NO LIMITS

NET2745

vSphere Distributed Switch: Technical Deep Dive

Jason Nash, Varrow
Chris Wahl, AHEAD

vmworld® 2014

vSphere Distributed Switch:

A Technical Deep Dive

NET2745

Chris Wahl

Senior Solutions Architect @ AHEAD

VCDX #104, vExpert

Blogger – WahlNetwork.com

Author – *Networking for VMware Administrators*

Author – Pluralsight IT Pro training

CCNA Data Center, vCloud Director

Host – *VUPaaS* and *IT Engine Builders* Podcasts

Jason Nash

Chief Technology Officer @ Varrow
VCDX #49, vExpert

Blogger – JasonNash.com
Author – Pluralsight IT Pro training
XtremIO, Cisco UCS, Nexus 1000v, vC Ops, + more

Advanced Technical Warning

Session Agenda

- Overview
- Requirements
- Construction
- Alternatives
- Tips
- Real World Use Cases

Are You Using the vSphere Distributed Switch (VDS)?

Requirements

- ✓ Enterprise Plus licensing
- ✓ VMware skillset
- ✓ ESXi host versions ≥ to VDS version

Select version

Specify a distributed switch version.

- (•) Distributed switch: 5.5.0
  This version is compatible with VMware ESXi version 5.5 and later. The following new features are available: Traffic Filtering and Marking, and enhanced LACP support.
- ( ) Distributed switch: 5.1.0
  This version is compatible with VMware ESXi version 5.1 and later. The following new features are available: Management Network Rollback and Recovery, Health Check, Enhanced Port Mirroring, and LACP.
- ( ) Distributed switch: 5.0.0
  This version is compatible with VMware ESXi version 5.0 and later. The following new features are available: User-defined network resource pools in Network I/O Control, NetFlow, and Port Mirroring.
- ( ) Distributed switch: 4.1.0
  This version is compatible with VMware ESX version 4.1 and later. The following new features are available: Load-Based Teaming and Network I/O Control.
- ( ) Distributed switch: 4.0
  This version is compatible with VMware ESX version 4.0 and later. The features supported by later distributed switch versions will not be available.

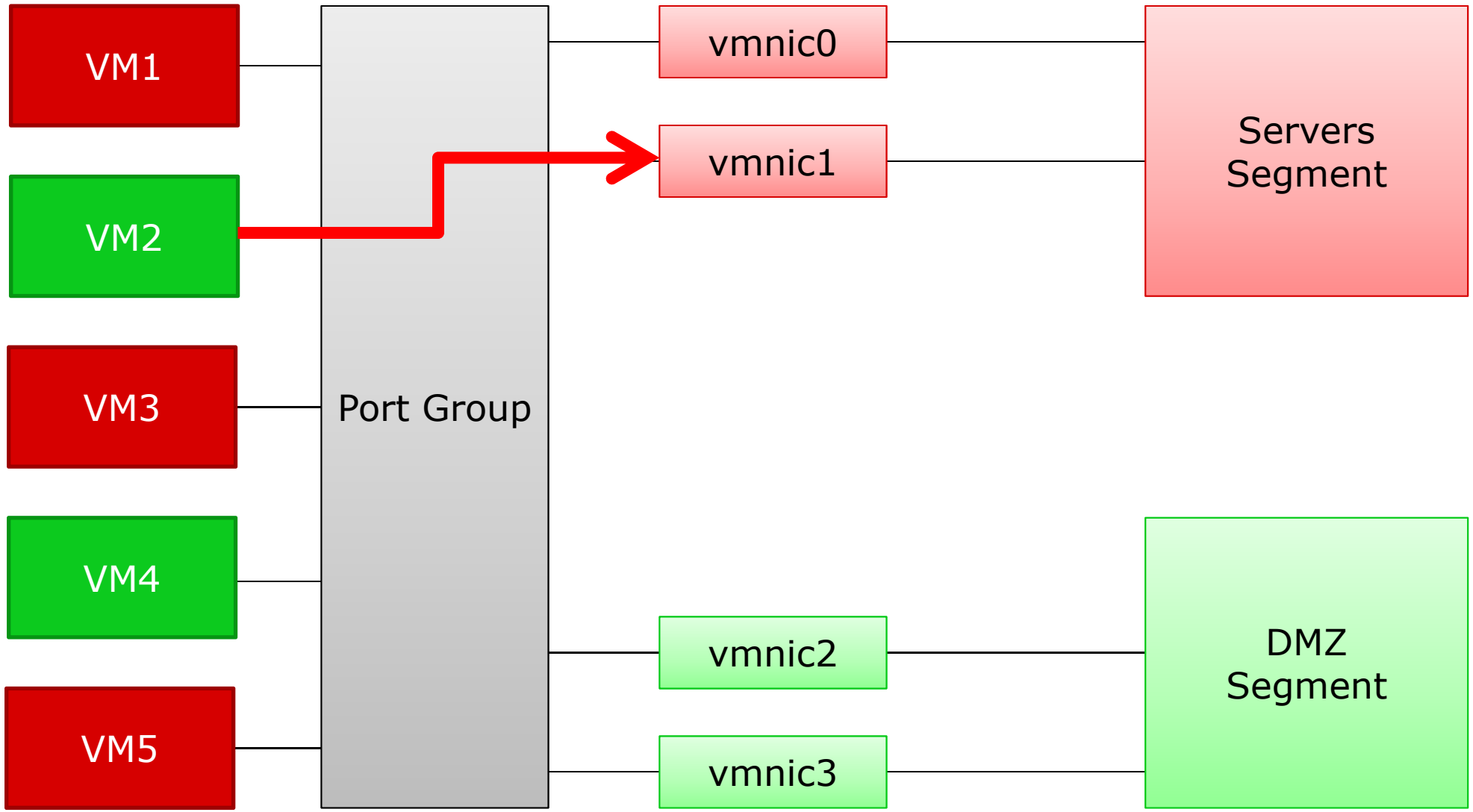
Control Plane

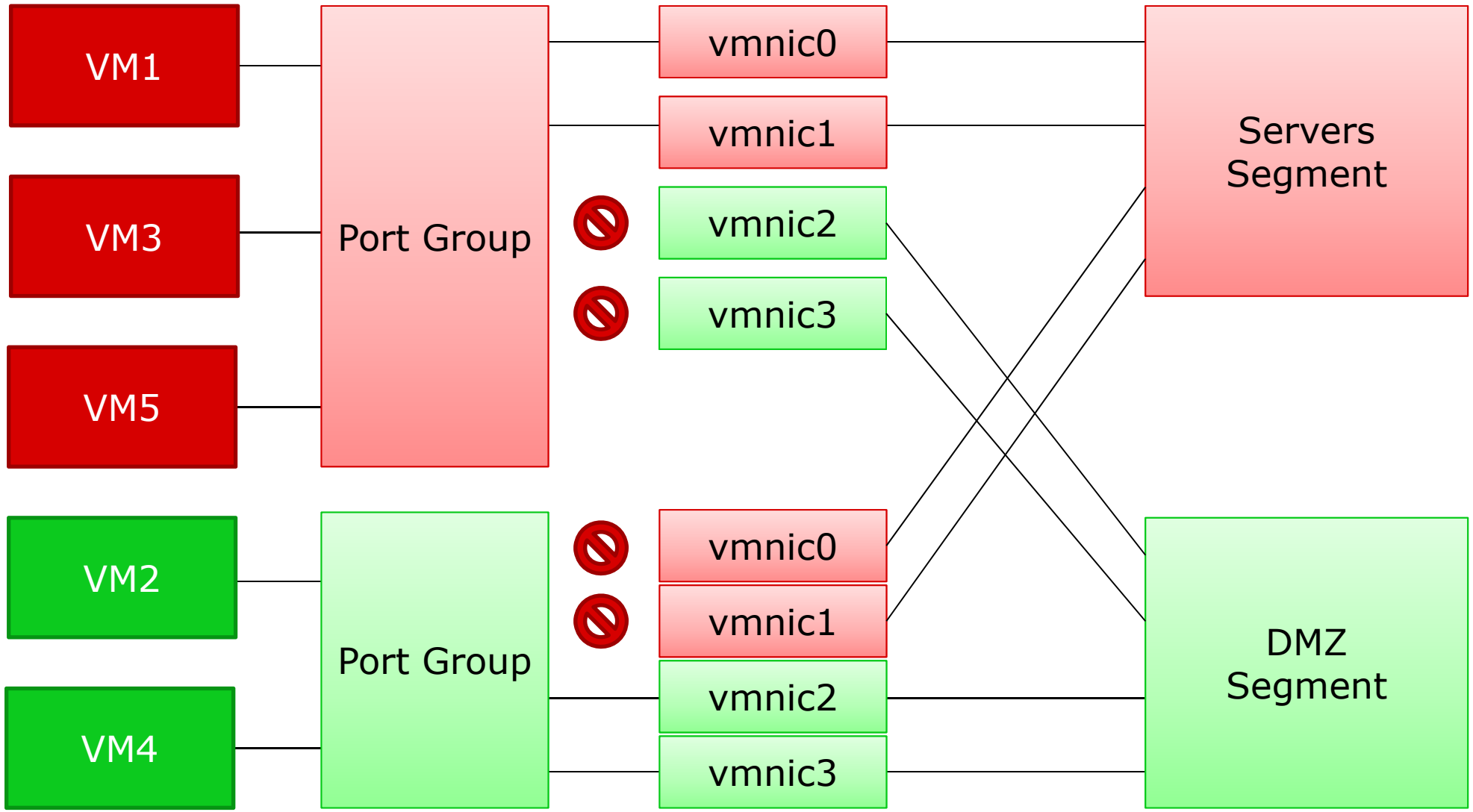
- Lives in vCenter
- All 5.1+ features are Web Client only ☹
- VDS is a Data Center level object
- VDS database syncs with each ESXi host

```
/etc/vmware # ls -l
total 1528
-r--r--r--    1 root     root          3041 Jun  7 19:30 BootbankFunctions.sh
-rw-r--r--    1 root     root          1061 Jul  3 20:59 config
-rw-r--r-T    1 root     root          9872 Jun  7 19:30 configrules
drwxr-xr-x    1 root     root           512 Jul  3 20:57 default.map.d
drwxr-xr-x    1 root     root           512 Jul  3 20:57 driver.map.d
-rw-r--r--    1 root     root         63488 Aug  3 13:38 dvsdata.db
-rw-------    1 root     root         35602 Jul 10 16:49 esx.conf
drwxr-xr-x    1 root     root           512 Jul  3 20:57 firewall
-r--r--r--    1 root     root            59 Jun  7 19:42 ft-vmk-version
-r--r--r--    1 root     root            60 Jun  7 19:42 ft-vmx-version
drwxr-xr-x    1 root     root           512 Jul 29 15:31 hostd
drwxr-xr-x    1 root     root           512 Jul  3 20:57 icu
drwxr-xr-x    1 root     root           512 Jul  3 20:57 ike
-rw-r--r-T    1 root     root            50 Jun  7 19:30 ima_plugin.conf
-rw-r--r--    1 root     root           310 Aug  3 13:35 license.cfg
-rw-r--r-T    1 root     root           440 Jun  7 19:30 localsas
-rw-r--r-T    1 root     root             0 Jun  7 19:30 lockdown
-rw-r--r-T    1 root     root            53 Jun  7 19:42 lockdown.conf
-rw-r--r--    1 root     root             0 Jul  3 20:58 locker.conf
drwxr-xr-x    1 root     root           512 Jul  3 20:57 microcode
drwxr-xr-x    1 root     root           512 Jul  3 21:03 netcpa
-rw-r--r-T    1 root     root           817 Jun  7 19:30 passthru.map
-r--r--r--    1 root     root        815377 Dec 11  2013 pci.ids
drwxr-xr-x    1 root     root           512 Jun  7 19:30 pciid
drwxr-xr-x    1 root     root           512 Jul  3 20:59 rhttpproxy
```

Data Plane

- Lives on the ESXi host
- We suggest
  - Use **Elastic ports**
  - Connect uplinks (vmnics) to a **single network segment**
- If you need multiple network segments
  - Possible, but requires workarounds


VM1
VM2
VM3
VM4
VM5
Port Group
vmnic0
vmnic1
vmnic2
vmnic3
Servers Segment
DMZ Segment


VM1
VM3
VM5
VM2
VM4
Port Group
Port Group
vmnic0
vmnic1
vmnic2
vmnic3
vmnic0
vmnic1
vmnic2
vmnic3
Servers Segment
DMZ Segment

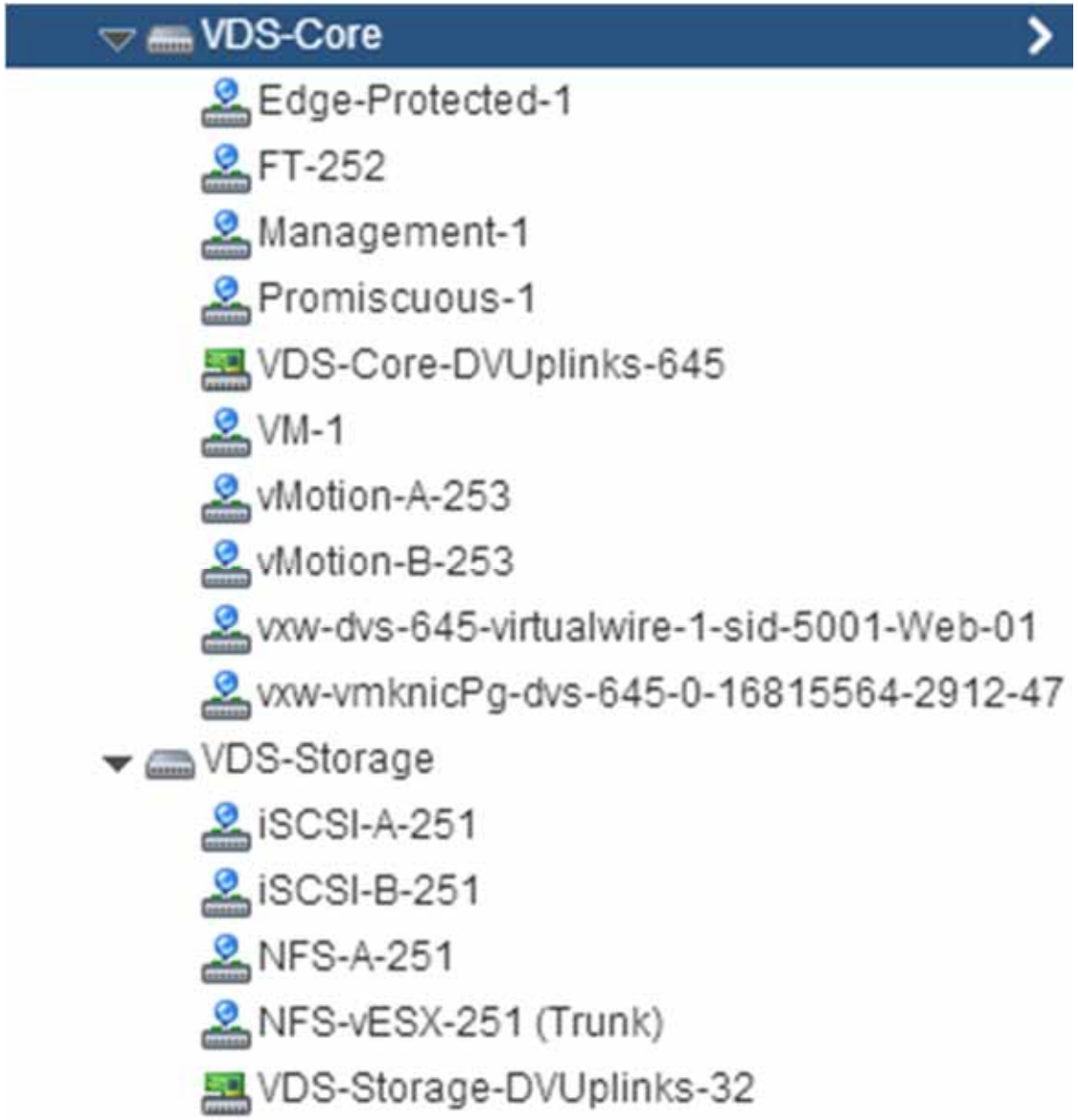
VDS Alternatives

- vSphere Standard Switch (VSS)
- Cisco Nexus 1000v
- IBM 5000V aka “Chupacabra”

VDS Quick Tips

- ✓ Use **802.1Q tags** for port groups
- ✓ **At least 2 vmnics** (uplinks) per VDS
- ✓ A 2x 10 GbE configuration can work fine
- ✓ Put **QoS tagging** in VDS or physical, not both
- ✓ Use **descriptive naming** everywhere
  - o No one knows what “dvPortGroup-1” does

Example of Descriptive Naming



Real World Use Cases – Deep Dive

Migration VSS to VDS

Mixing 1Gb and 10Gb Hosts

Handling vMotion Saturation

vSphere Replication Bandwidth

Quality of Service Tagging

Load Based Teaming vs Link Aggregation

Use Case – Migration from VSS to VDS

Triggers:

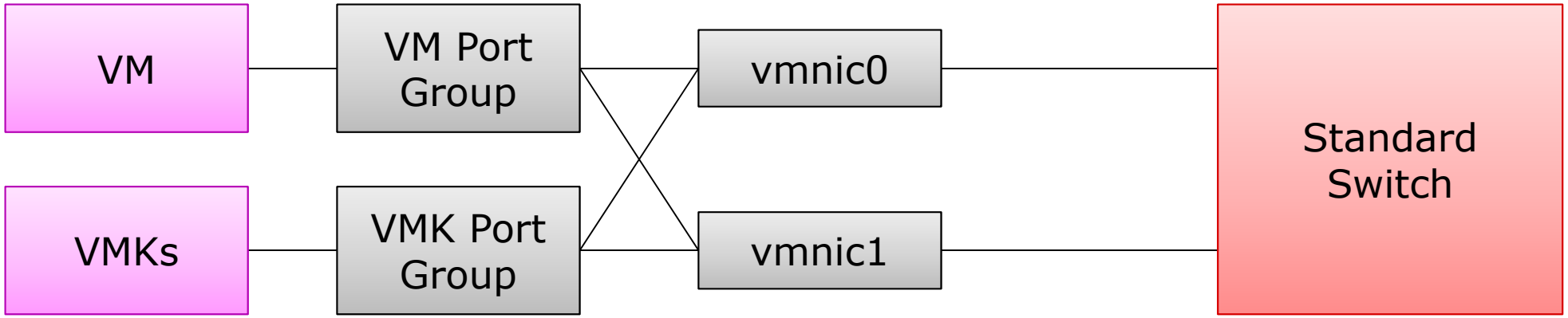
- Licensing (purchased Enterprise Plus)
- Consume features found only in VDS
- Reduce operational overhead
- Separate control planes and related responsibilities

Use Case – Migration from VSS to VDS

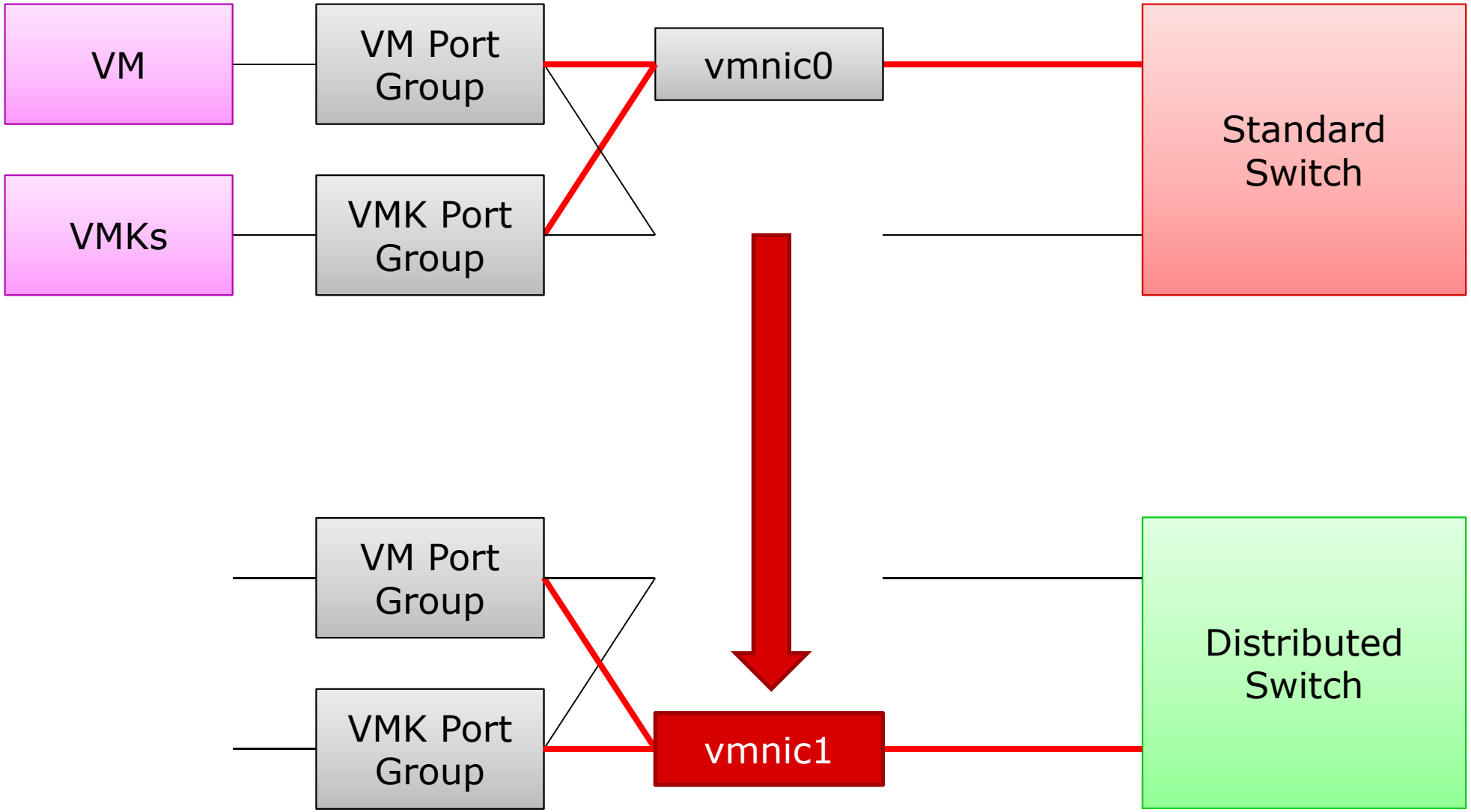
Tips and Advice:

- Have a **detailed plan** in place
- **Test the process** on a single host with non-prod VMs
- **Test network convergence** time and ping drops
- Become **comfortable** with the steps
- Put in a **change control**
- Execute change during **maintenance window**

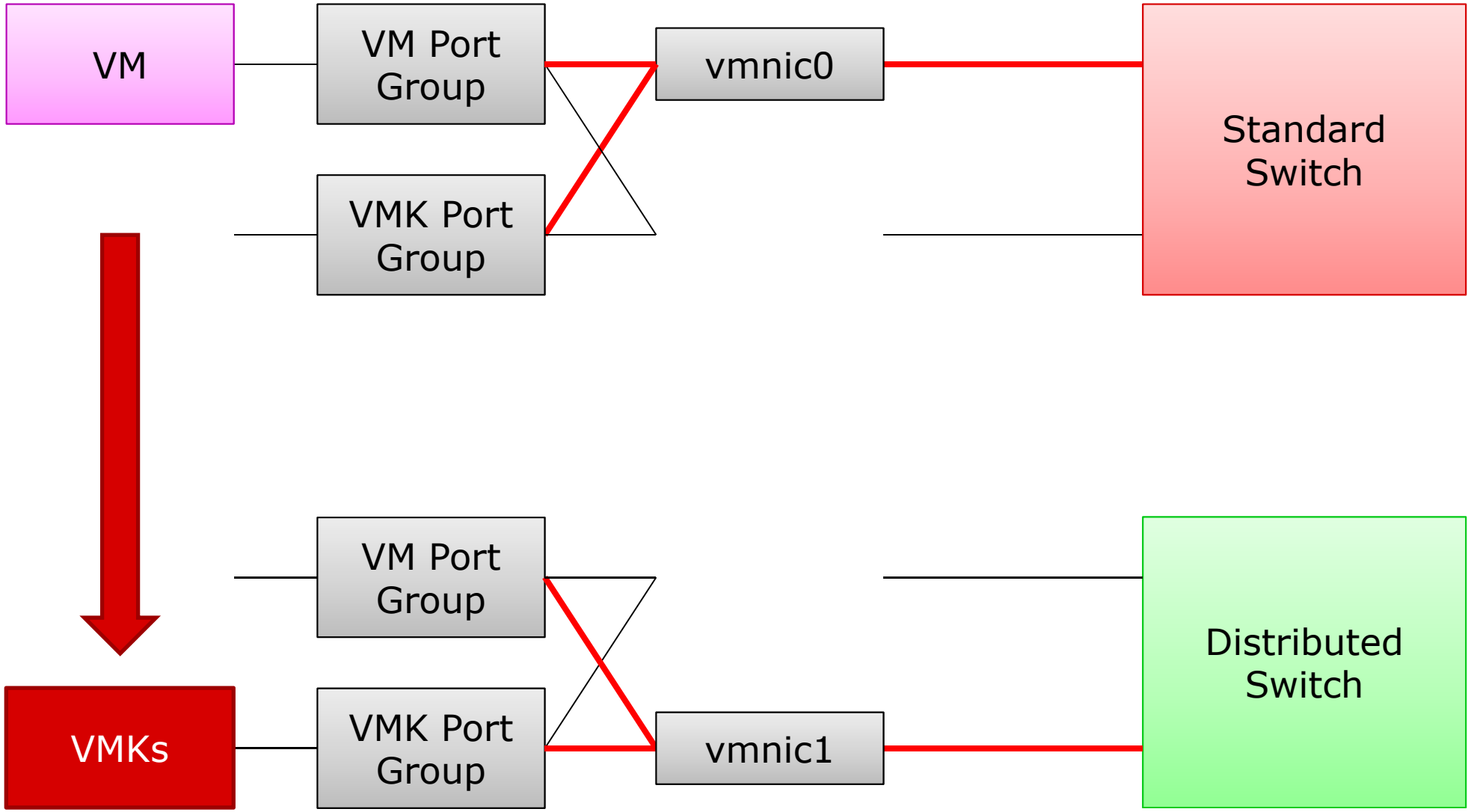
Step 1 – Create the VDS Structure



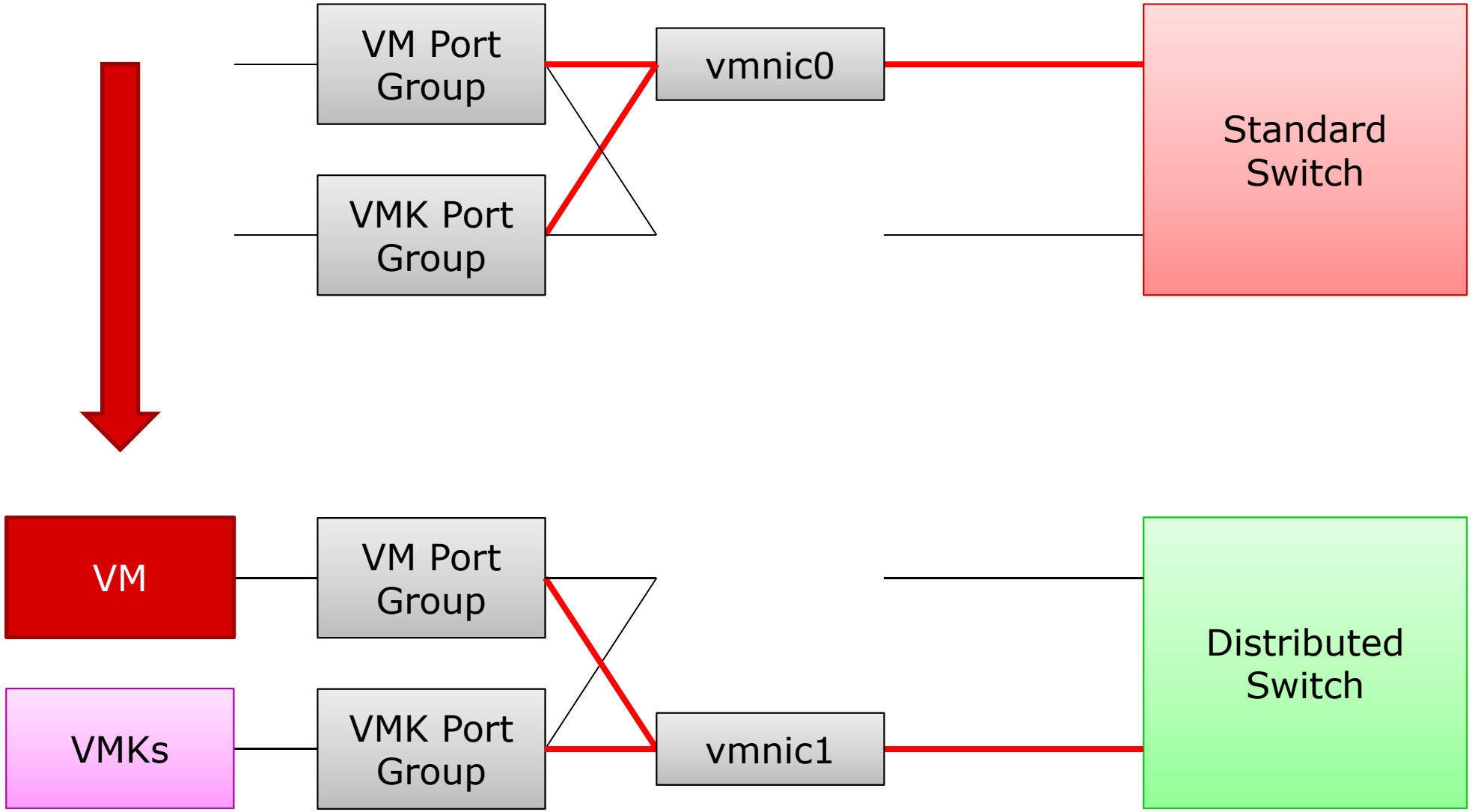
Step 2 – Migrate one vmnic to the VDS



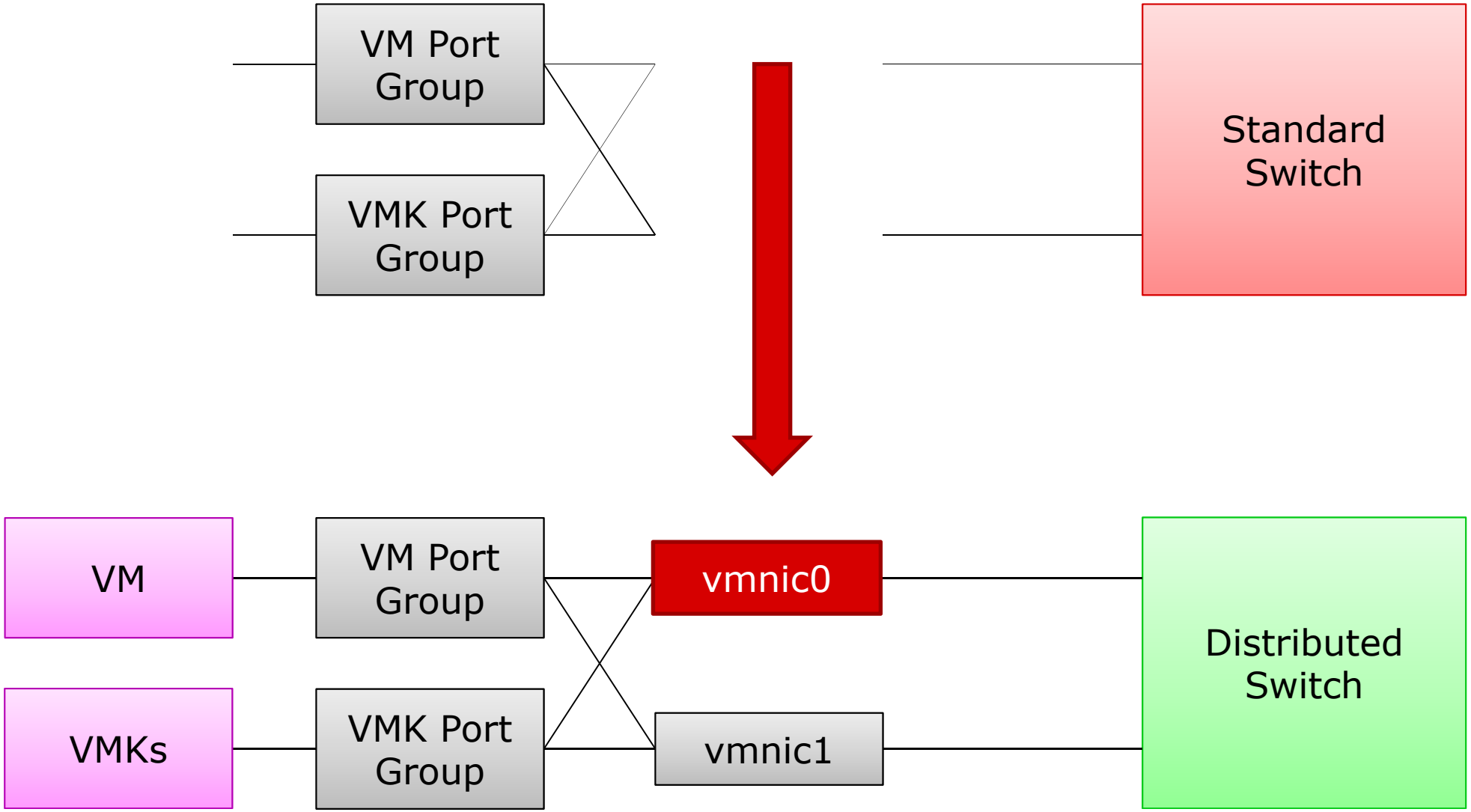
Step 3 – Migrate the VMK interface(s)



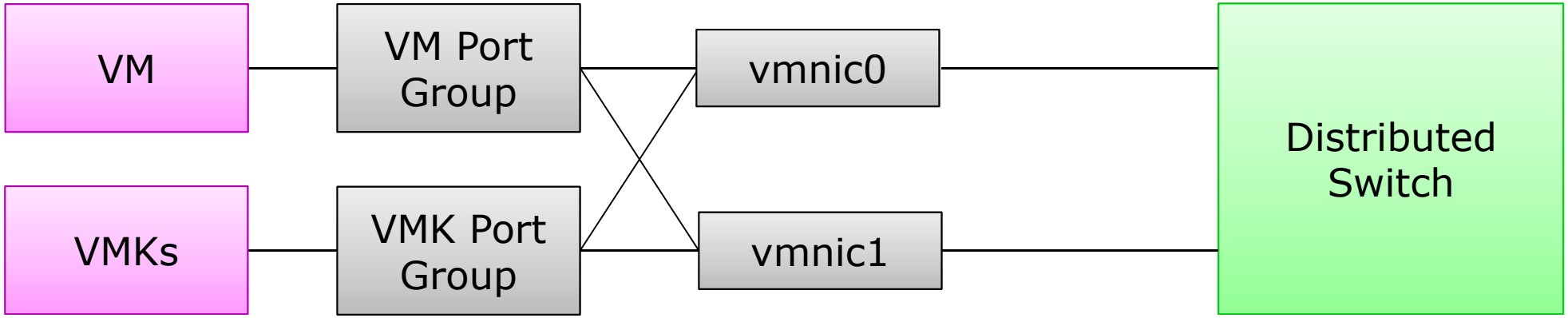
Step 4 – Migrate the VMs



Step 5 – Migrate the remaining vmnic



Step 6 – Destroy the Standard Switch



Use Case – Mixing 1 GbE and 10 GbE Hosts

Triggers:

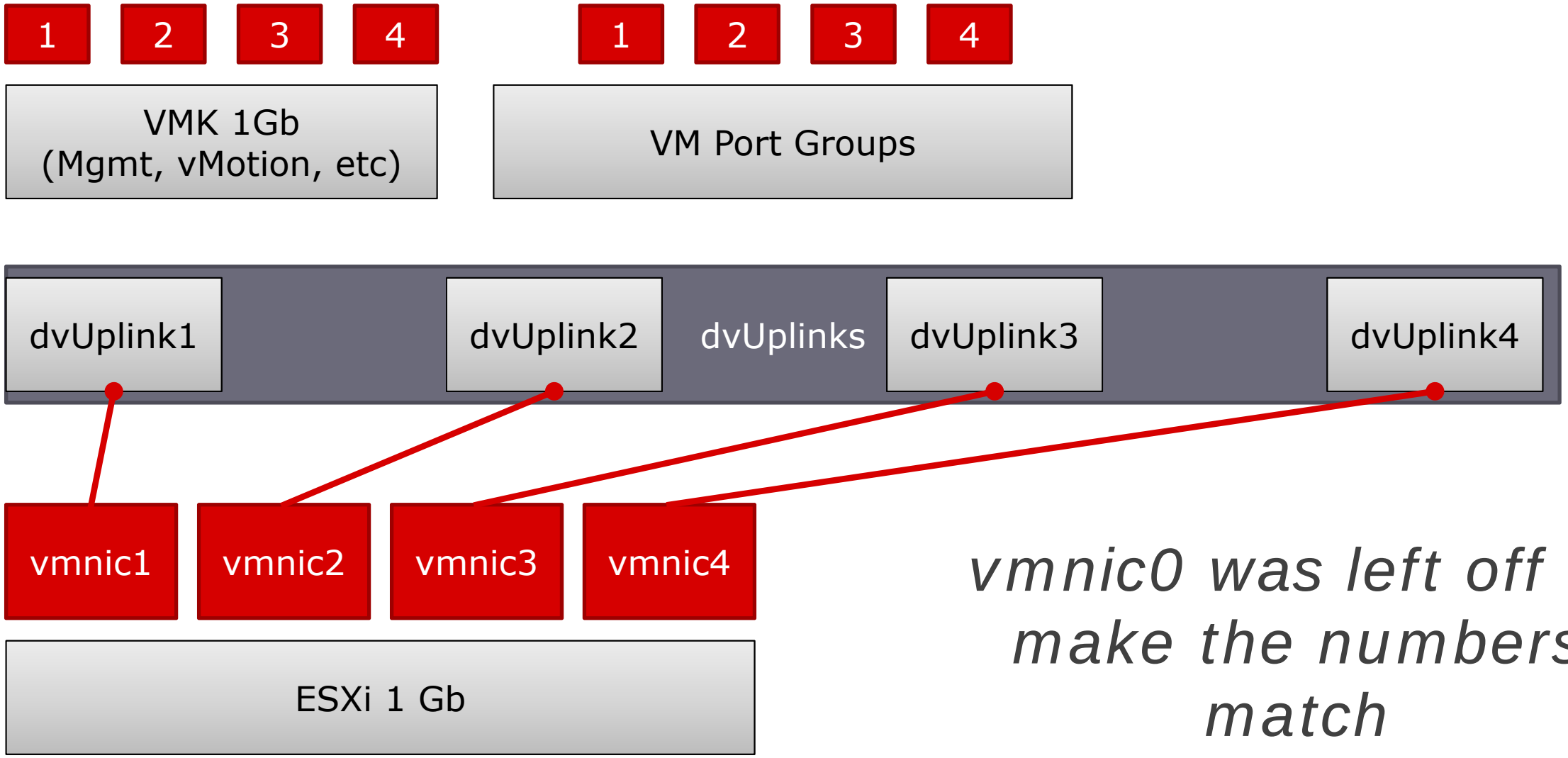
- Purchase of new server / switch hardware
- Staged migration to 10 GbE
- Data Center transformation process

Use Case – Mixing 1 GbE and 10 GbE Hosts

Tips and Advice:

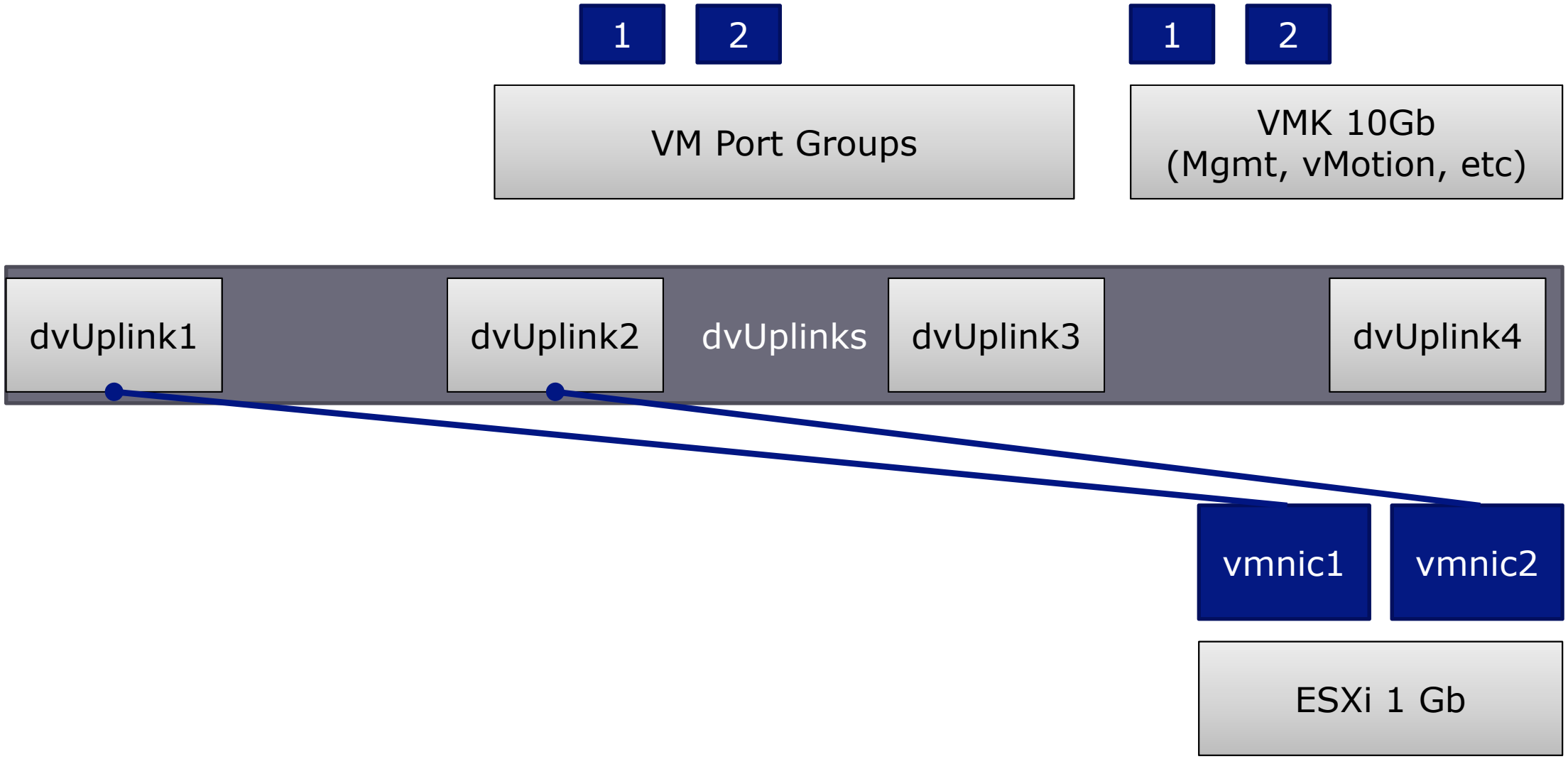
- Use a **single network segment**
- Use a **single VDS**
- Hosts should be **entirely** 1 GbE or 10 GbE
- VM Traffic can traverse **any uplink**
- Control teaming policies on **VMK networks**

VDS with 1 Gb NICs

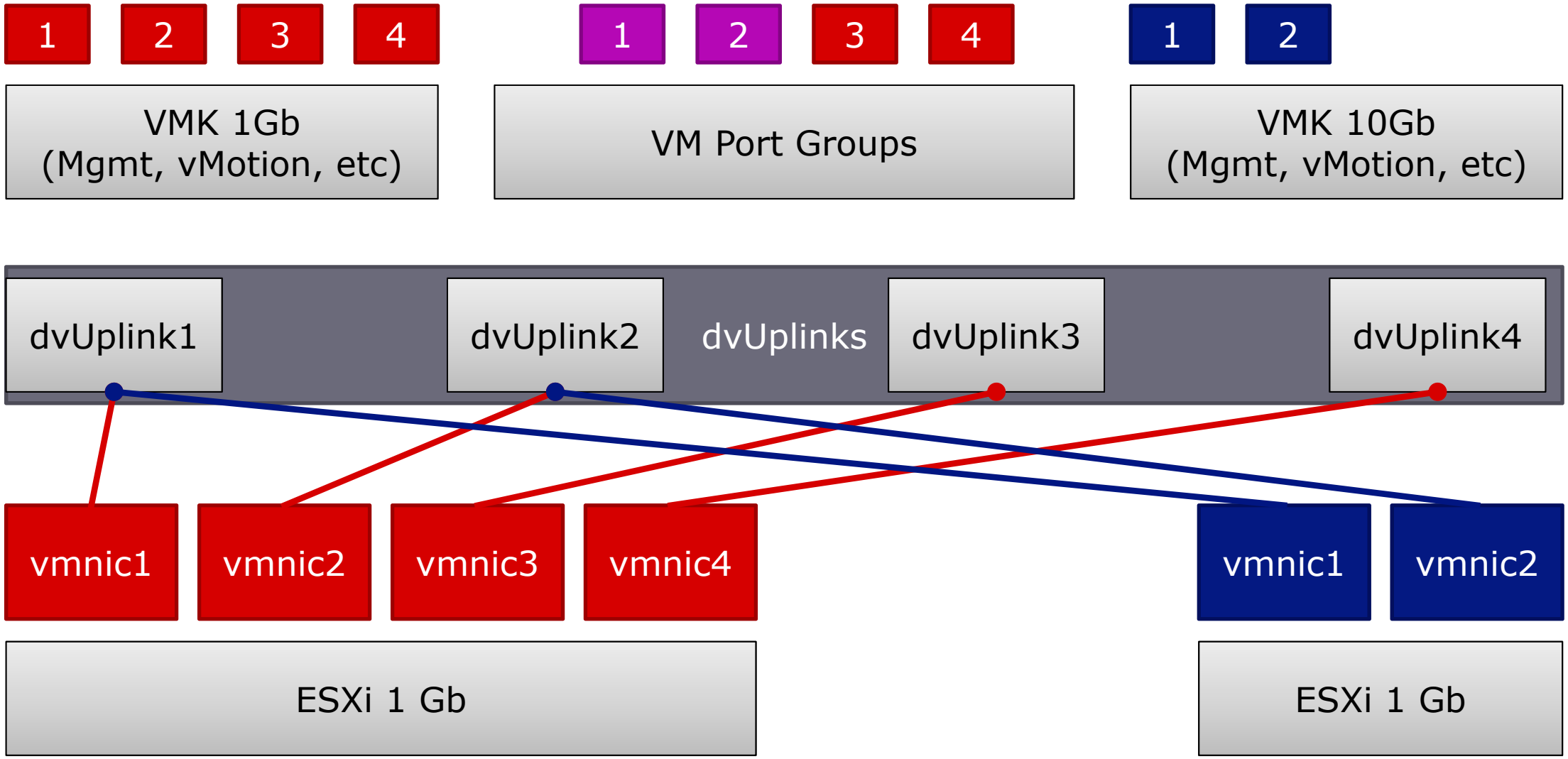


vmnic0 was left off to make the numbers match

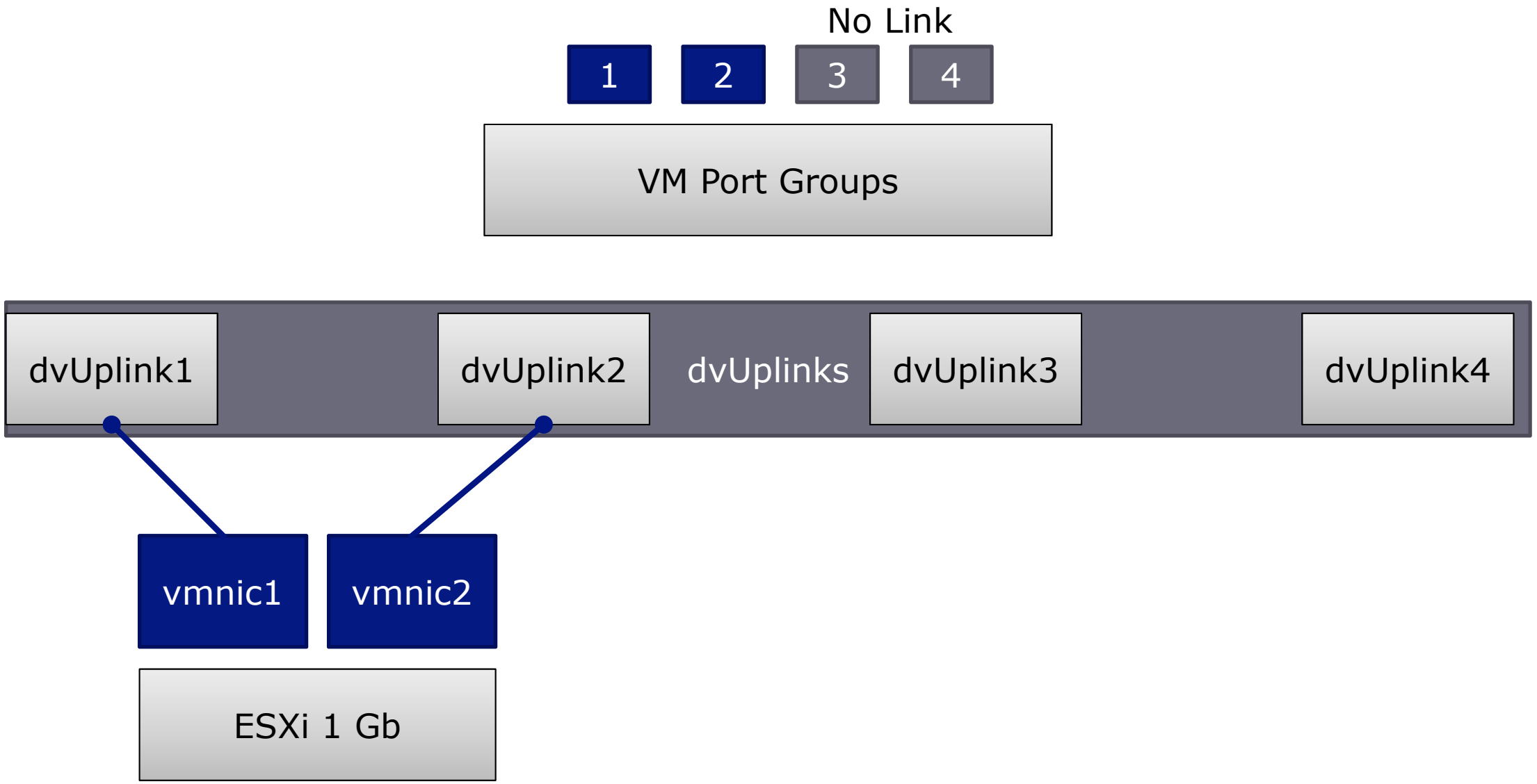
VDS with 10 Gb NICs



VDS with Mixed NIC Speeds



VM Port Groups on 10 GbE Host



Use Case – Multiple vMotion Host Saturation

Triggers:

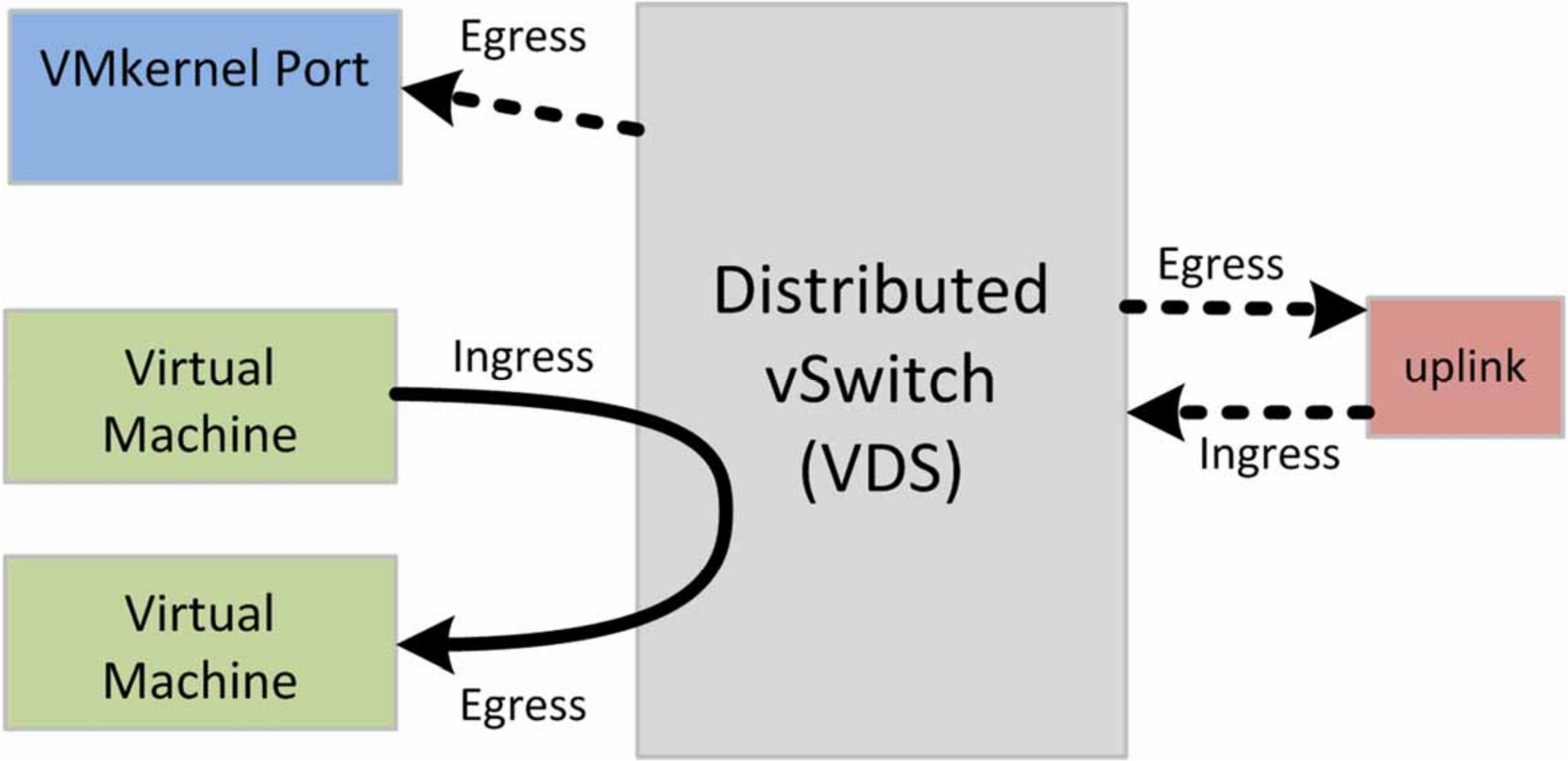
- Multiple hosts migrate VMs to a single host
- 2+ host maintenance mode
- DRS migrations
- DRS affinity and anti-affinity rules

Use Case – Multiple vMotion Host Saturation

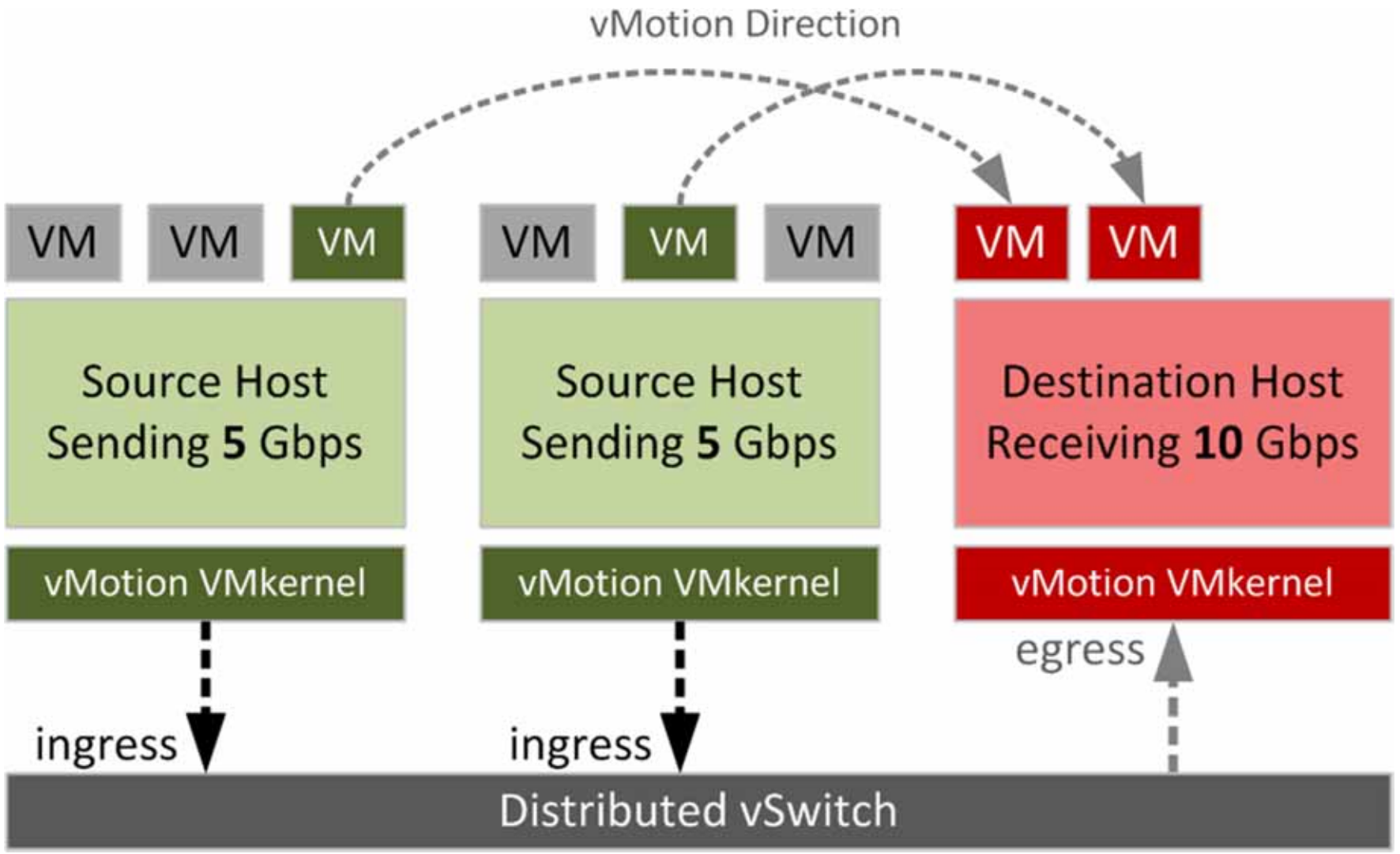
Tips and Advice

- Know how **Ingress** vs **Egress** works in VDS
- Use **NIOC** for source-based control
- Use **Traffic Shaping** for destination-based control

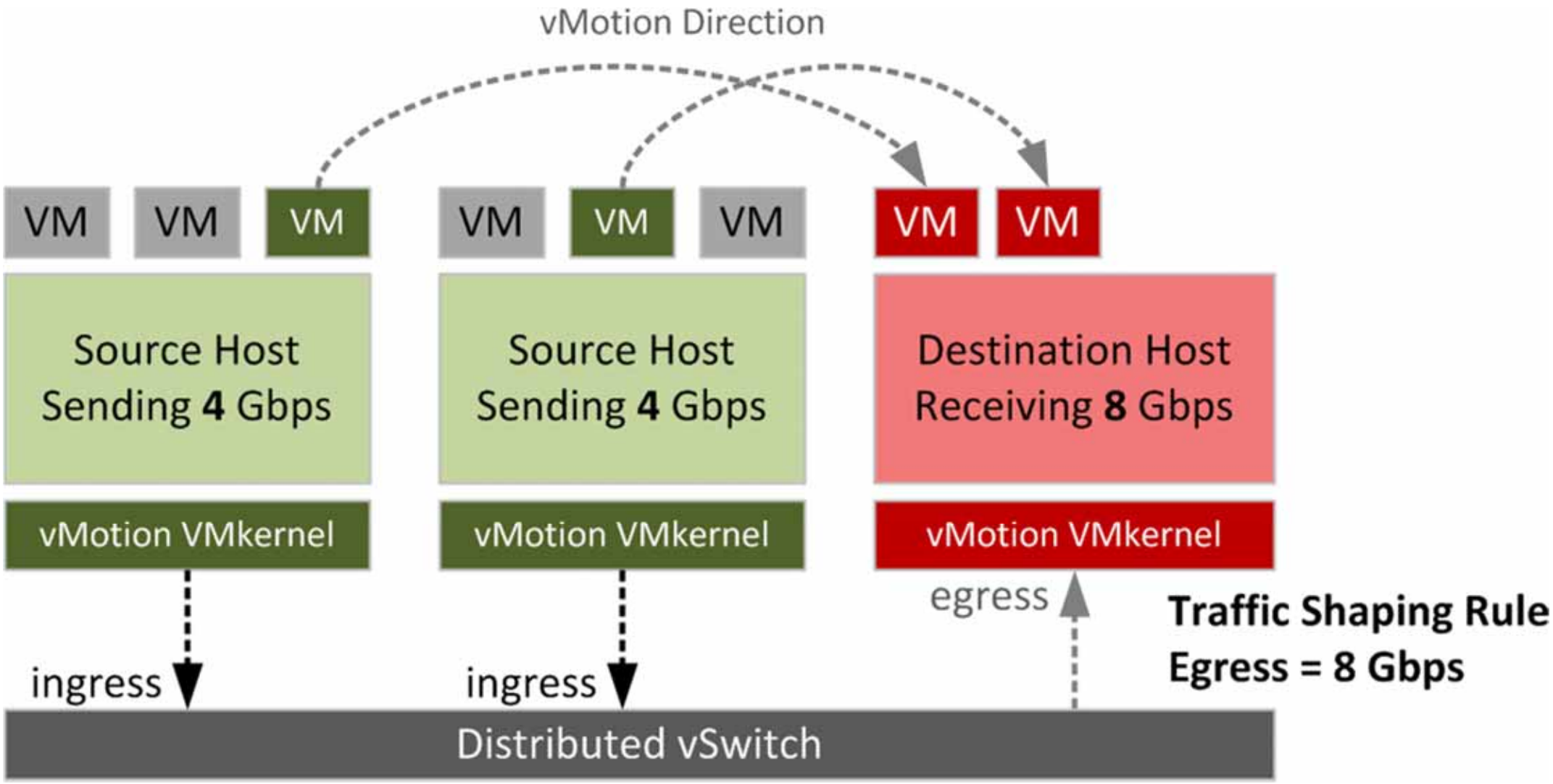
VDS Point of View: Ingress vs Egress



Traffic Shaping – The Issue



Traffic Shaping – The Solution



Use Case – Limiting vSphere Replication Traffic (SRM)

Triggers:

- Sharing a WAN pipe with other traffic
- Paying for bandwidth at certain % of peak
- Multiple VR tenants between data centers
- Contention with other backup or replication jobs

Use Case – Limiting vSphere Replication Traffic (SRM)

Tips and Advice:

- Use **NIOC** with Limits (per vmnic)
- Use **Network Resource** pool for VR
- Alternative is to limit based on VR **ports**

Setting NIOC for vSphere Replication

NET2745 | Actions

1

Summary | Monitor | Manage | Related Objects

2

Settings | Alarm Definitions | Tags | Permissions | Application Services | Hyperic Agents | Network Protocol Profiles | Ports | Resource Allocation

Physical network adapters: --

Bandwidth capacity: --

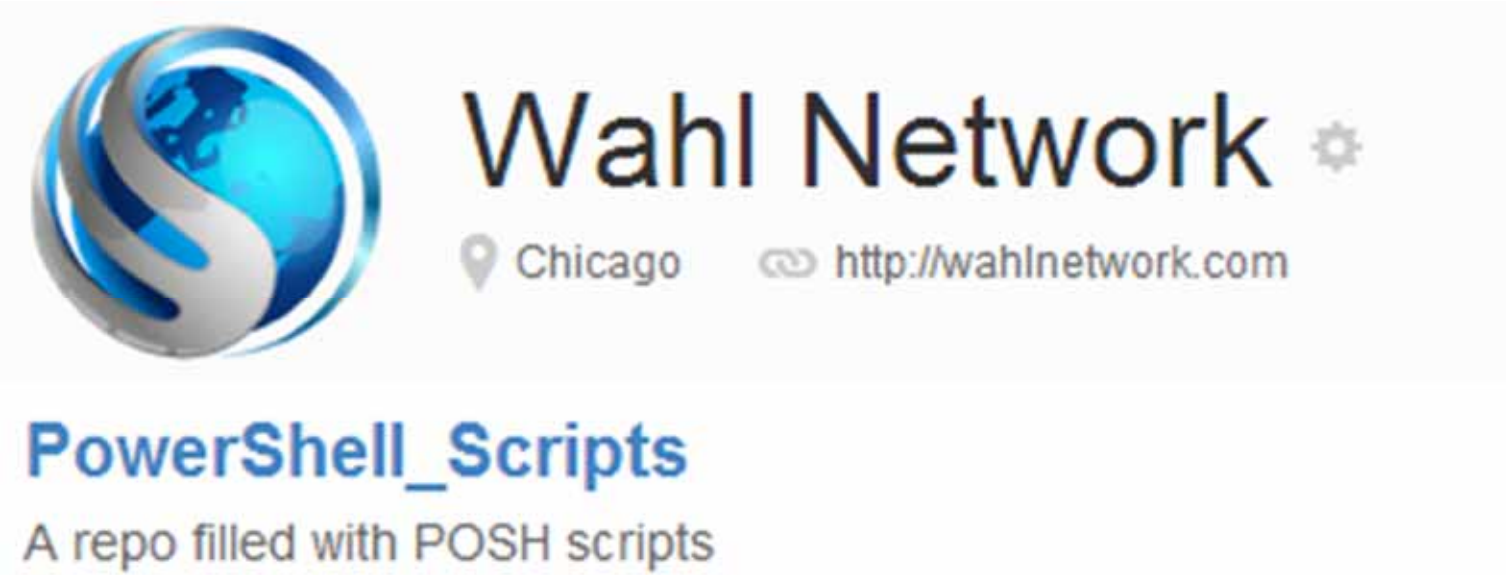
Network I/O Control: Enabled 3

Network Resource Pool	Limit (Mbps)	Physical Adapter Sha...	Shares Value	QoS Priority Tag
System network resource pools				
NFS Traffic	Unlimited	Normal	50	
Management Traffic	Unlimited	Normal	50	
vMotion Traffic	Unlimited	Normal	50	
Virtual SAN Traffic	Unlimited	Normal	50	
4 vSphere Replication (VR) Traffic	500	Normal	50	
iSCSI Traffic	Unlimited	Normal	50	
Virtual Machine Traffic	Unlimited	High	100	
Fault Tolerance (FT) Traffic	Unlimited	Normal	50	

PowerCLI – Automatically Updating VR Limits

- Run script for different limits during day/night
- Requires PowerCLI and vCenter service account

Short URL = http://goo.gl/dAgqBz



Use Case – Quality of Service

Triggers:

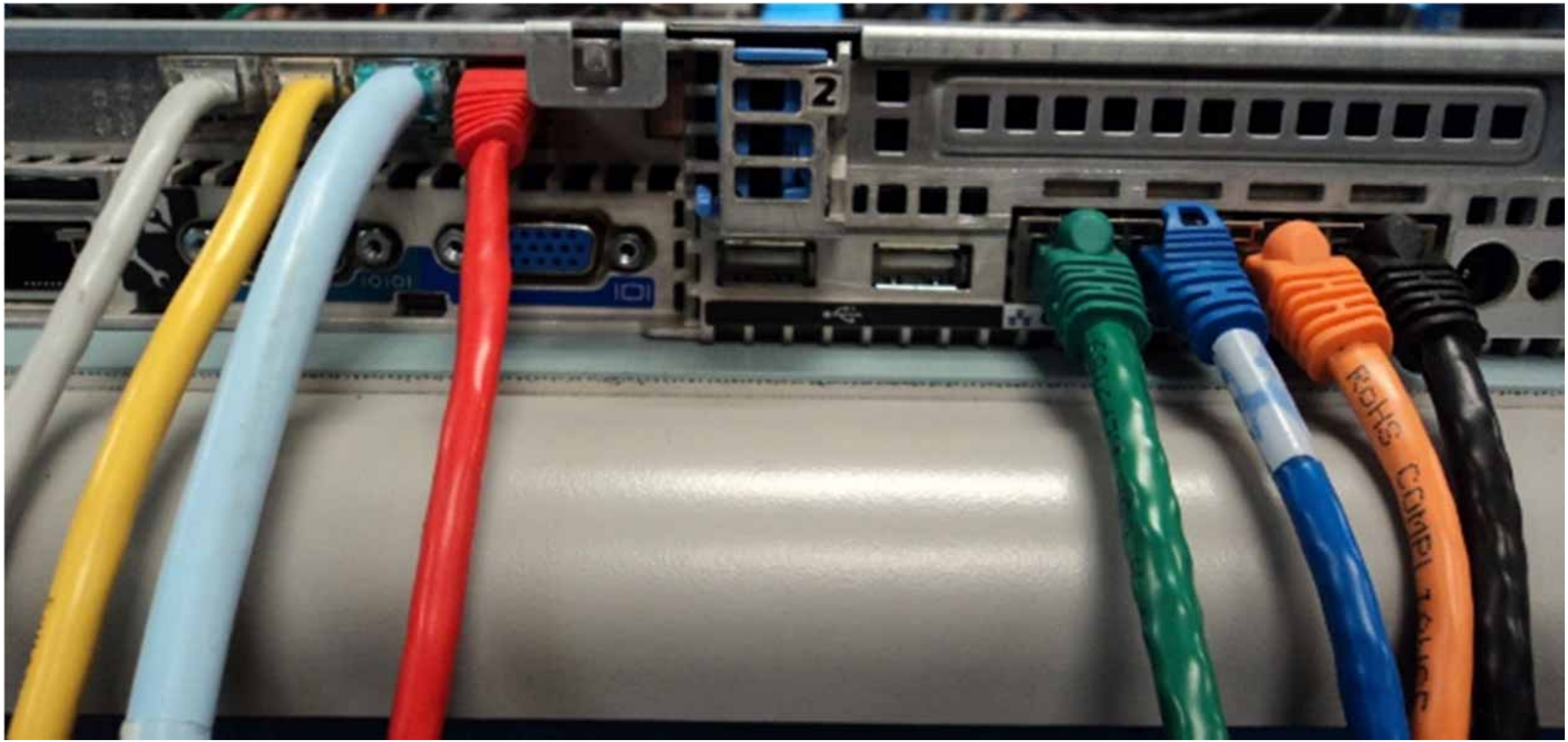
- Tag traffic for various SLAs
- Use L2 Priority Code Point (PCP)
- Use L3 Differentiated Services Code Point (DSCP)
- Data Center Bridging extensions in 802.1
  - Priority-based Flow Control (PFC) – 802.1Qbb
  - Enhanced Transmission Selection (ETS) – 802.1Qaz

Use Case – Quality of Service

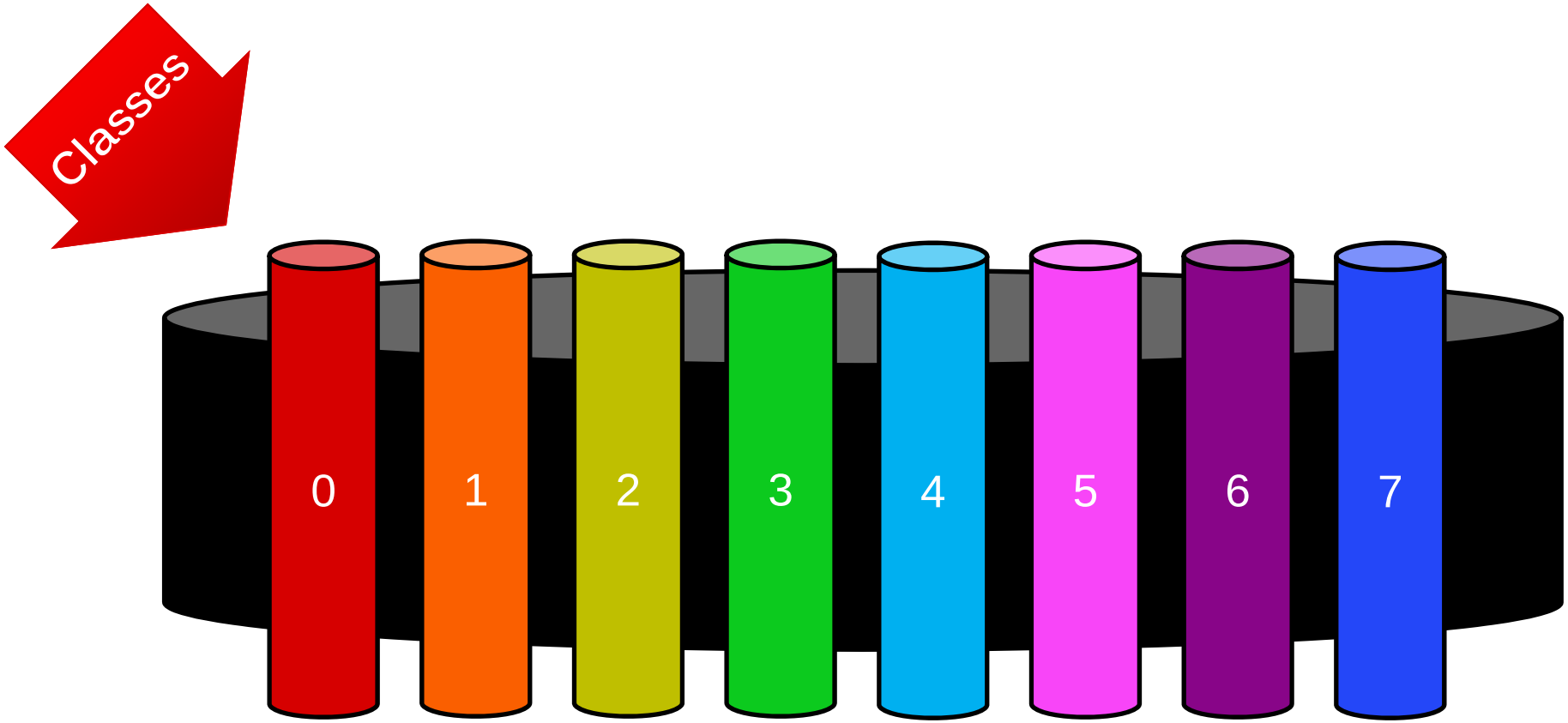
Tips and Advice:

- **KISS**: QoS solves contention problems
- Pick a place to tag traffic – **virtual or physical**
- Try not to enforce QoS in **too many** ways
- Use **clearly defined** tagging when **needed**
- **Avoid hard limits** on traffic flows

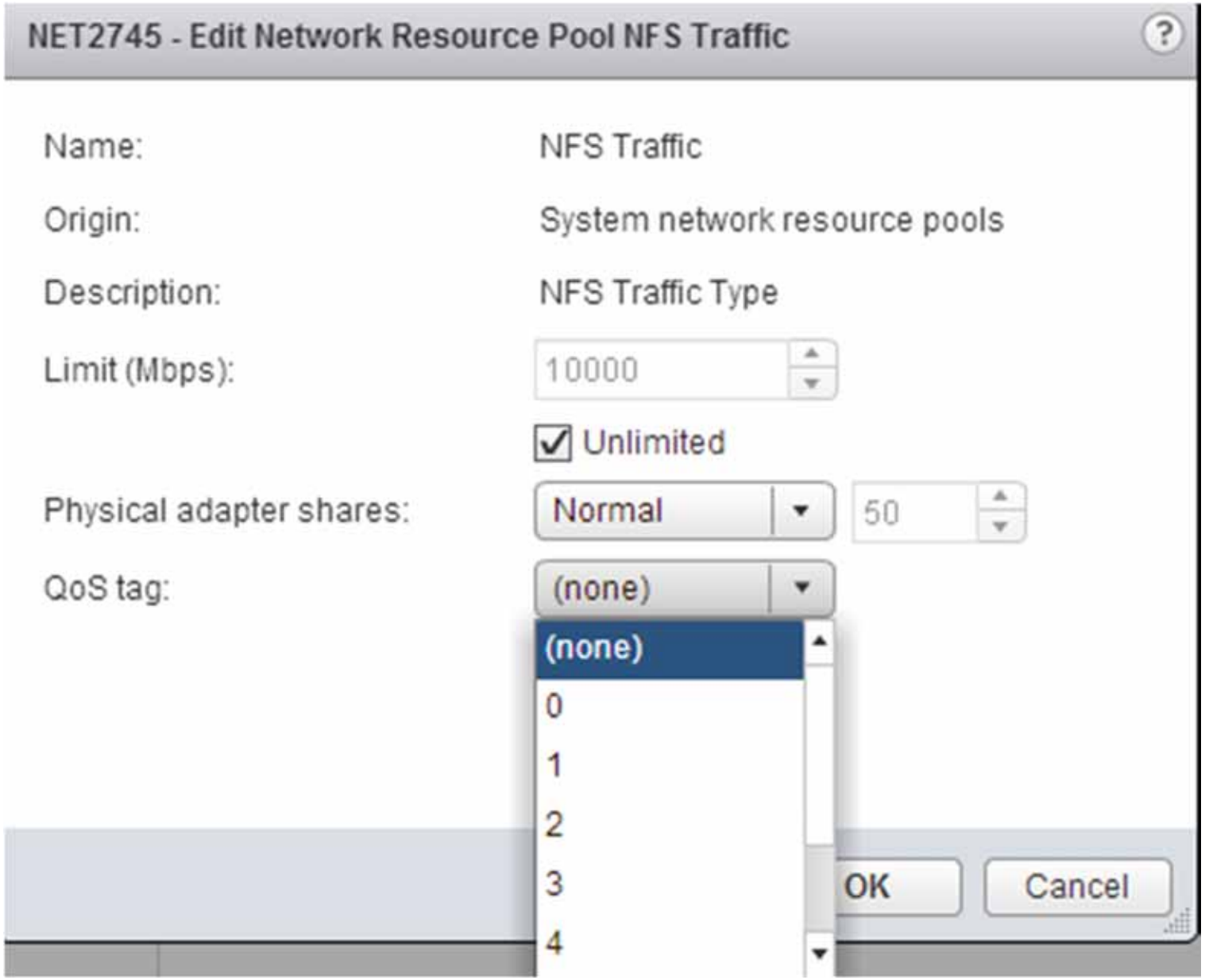
Old Method



New Method



Layer 2 QoS - PCP



Layer 3 QoS - DSCP

NET2745-DSCP_PG - Edit Settings 1

- General
- Advanced
- Security
- Traffic shaping
- VLAN
- Teaming and failover
- Monitoring
- Traffic filtering and marking 2
- Miscellaneous

Status: Enabled 3

No	Rule Name	Action	Traffic Direction
1	Spongebob_Videos 4	Tag (DSCP: 43)	Ingress/Egress

Traffic qualifiers 5

Type	Protocol/Traffic Type	Source Port	Destination Port
IP	TCP	*any*	80

Hardware QoS Example – Cisco UCS

Priority	Enabled	CoS	Packet Drop	Weight	Weight (%)	MTU	Multicast Optimized
Platinum	☐	5	☑	10	N/A	9000	☐
Gold	☐	4	☑	9	N/A	normal	☐
Silver	☑	2	☐	8	44	9216	☐
Bronze	☐	1	☑	7	N/A	normal	☐
Best Effort	☑	Any	☑	5	27	9216	☐
Fibre Channel	☑	3	☐	5	29	fc	N/A

Use Case – Load Distribution vs Load Balancing

Triggers:

- Network and Server teams not cooperating
  - *Pop out of those silos!*
- Poor convergence times during link failover
- Poor use of uplink throughput
- Excessive Topology Change Notifications (TCN)
- Excessive vMotion activity

What Is The Difference:
Load Distribution vs Load Balancing

Load Distribution

- Assigning workloads to uplinks based on identifiers
- Example: L2, L3, L4, and VLAN values

Load Balancing

- Assigning workloads to uplinks based on traffic
- Example: “Route based on physical NIC load”

Decision Tree
Load Distribution vs Load Balancing

Load Distribution

- **No** iSCSI Binding or Multi-NIC vMotion
- Potential Layer 2 Path Optimization

Load Balancing

- Imbalanced NIC saturation
- Network can tolerate TCN and MAC table updates

MAC Learning Through RARPs

```
⊟ Ethernet II, Src: 00:50:56:a2:67:90 (00:50:56:a2:67:90)
  ⊞ Destination: ff:ff:ff:ff:ff:ff (ff:ff:ff:ff:ff:ff)
  ⊞ Source: 00:50:56:a2:67:90 (00:50:56:a2:67:90)
    Type: RARP (0x8035)
    Trailer: 000000000000000000000000000000000000
⊞ Address Resolution Protocol (reverse request)
```

Options:
Load Distribution vs Load Balancing

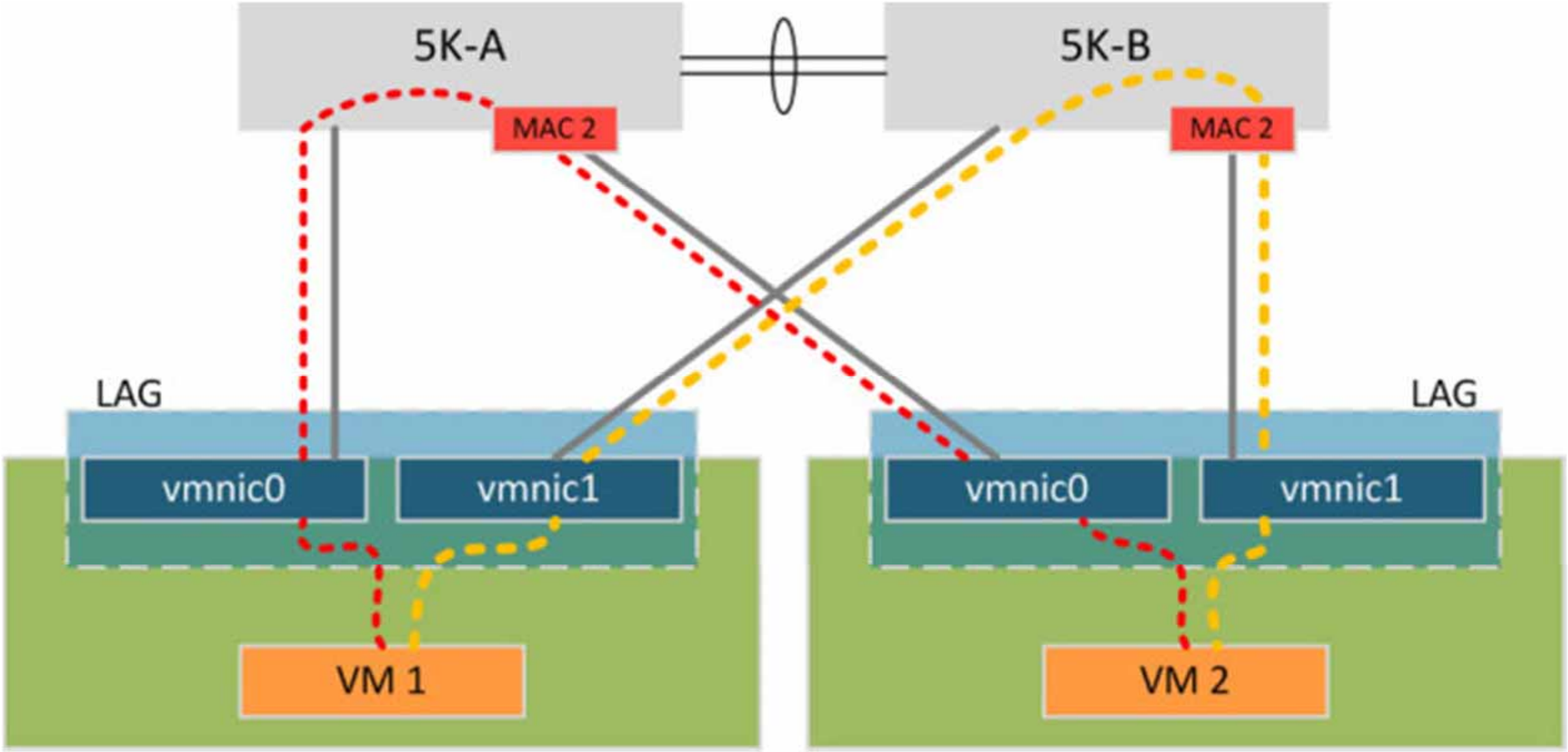
Load Distribution

- **Link Aggregation Group** (LAG)
- **Static** (EtherChannel) or **Dynamic** (LACP)

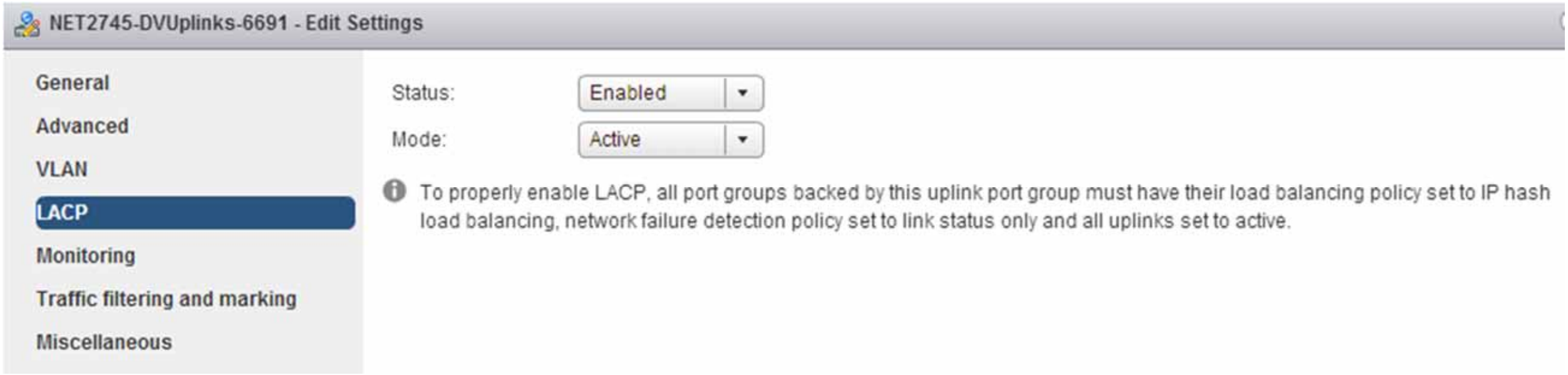
Load Balancing

- Set PG to *Route based on physical NIC load*
- Also known as **Load Based Teaming** (LBT)

Load Distribution – Topology View



Load Distribution – Step 1: Enable LACP on DVUplinks



Load Distribution – Step 2: Enable Enhanced LACP Support

NET2745 - Enhance LACP Support

1 Overview

2 Validate prerequisites

3 Ready to complete

Overview

Review the new features provided when enhancing LACP support on the distributed switch.

It is recommended that you backup the configuration of the distributed switch before proceeding.

Export configuration

Features of the enhanced LACP support

- Support for configuring multiple link aggregation groups (LAGs).
- LAGs are represented as uplinks in the teaming and failover policy of distributed ports or port groups. You can create a distributed switch configuration that uses both LACP and existing teaming algorithms on different port groups.
- Multiple load balancing options for LAGs.
- Centralized switch-level configuration available under *Manage* > *Settings* > *LACP*.

Load Distribution – Step 3: Pass Validation Checks

NET2745 - Enhance LACP Support

1 Overview

2 Validate prerequisites

3 Create a Link Aggregation Group

4 Ready to complete

Validate prerequisites

To enhance the LACP support, certain prerequisites must be met.

Overall validation status: Passed

Prerequisite Check	Status
Port group accessibility	Passed
LACP configuration	Passed
Uplink teaming policy override	Passed
Host accessibility	Passed
Host connectivity	Passed

Load Distribution – Step 4: Create a LAG

NET2745 - Enhance LACP Support

- ✓ 1 Overview
- ✓ 2 Validate prerequisites
- 3 Create a Link Aggregation Group
- 4 Ready to complete

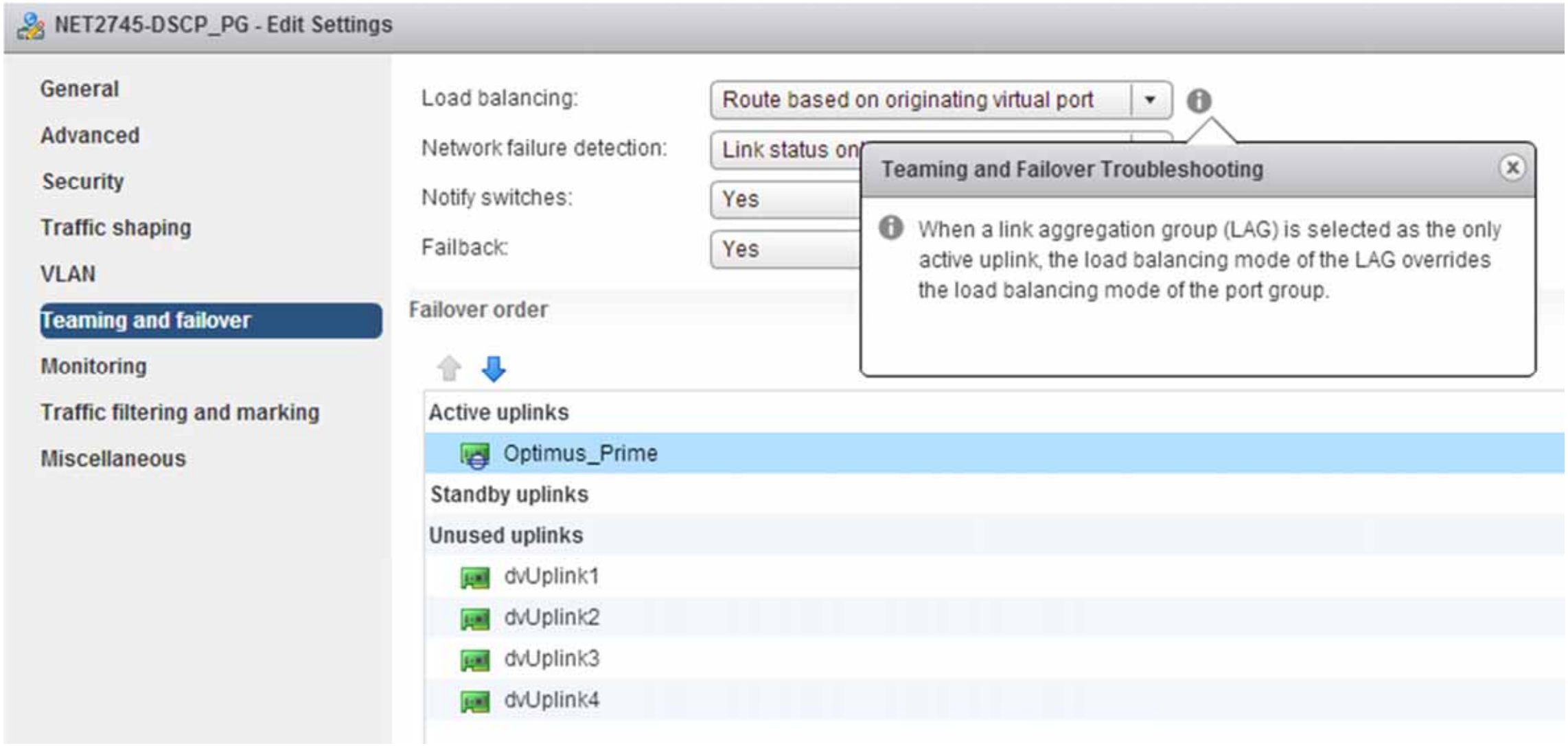
Create a Link Aggregation Group

Enter a name for the new link aggregation group.

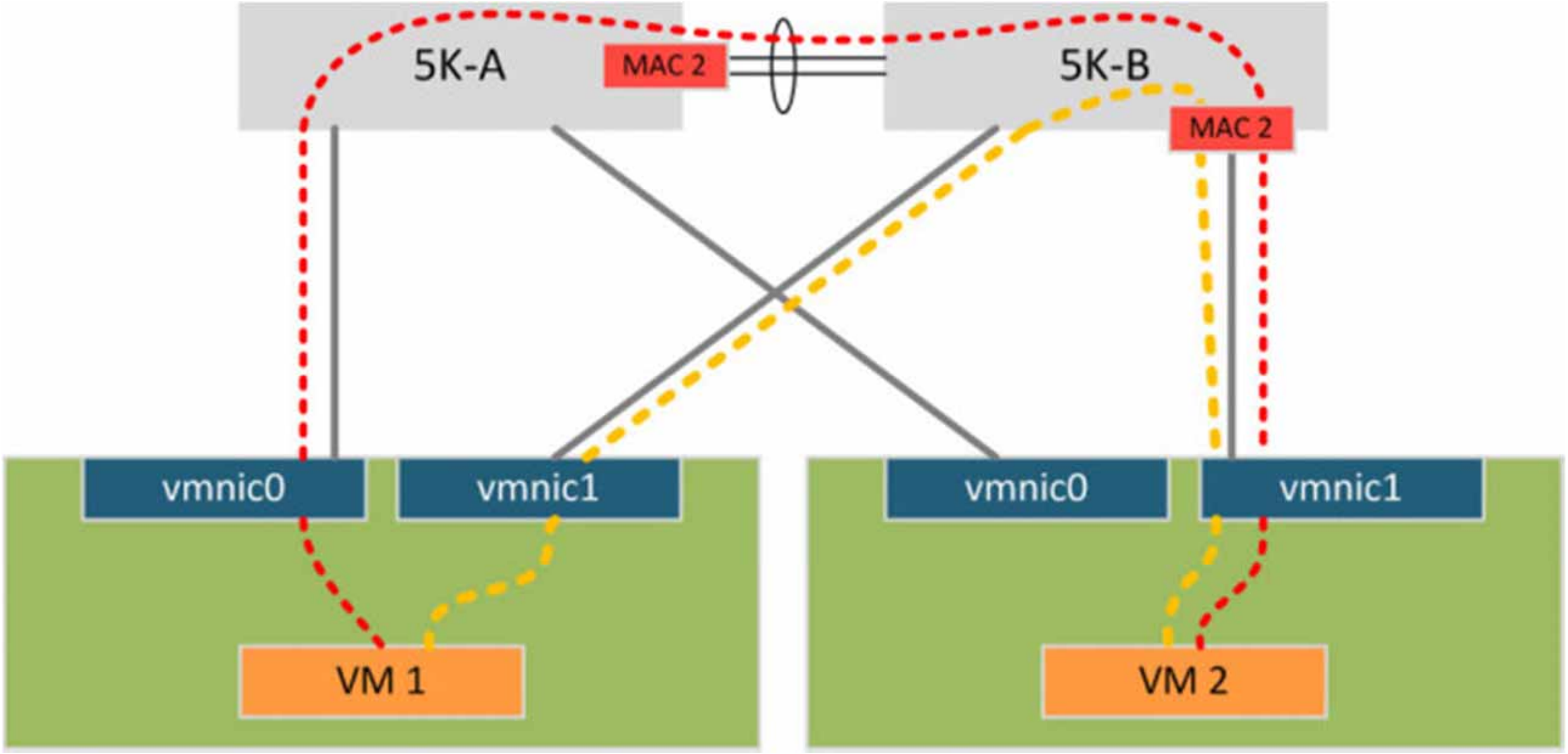
LACP is enabled on the distributed switch. The LACP configuration on the uplink port group will be converted to a LAG.

Name:	Optimus_Prime
Number of ports:	4
Mode:	Active
Load balancing mode:	Source IP address
Port policies	
VLAN:	Inherited from uplink port group
NetFlow:	Inherited from uplink port group

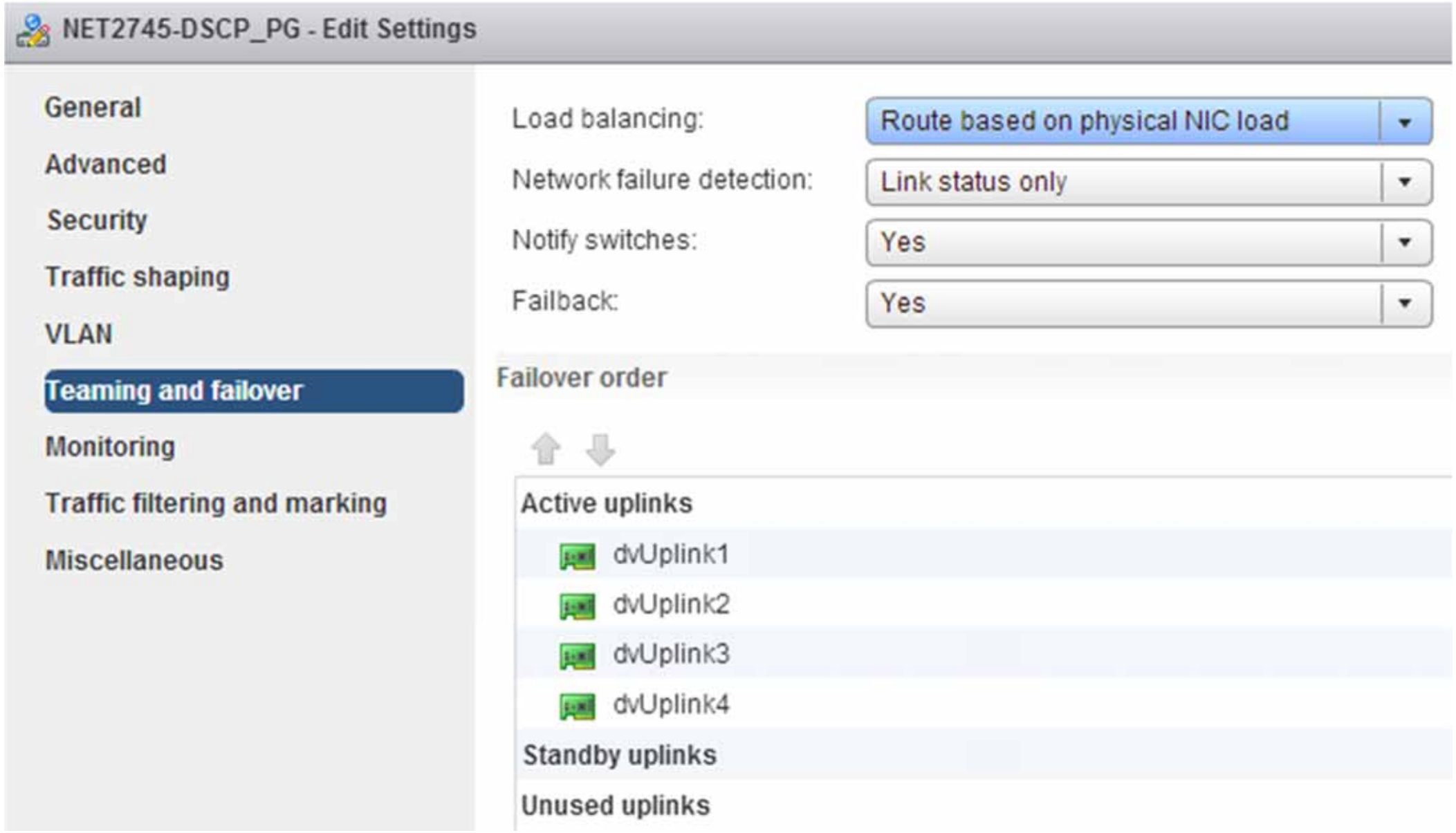
Load Distribution – Step 5: Use The LAG Uplink



Load Balancing – Topology View



Load Balancing – One Step!



More Content Available at …

Chris Wahl
@ChrisWahl
WahlNetwork.com

Jason Nash
@TheJasonNash
JasonNash.com

Thank You

vmworld® 2014

Fill out a survey

Every completed survey is entered into a drawing for a $25 VMware company store gift certificate

NO LIMITS

NET2745

vSphere Distributed Switch: Technical Deep Dive

Jason Nash, Varrow
Chris Wahl, AHEAD

vmworld® 2014