IPv6 Microsegmentation

Ivan Pepelnjak (ip@ipSpace.net)
Network Architect

ipSpace.net AG

Who is Ivan Pepelnjak (@ioshints)

Past

- Kernel programmer, network OS and web developer
- Sysadmin, database admin, network engineer, CCIE
- Trainer, course developer, curriculum architect
- Team lead, CTO, business owner

Present

- Network architect, consultant, blogger, webinar and book author
- Teaching the art of Scalable Web Application Design

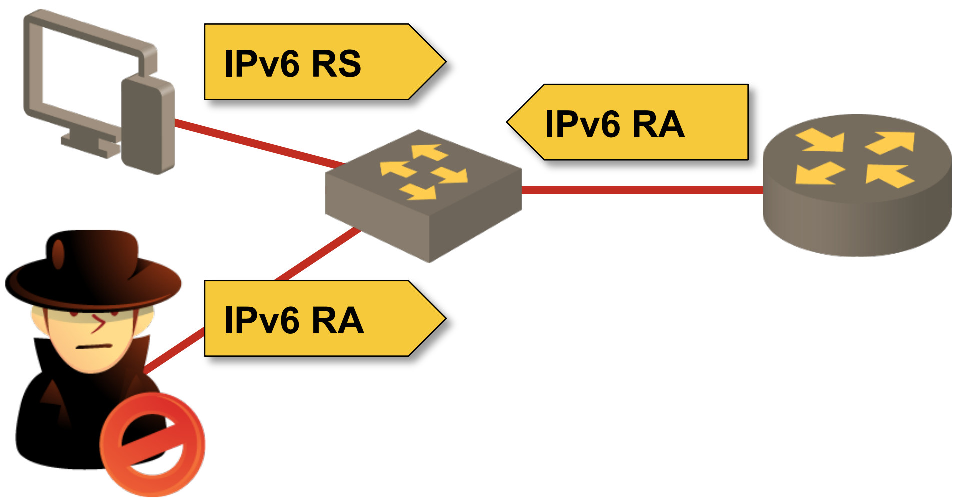
Focus

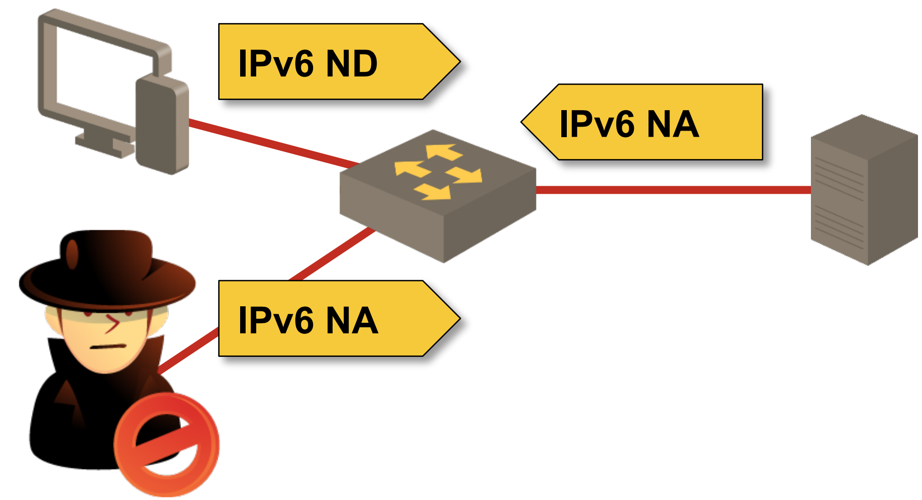
- Large-scale data centers, clouds and network virtualization
- Scalable application design
- Core IP routing/MPLS, IPv6, VPN

More @ ipSpace.net/About

IPv6 Layer-2 Security Challenges

The Problem



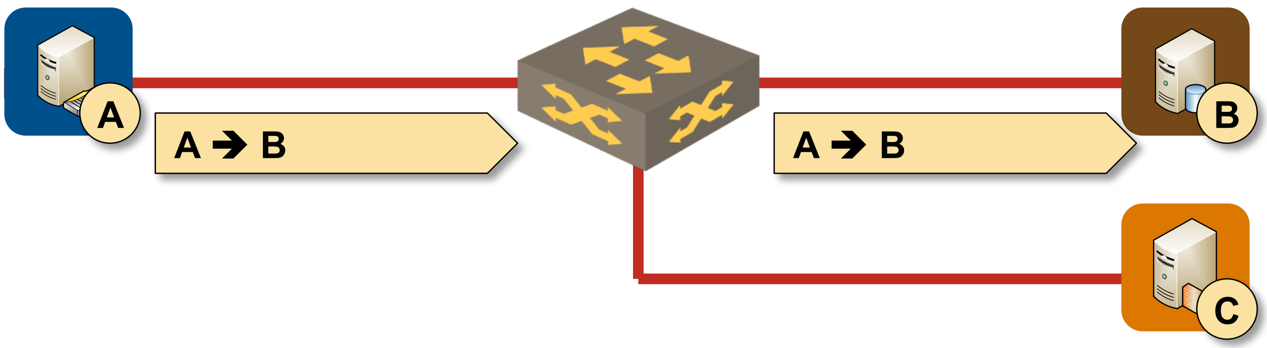


- **Assumption:** one subnet = one security zone
- **Corollary**: intra-subnet communication is not secured
- **Consequences**: multiple first-hop vulnerabilities

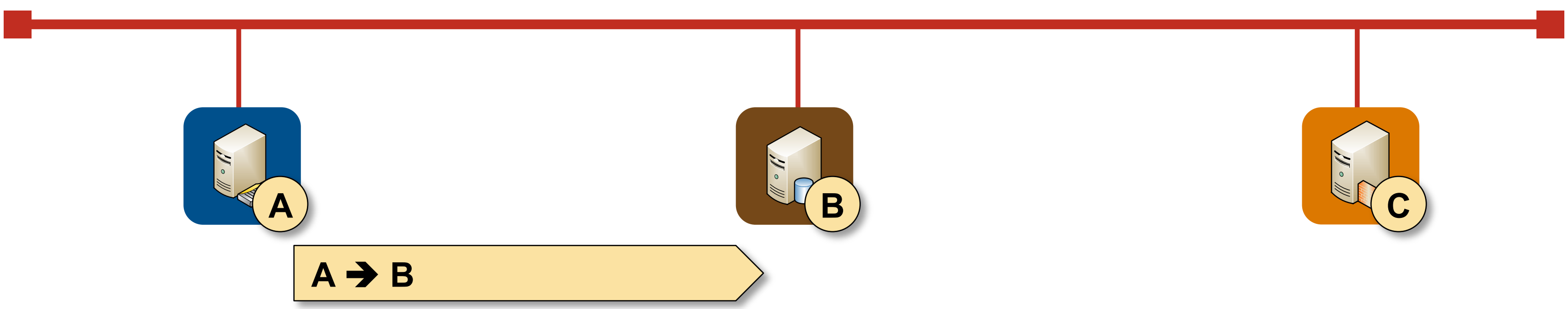
Sample vulnerabilities:

- RA spoofing
- NA spoofing
- DHCPv6 spoofing
- DAD DoS attack
- ND DoS attack

Root Cause

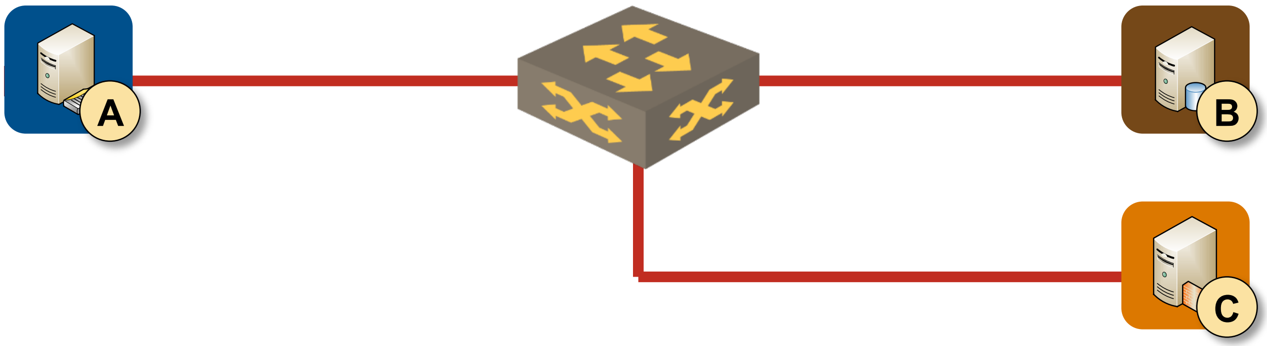


≈



All LAN infrastructure we use today emulates 40 year old thick coax cable

The Traditional Fix: Add More Kludges



Typical networking industry solution

- Retain existing forwarding paradigm
- Implement layer-2 security mechanisms

Sample L2 security mechanisms

- RA guard
- DHCPv6 guard
- IPv6 ND inspection
- SAVI

Benefits

- Non-disruptive deployment (clusters and Microsoft NLB still works)
- No need to educate customers

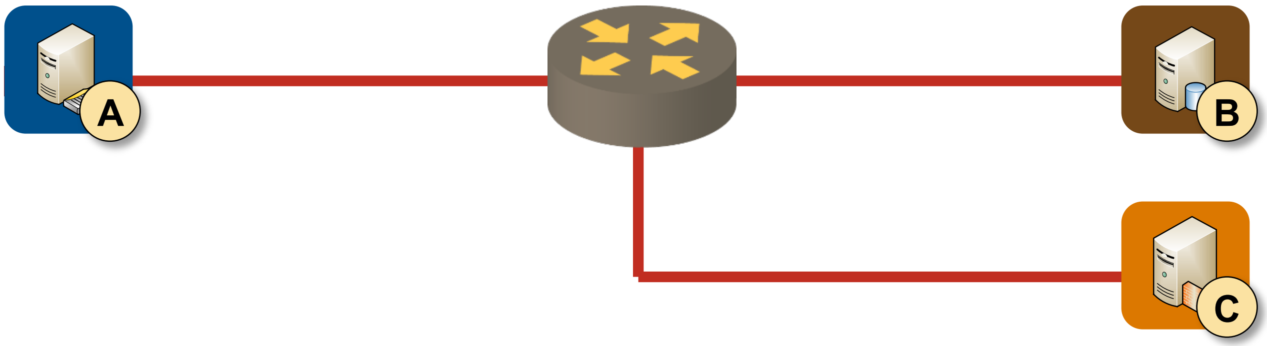
Drawbacks

- Not available on all platforms
- Expensive to implement in hardware
- Exploitable by infinite IPv6 header + fragmentation creativity

Can we do any better than that?

Layer-3-Only IPv6 Networks

Goal: Remove Layer-2 from the Network



Change the forwarding paradigm

- First-hop network device is a router (layer-3 switch in marketese)
- Fake router advertisements or ND/NA messages are not propagated to other hosts

Simplistic implementation

- Every host is in a dedicated /64 subnet
- Results in IPv6 routing table explosion (most data center switches have very limited IPv6 forwarding tables)
- Exceedingly complex in virtualized environments

Can we do any better than that?

Arista Spline Switches

Switch model	Ports	MAC	IPv4	ARP	IPMC	IPv6
7304	128 x 40GbE 512 x 10GbE 192 x 10GBASE-T					
7308	256 x 40GbE 1024 x 10GbE 384 x 10GBASE-T	288K	16K	208K	104K	8K
7316	512 x 40GbE 2048 x 10GbE 768 x 10GBASE-T					

Brocade VDX ToR Switches

Port density

Switch model	GE ports	10GE ports	40GE ports	FC ports
VDX 6710	48	6	-	-
VDX 6720-24	24		-	-
VDX 6720-60	60		-	-
VDX 6730-32	24		-	8
VDX 6730-76	60		-	16
VDX 6740 New	48		4	

Table sizes

Switch	MAC	IPv4	ARP	IPv6
VDX 6740	160K	12K	32K	3K
VDX 67xx	32K	2K	12K	-

Nexus 6000 and 9300 Series Overview

Port density

Switch	1G	10GE	40GE
9396PX New	48 (SFP+)	48	12
9396TX New	48 (10GBASE-T)	48	12
9336PQ New			36
93128PX New	96 (10GBASE-T)	96	8
Nexus 6001 (48 x SFP+, 4 x QSFP)	48	64	4
Nexus 6004 (96 x QSFP)		384	96

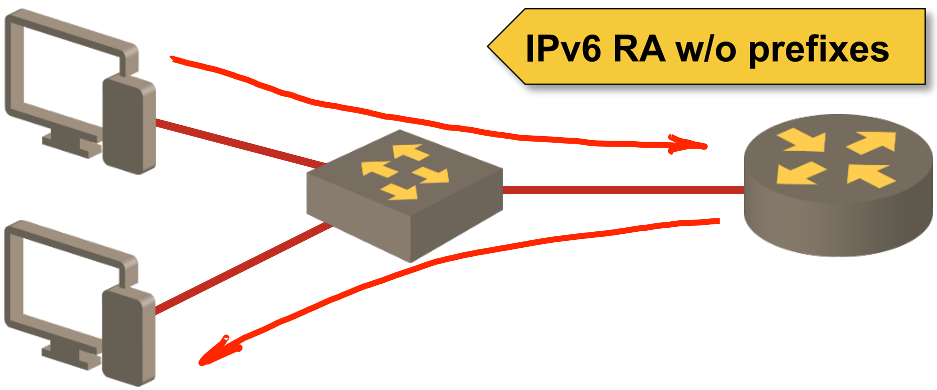
Table sizes

Switch	MAC	IPv4	ARP	IPv6	ND
Nexus 9300	96K	16K	88K	6K	20K
Nexus 6000	115K	24K	64K	8K	32K

Fixed Data Center Switches – EX Series

Model	EX4200	EX4300 New	EX4500	EX4550
Typical role	ToR	ToR	Tor/Core	ToR/Core
Max ports	48 x 1GE 2 x 10GE	24 / 48 GE 4 / 8 10GE	40 – 48 x 10GE	32 – 48 x 10GE 2 x 40GE
MAC table	32K	64K	32K	32K
IPv4 table	16K	4K	10K	10K
ARP	16K	64K	8K	8K
IPMC	8K	8K	4K	4K
IPv6 table	4K	1K	1K	1K
IPv6 ND	16K (shared)	32K	1K	1K

Tweaking On-net Determination



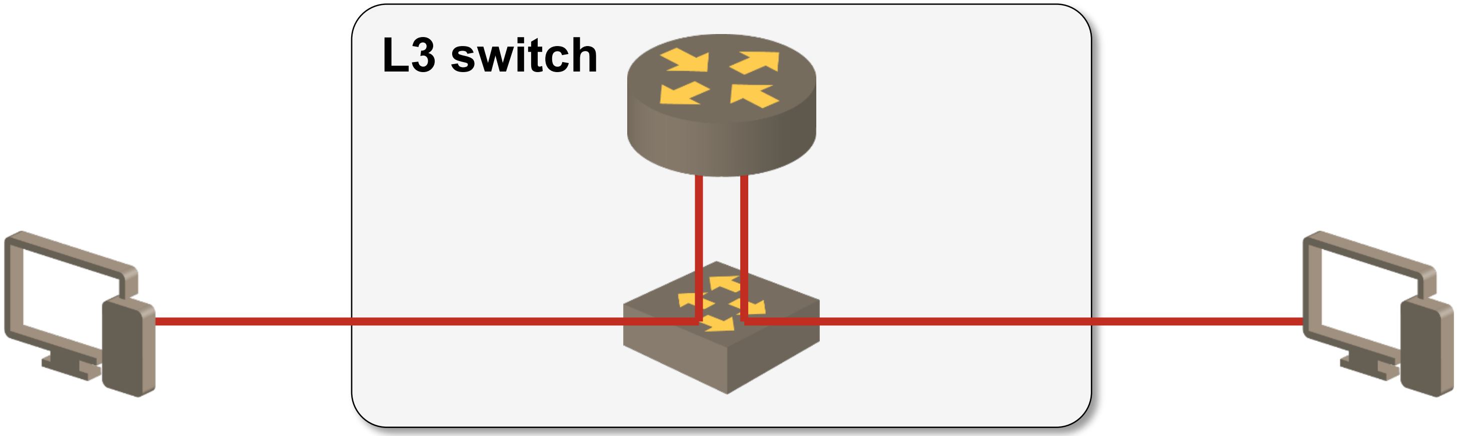
Local subnet is not advertised in RA messages

- IPv6 hosts cannot perform on-net check
- All intra-subnet traffic goes through the first-hop router
- Access lists on first-hop router enforce segmentation

Drawbacks

- Relies on proper IPv6 host behavior
- RA and ND attacks are still possible without IPv6 first-hop security

Tweaking On-net Determination + PVLAN



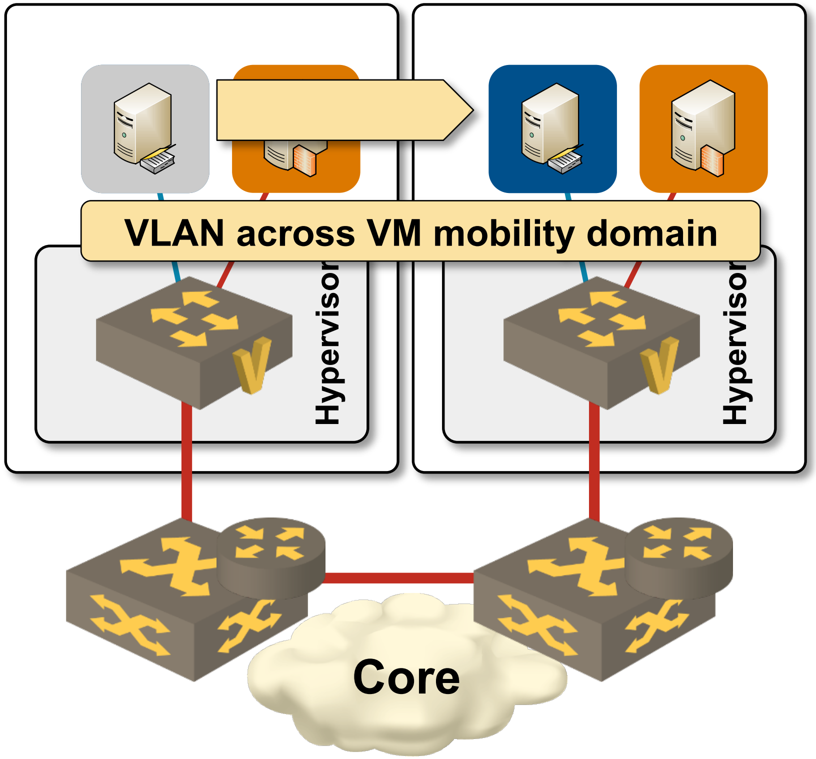
Private VLANs can be used to enforce L3 lookup

- Force traffic to go through L3 device
- Potential solution for campus environments with low-cost L2-only switches or virtualized environments
- L3 device **must not** perform mixed L2/L3 forwarding (hard to implement on a L2/L3 switch)

Implications of Live VM Mobility

Challenges

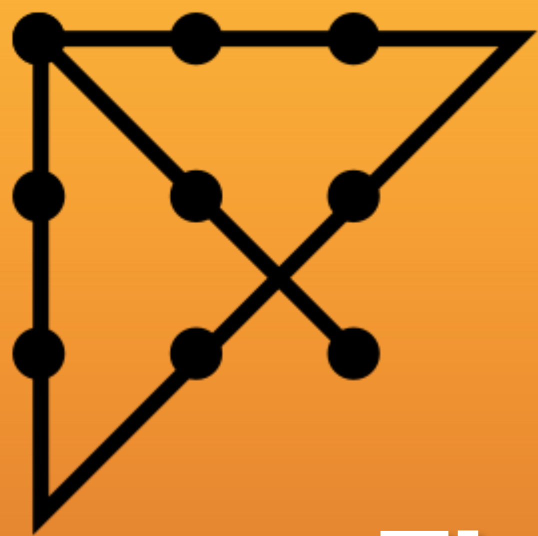
- VM moved to another server must retain its IPv6 address and all data sessions
- Existing L3 solutions are too slow for non-disruptive VM moves
- Live VM mobility usually relies on L2 connectivity between physical servers



Integration with IPv6 Microsegmentation

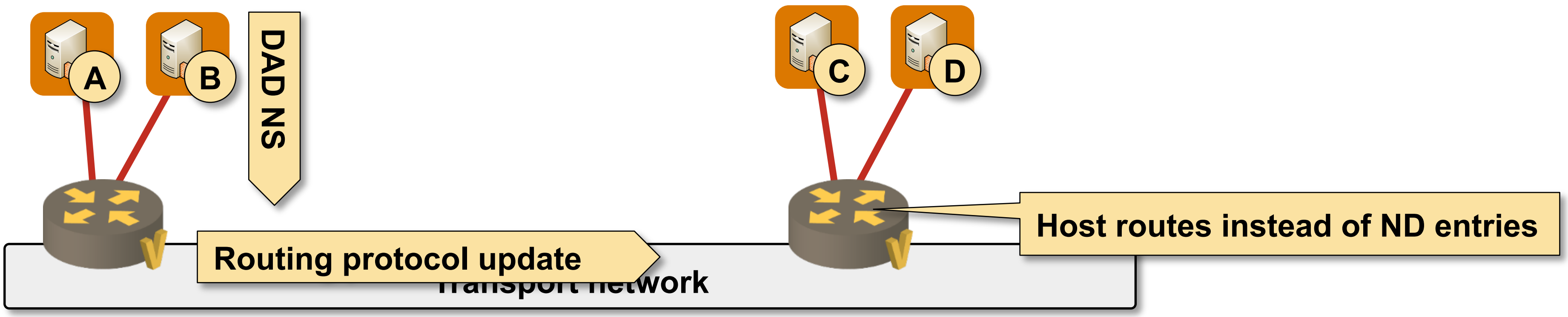
- PVLAN or VLAN-per-VM
- L3 lookup on core switches or anycast first-hop gateway
- East-west traffic always traverses network core

We still need something better



Thinking Outside of the Box

Intra-Subnet (Host Route) Layer-3 Forwarding

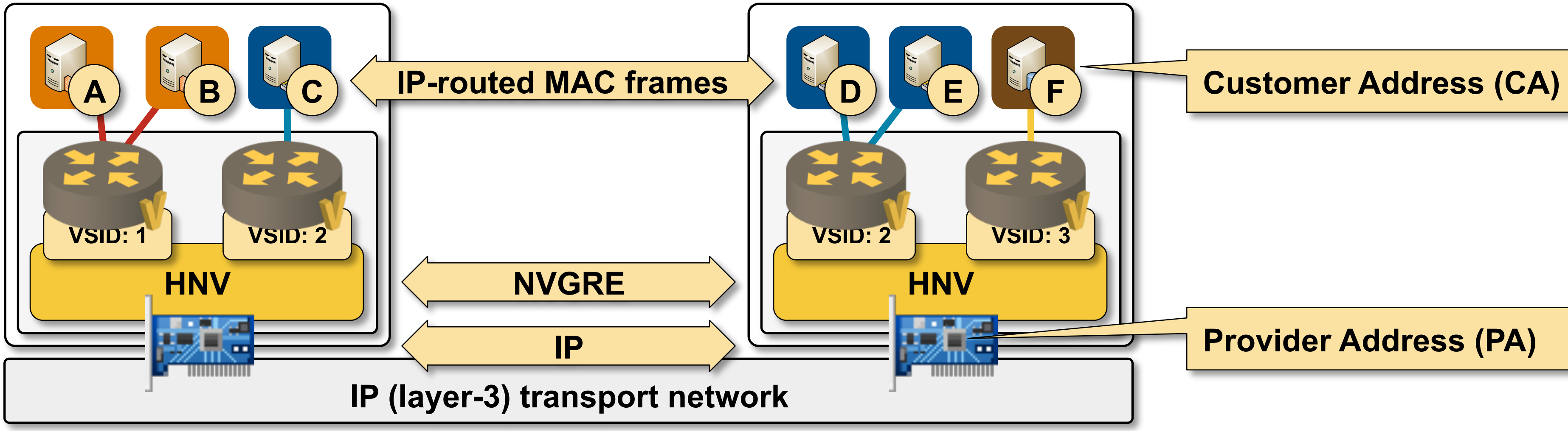


- Hosts are connected to layer-3 switches (routers)
- Numerous hosts share a /64 subnet
  ➔ a /64 subnet spans multiple routers
- First-hop router creates a host route on DAD or DHCPv6 transaction
- IPv6 host routes are propagated throughout the local routing domain
- Host-side IPv6 addressing and subnet semantics are retained
- IPv6 ND entries are used instead of IPv6 routing table entries

Fixed Data Center Switches – EX Series

Model	EX4200	EX4300 New	EX4500	EX4550
Typical role	ToR	ToR	Tor/Core	ToR/Core
Max ports	48 x 1GE 2 x 10GE	24 / 48 GE 4 / 8 10GE	40 – 48 x 10GE	32 – 48 x 10GE 2 x 40GE
MAC table	32K	64K	32K	32K
IPv4 table	16K	4K	10K	10K
ARP	16K	64K	8K	8K
IPMC	8K	8K	4K	4K
IPv6 table	4K	1K	1K	1K
IPv6 ND	16K (shared)	32K	1K	1K

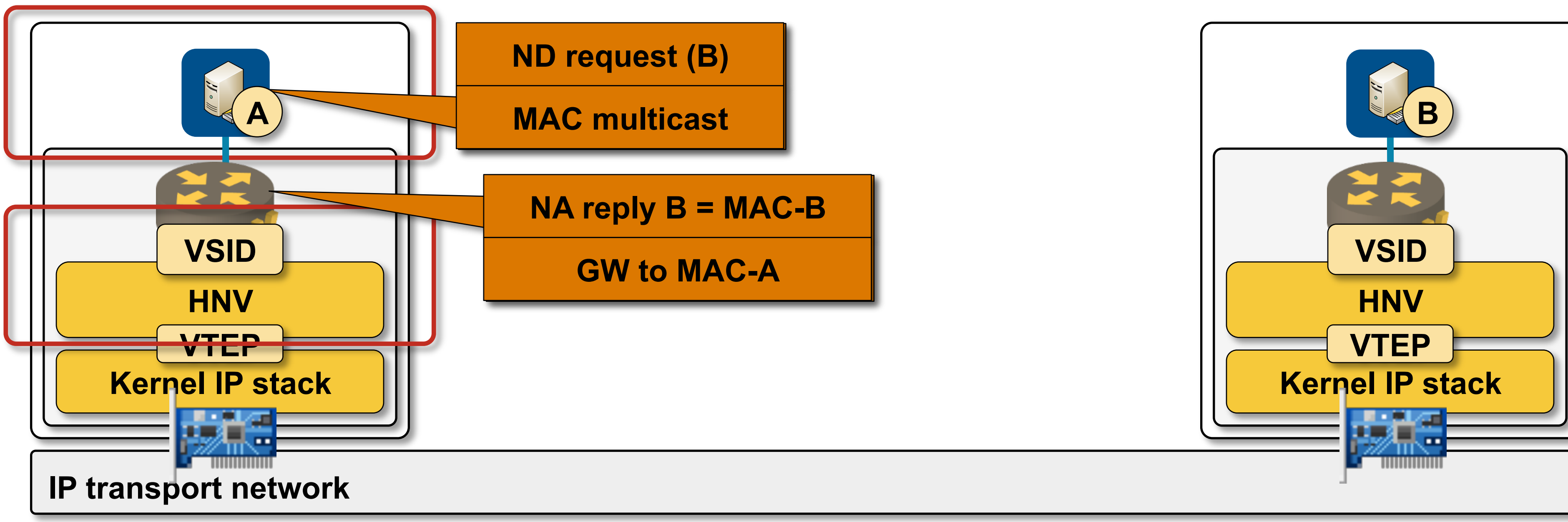
Example: Hyper-V Network Virtualization



Full layer-3 switch in the hypervisor (distributed routing functionality)

- L3-only switching for intra-hypervisor and inter-hypervisor traffic
- IPv4 and IPv6 support in customer (virtual) and provider (transport) network
- ARP and ND proxies ➔ no ARP or unknown unicast flooding
- Source node flooding or Customer ➔ Provider IP multicast mapping

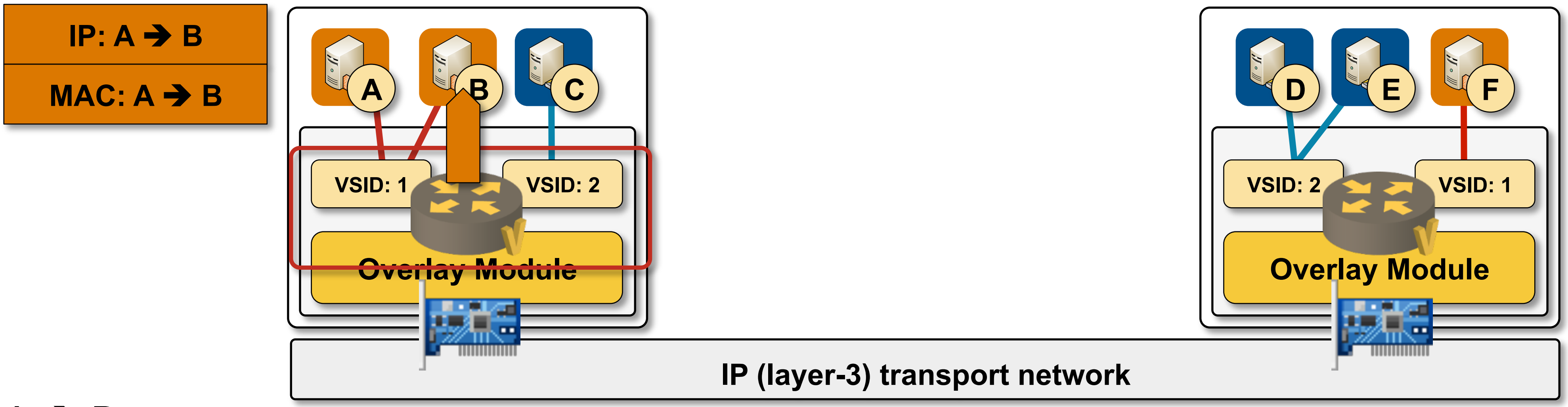
Hyper-V Network Virtualization ND Proxy



- VM generates ND multicast
- L2 broadcast/multicast intercepted by Hyper-V kernel module
- Local Hyper-V replies to ND request with MAC address of remote VM
- Remote hypervisor is not involved
- Unicast ND requests are forwarded to target VM (NUD probes)

Other implementations might use GW MAC address in NA replies

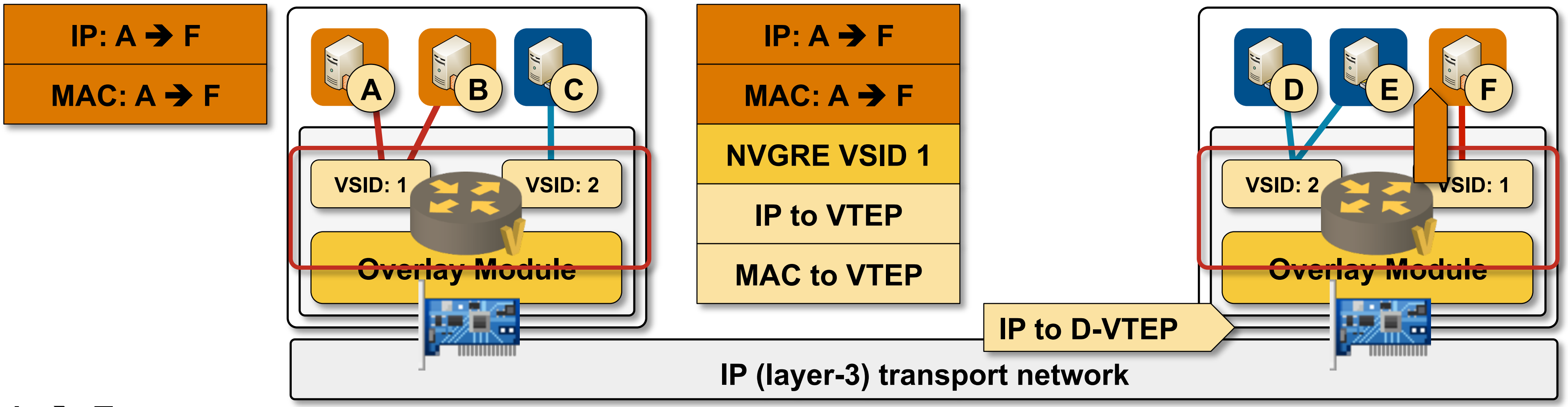
HNV Local Switching



A ➔ B

- On-link, sent directly to MAC-B
- L3 switched within the hypervisor (based on destination IPv6 address)
- IPv4, IPv6 and ARP packets are forwarded, all other traffic is dropped
- Ethernet frame delivered to target VM

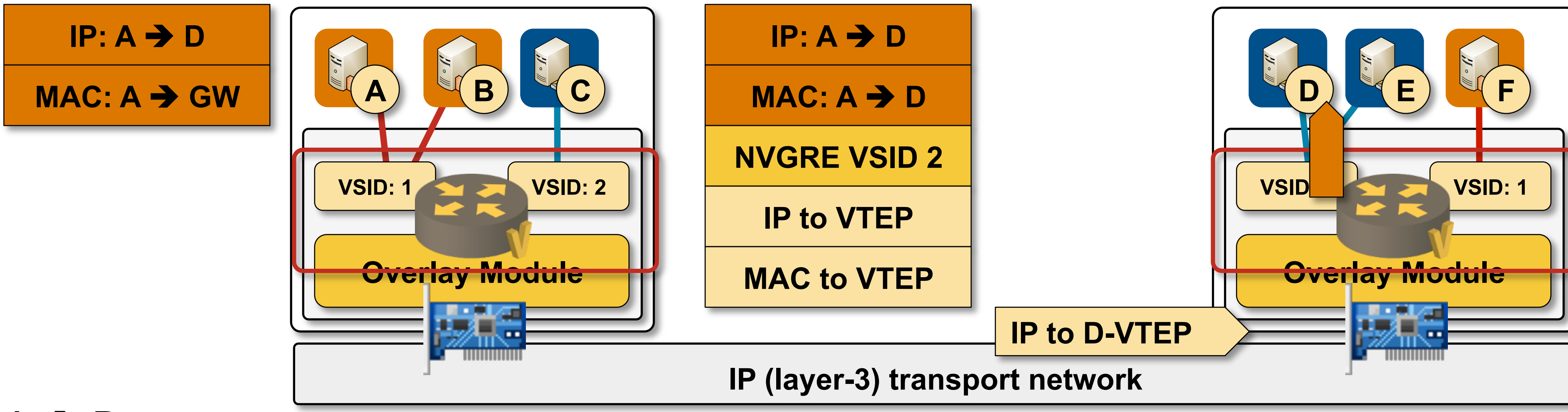
HNV Remote Switching within a Subnet



A ➔ F

- On-link, sent directly to MAC-F
- L3 switched within the hypervisor (based on destination IPv6 address)
- Destination VTEP is remote ➔ build NVGRE envelope and send packet
- Packet received by remote hypervisor
- L3 switching within the routing domain (based on NVGRE VSID)
- Ethernet frame delivered to target VM

HNV Remote Switching across Subnets



A → D

- Off-link, sent to GW MAC address
- L3 switched within the hypervisor (based on destination IPv6 address)
- Switching across subnets → MAC rewrite
- Destination VTEP is remote → build NVGRE envelope and send packet
- Packet received by remote hypervisor
- L3 switching within the routing domain (based on NVGRE VSID)
- Ethernet frame delivered to target VM

HNV does not rewrite source MAC address or decrement TTL

Implementations of Host Route-Based Forwarding

IPv6 and IPv4

- Hyper-V Network Virtualization
- Juniper Contrail
- Cisco Dynamic Fabric Automation (DFA)

IPv4 only

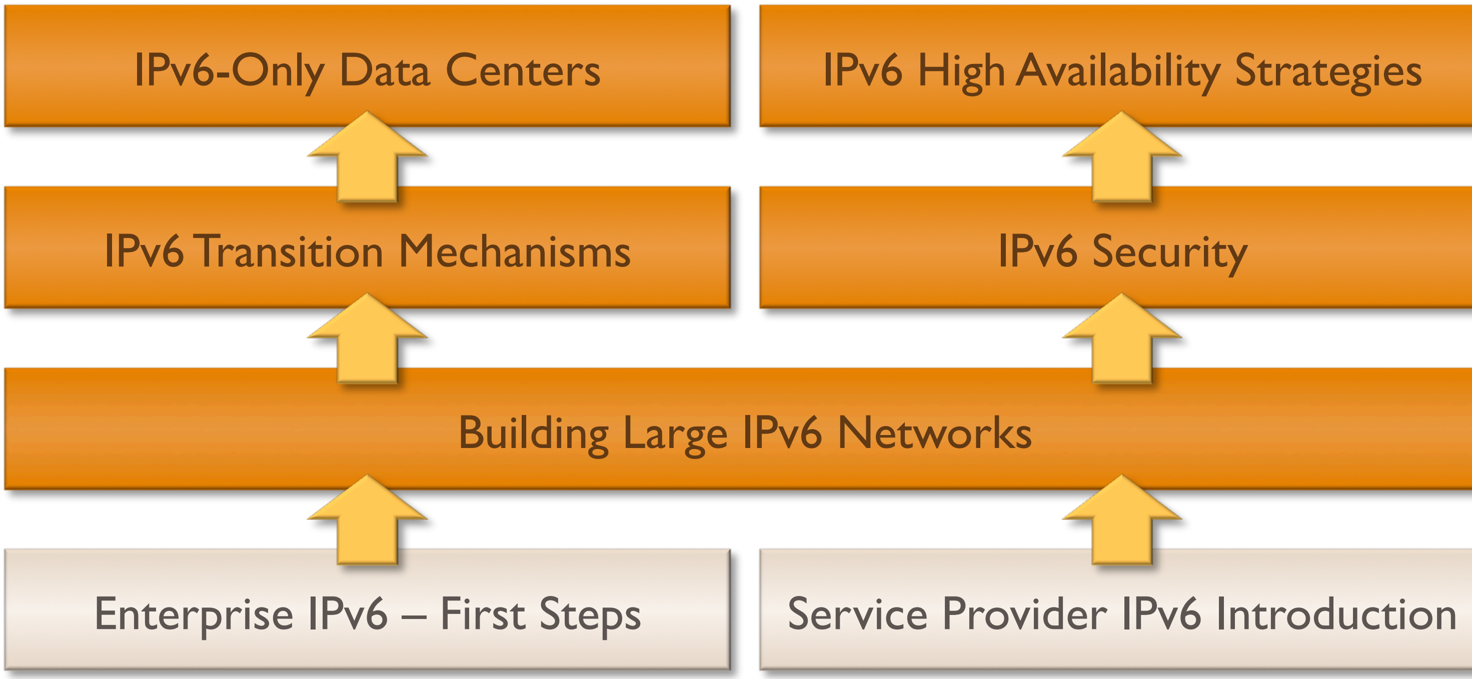
- Nuage Virtual Services Platform (VSP)
- Cisco Application Centric Infrastructure (ACI)

Unrelated honorable mention

- IPv6 RA guard and ND inspection implemented on VMware NSX

Hint: vote with your wallet!

More Information



Availability

- Live sessions
- Recordings of individual webinars
- Yearly subscription

Other options

- Customized webinars
- ExpertExpress
- On-site workshops

More information @ http://www.ipSpace.net/IPv6

Stay in Touch

Web:	ipSpace.net
Blog:	blog.ipSpace.net
Email:	ip@ipSpace.net
Twitter:	@ioshints

SDN:	ipSpace.net/SDN
Webinars:	ipSpace.net/Webinars
Consulting:	ipSpace.net/Consulting

