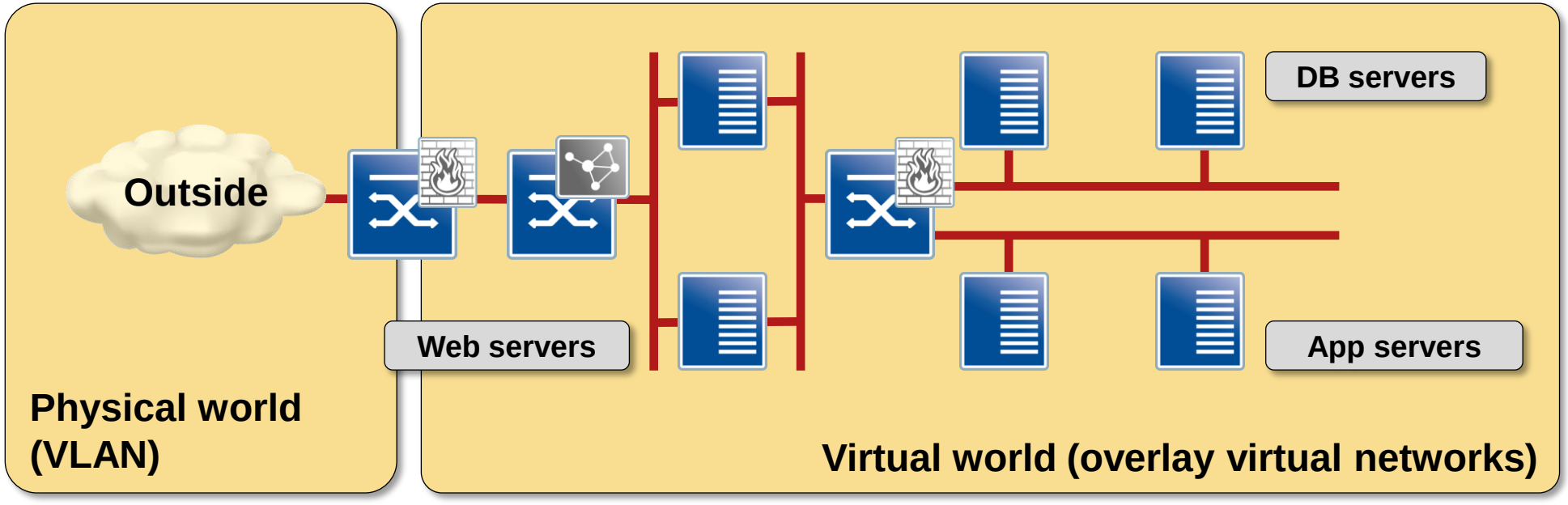
- Per-VM ACLs and QoS in ToR switches
- ACL and QoS applied based on VM MAC address (multiple ACLs per interface)
- ACL and QoS moved autmatically with the VM (for VM tracking and orchestration with VM management needs DCM)

No readily available solution for load balancers ➔ virtualize and move them

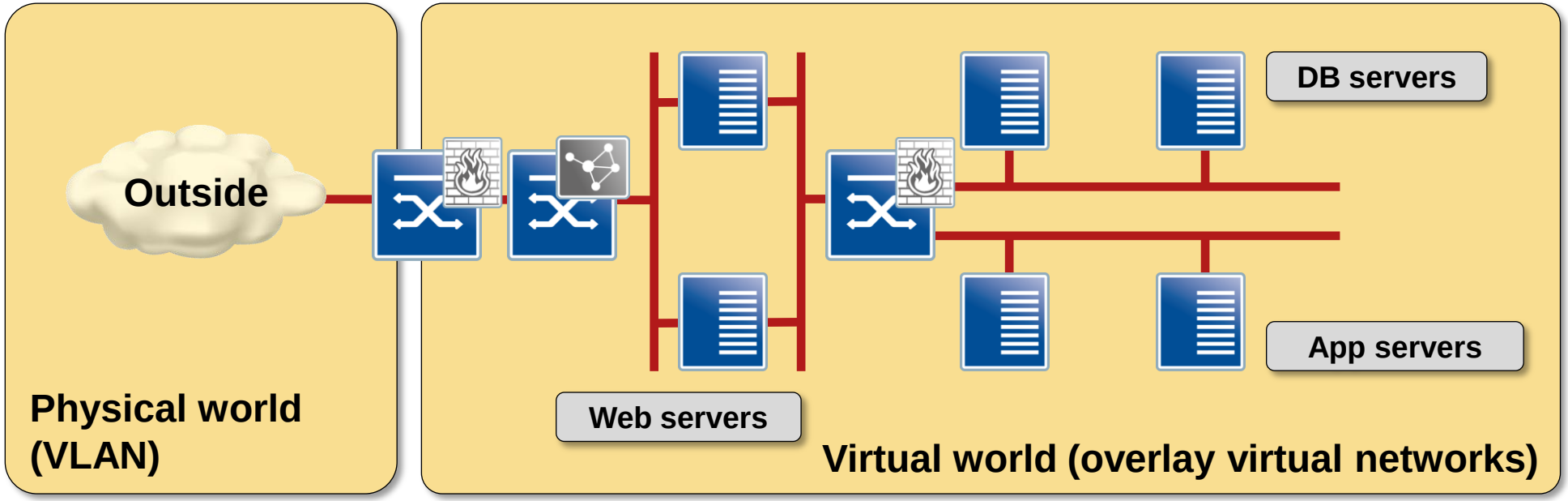
Interaction with Overlay Virtual Networking

Overlay Virtual Networking Principles

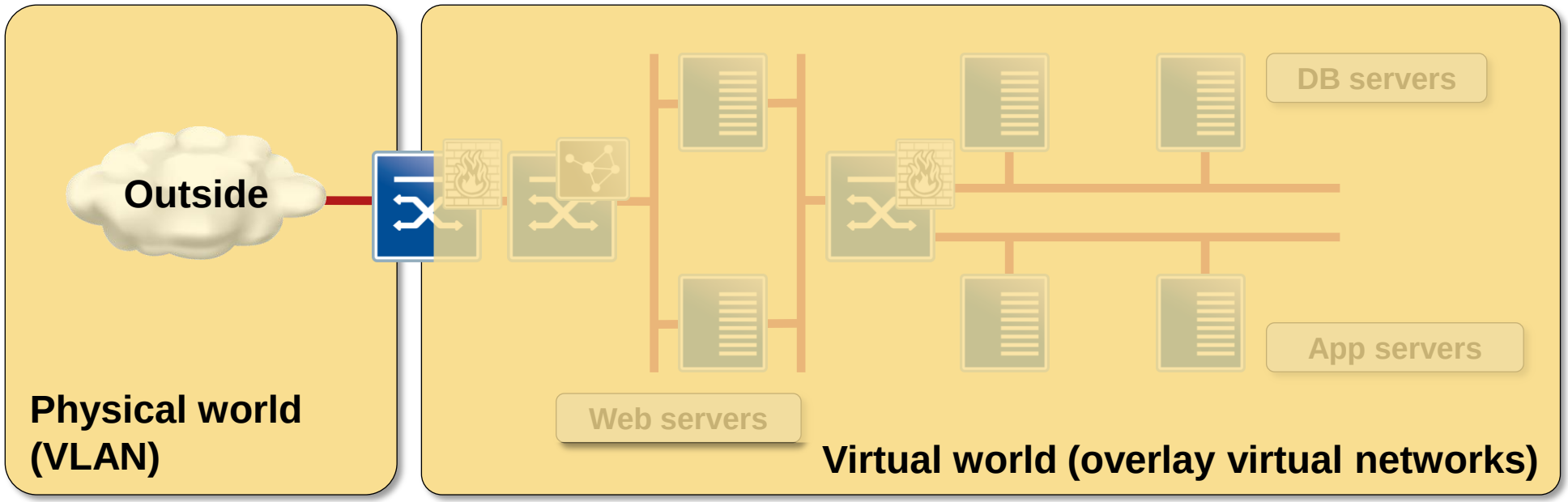
- Physical network provides simple IP transport
- Virtual segments implemented in hypervisor switches
- Virtual network traffic encapsulated in IP datagrams (MAC-over-X-over-IP)
  ➔ just another IP application
- L4-7 functionality implemented in hypervisors or VM appliances
- No VLANs, no VLAN-related troubles, no L2 DCI



Overlay Virtual Networking and Fabric/Host Routing

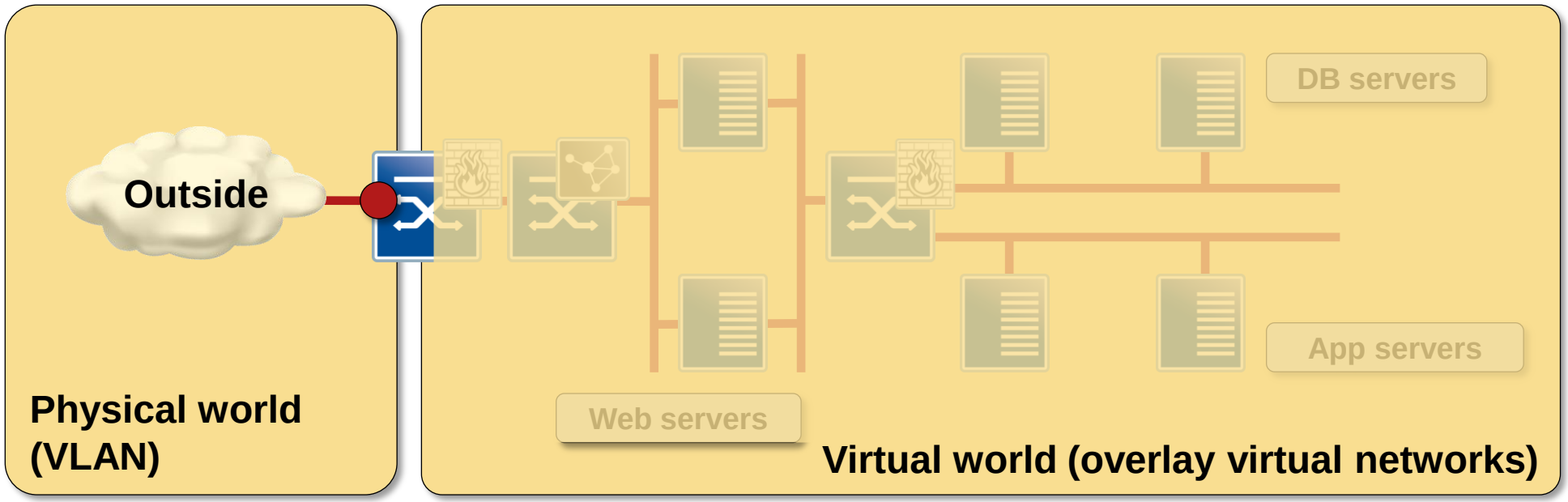


Overlay Virtual Networking and Fabric/Host Routing



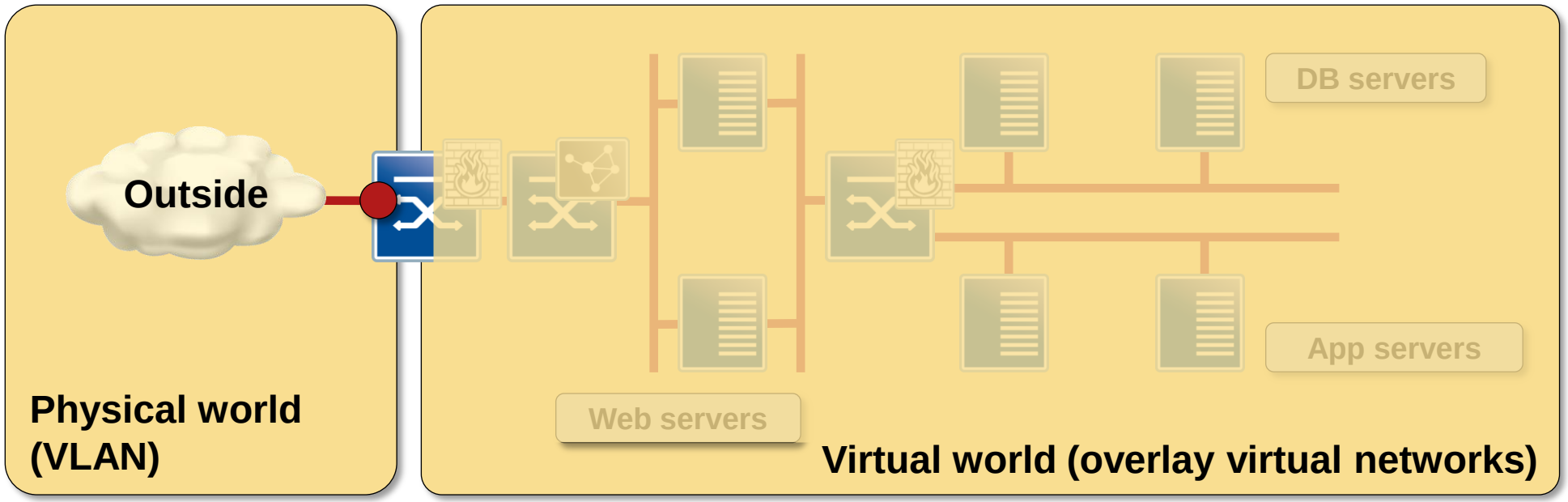
- Overlay networks simplify physical network design, deployment and operations (and increase the complexity of the hypervisor software)

Overlay Virtual Networking and Fabric/Host Routing



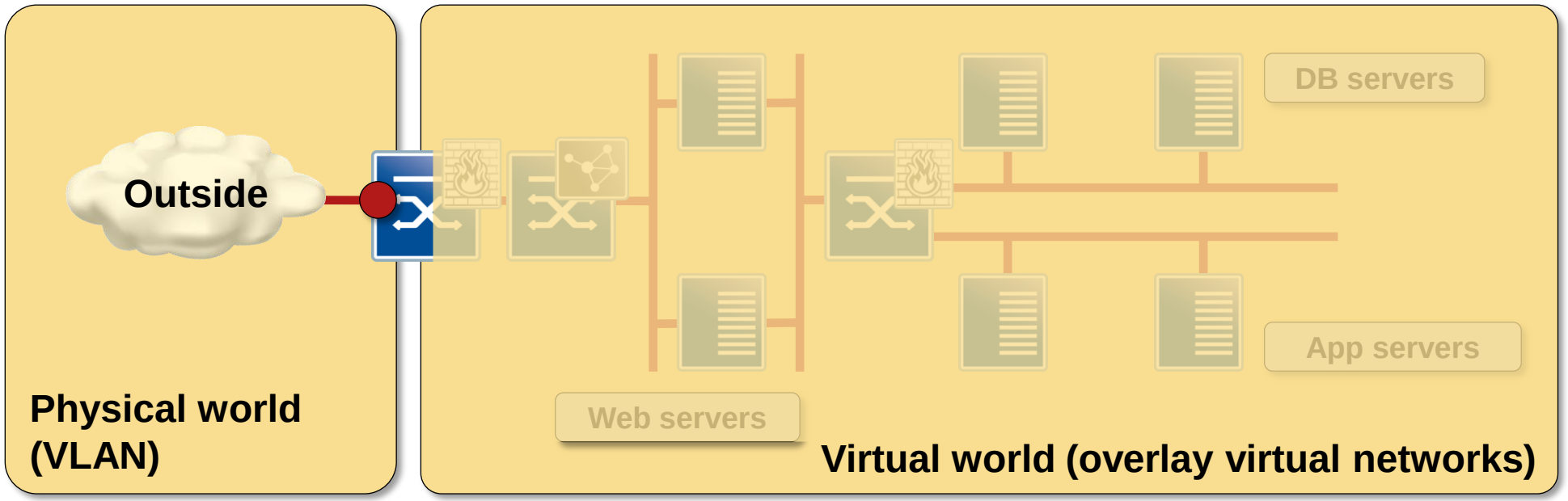
- Overlay networks simplify physical network design, deployment and operations (and increase the complexity of the hypervisor software)
- Interaction with physical world: outside IP address

Overlay Virtual Networking and Fabric/Host Routing



- Overlay networks simplify physical network design, deployment and operations (and increase the complexity of the hypervisor software)
- Interaction with physical world: outside IP address
- Fabric/host routing and IP address migration still very relevant

Overlay Virtual Networking and Fabric/Host Routing



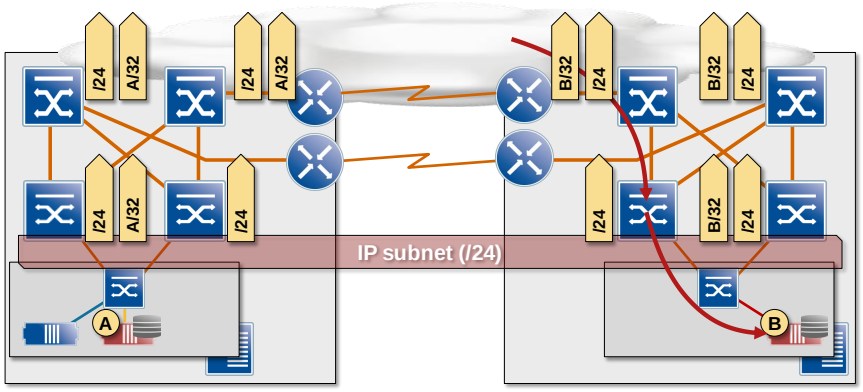
- Overlay networks simplify physical network design, deployment and operations (and increase the complexity of the hypervisor software)
- Interaction with physical world: outside IP address
- Fabric/host routing and IP address migration still very relevant

Still a few years before overlay networks become mainstream technology

Summary

Summary

Technologies



- **Fabric routing**: optimal VM-to-fabric routing
- **Host routing**: optimal fabric-to-VM routing (requires secure edge)
- **Virtual Private Port Services (with Layer2/MAC-over-GRE**: L2 DCI over any transport
- **Virtual Private Ethernet Services (with SPB-over-GRE** autumn): unified L2 ECMP fabric

VM mobility with layer-3 data center interconnects

- Split subnet + fabric/host routing + proxy ARP: available today
- Foreign subnet with local proxy arp: autumn 2013
- Robust L3-only implementation without bridging or overlays

No new technologies, works with all hypervisors, available today

Questions?

Questions?

Send them to ip@ipSpace.net or @ioshints