KVM FORUM

# Implementing SR-IOV Failover for Windows Guests During Migration

Annie Li - Principal Software Engineer, Oracle

Yan Vugenfirer - CEO, Daynix

http://github.com/virtio-win/

# Agenda

- Virtio-win drivers
- Windows guest terminology
- The problem
- Different solutions
- Failover solution with virtio-net on Windows guest

# Drivers for Windows

- Upstream: https://github.com/virtio-win/kvm-guest-drivers-windows/
- Drivers for the major virtio devices:
  - virtio-net
  - virtio-blk, virtio-scsi
  - virtio-balloon, virtio-serial, virtio-vsock, virtio-input, virtio-rng
- Panic, fw-cfg
- INF files (pci-serial, sm-bus on Q35)

# VirtIO Drivers for Windows

- WDF drivers for the “simple” devices
- Miniport architecture for network and storage
  - NDIS
  - Storport
  - Scsiport

# VirtIO Drivers for Windows

- Supported OS
  - Windows XP, Vista, 7, 8, 8.1, 10 (up to recent builds)
  - Widows Server 2003, 2008, 2008R2, 2012, 2012R2, 2016, 2019

# How to Contribute

- Send PRs - https://github.com/virtio-win/kvm-guest-drivers-windows/pulls
- Code changes should pass WHQL
- We are running WHQL CI on upstream (HCK-CI)

# Contributors

redhat

Virtuozzo

ORACLE

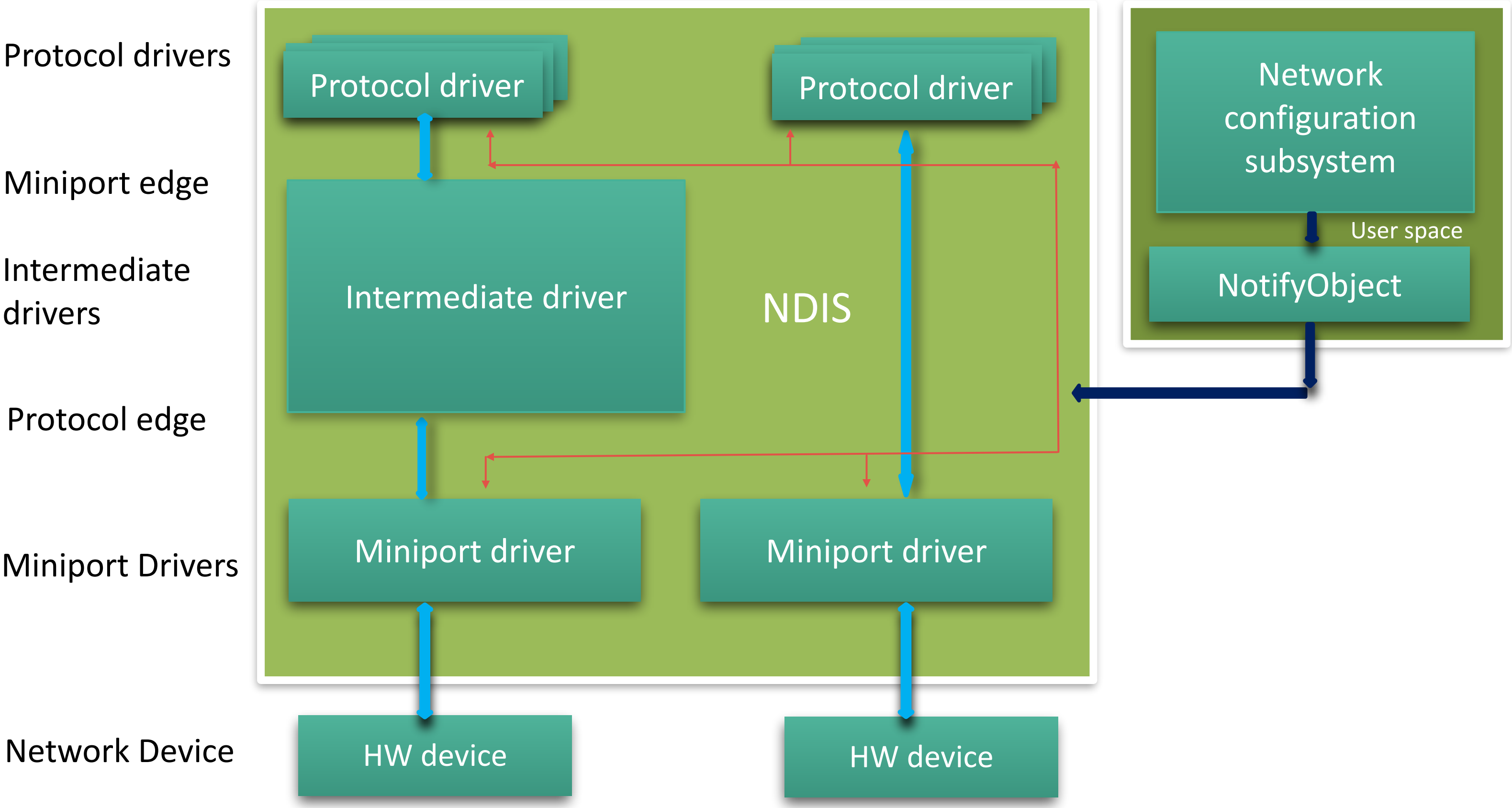
Google

Microsoft

aws

And others

# NDIS (Network Driver Interface Specification) Architecture



Protocol drivers

Miniport edge

Intermediate drivers

Protocol edge

Miniport Drivers

Network Device

Protocol driver

Protocol driver

Intermediate driver

NDIS

Miniport driver

Miniport driver

HW device

HW device

Network configuration subsystem

User space

NotifyObject

# Virtio-net (NetKVM) Driver for Windows

- NDIS miniport driver
- Basic driver package:
  - INF file – installation description
  - SYS file – driver binary
  - PDB file – symbols for debugging
  - CAT file - package digital signature

# The Problem – Live Migration and SR-IOV

# Overview of SR-IOV Migration Solutions

- Previous efforts and vendor specific HW solutions
- Hyper-V and Windows
- Linux and VirtIO

# Previous Efforts

- KVM Forum 2015 Live Migration with SR-IOV Pass-through - Weidong Han, Huawei
- KVM Forum 2018 - Live Migration Support for GPU with SR-IOV - Zheng Xiao, Alibaba Cloud; Jerry Jiang & Ken Xue, AMD
- KVM Forum 2020 (parallel session) -  Device Keepalive State for Local Live Migration and VMM Fast Restart - Jason Zeng, Intel

# Overview of Software Solutions

- Windows NIC Teaming
- Windows MUX Intermediate driver
- Hyper-V Solution

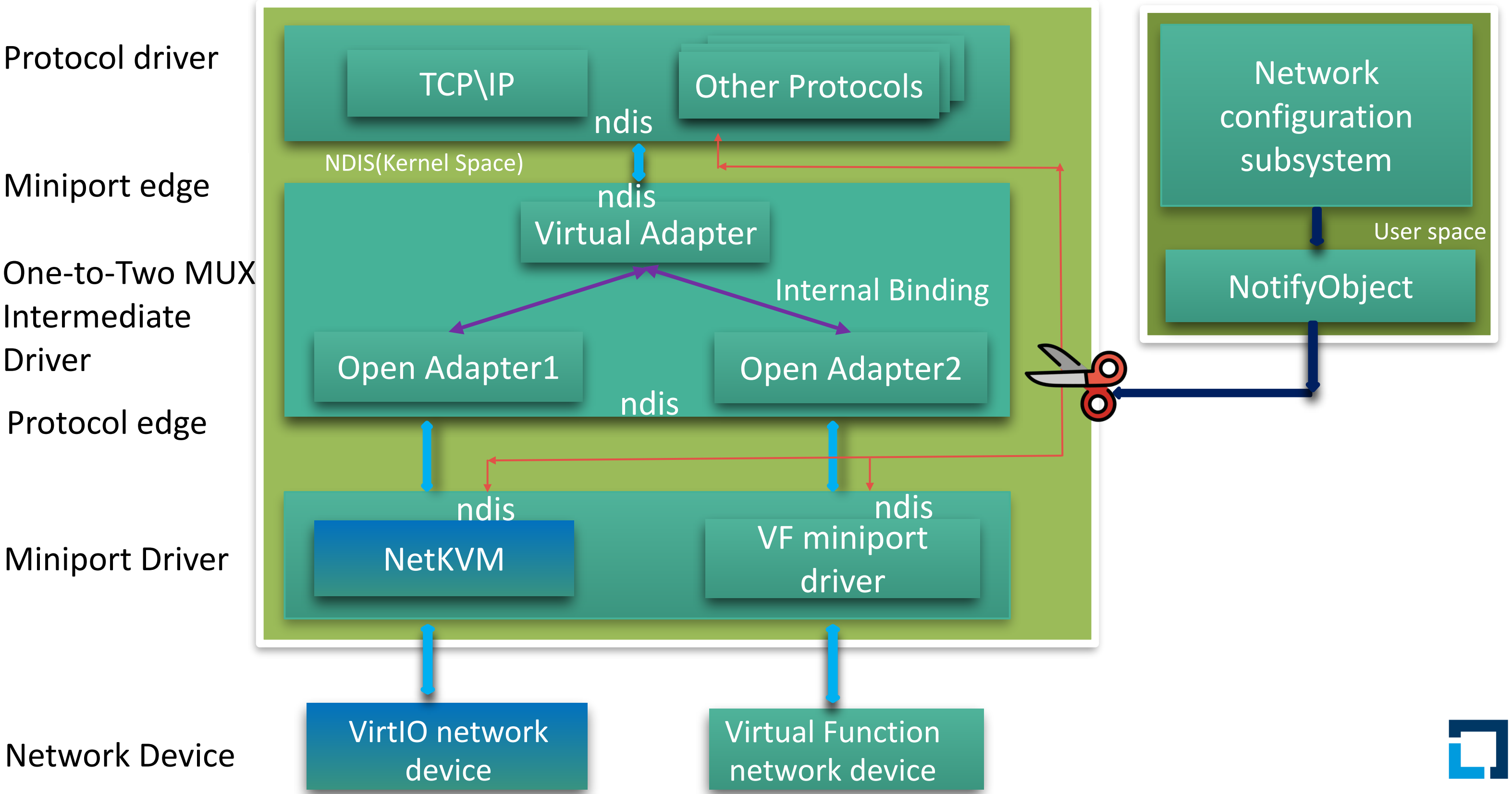
# Windows NIC Teaming

- Similar to bond in Linux
- Provides failover capability
- Configured through GUI or Powershell Cmdlets in user space

# Windows MUX Intermediate Driver

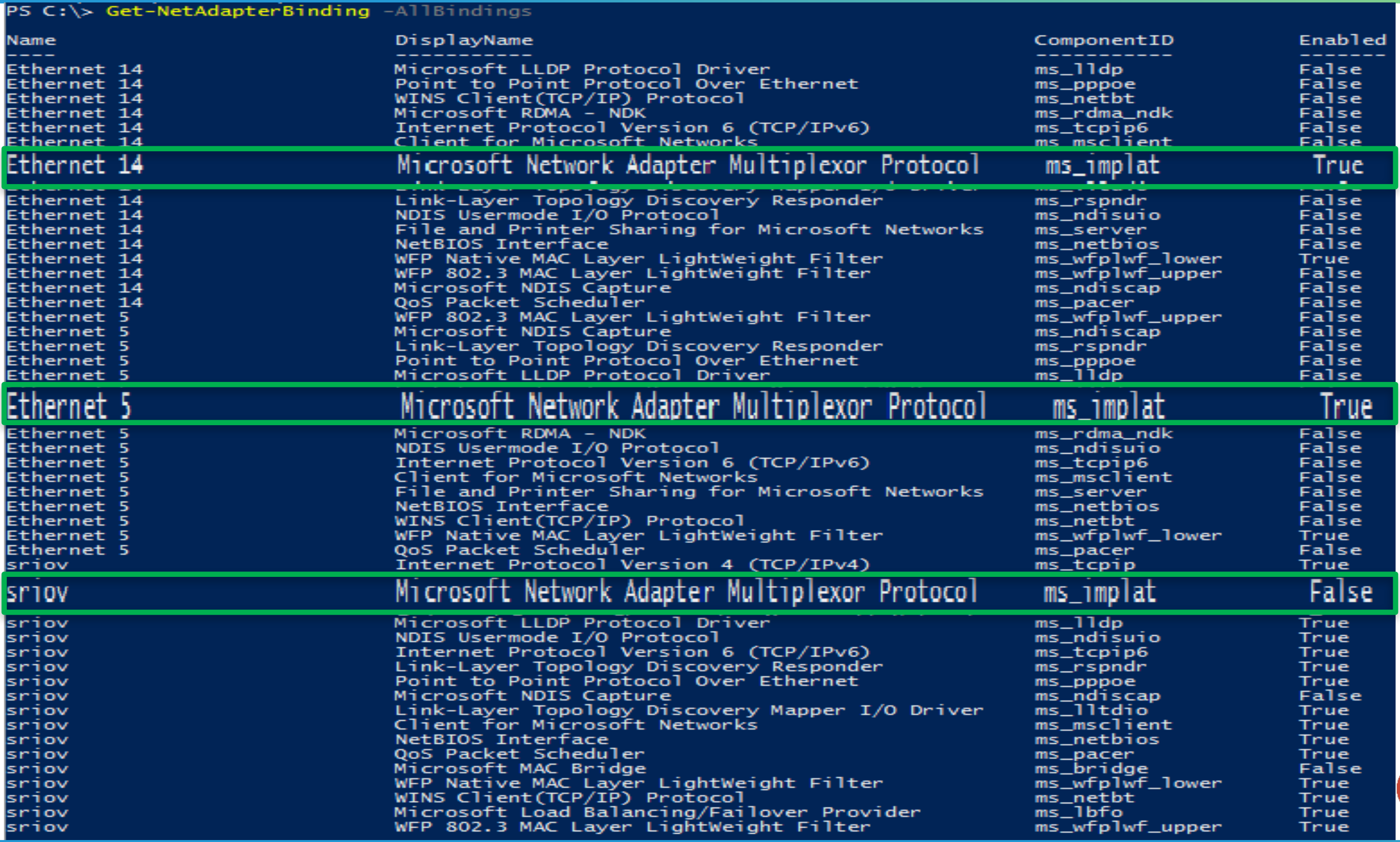
- Kernel space solution with various models
- One-to-two model for SR-IOV live migration

# Windows MUX Intermediate Driver



- Netkvm is Open Source driver code for VirtIO network
- VF miniport driver is provided by vendor

# Network Binding of NIC Teaming or MUX



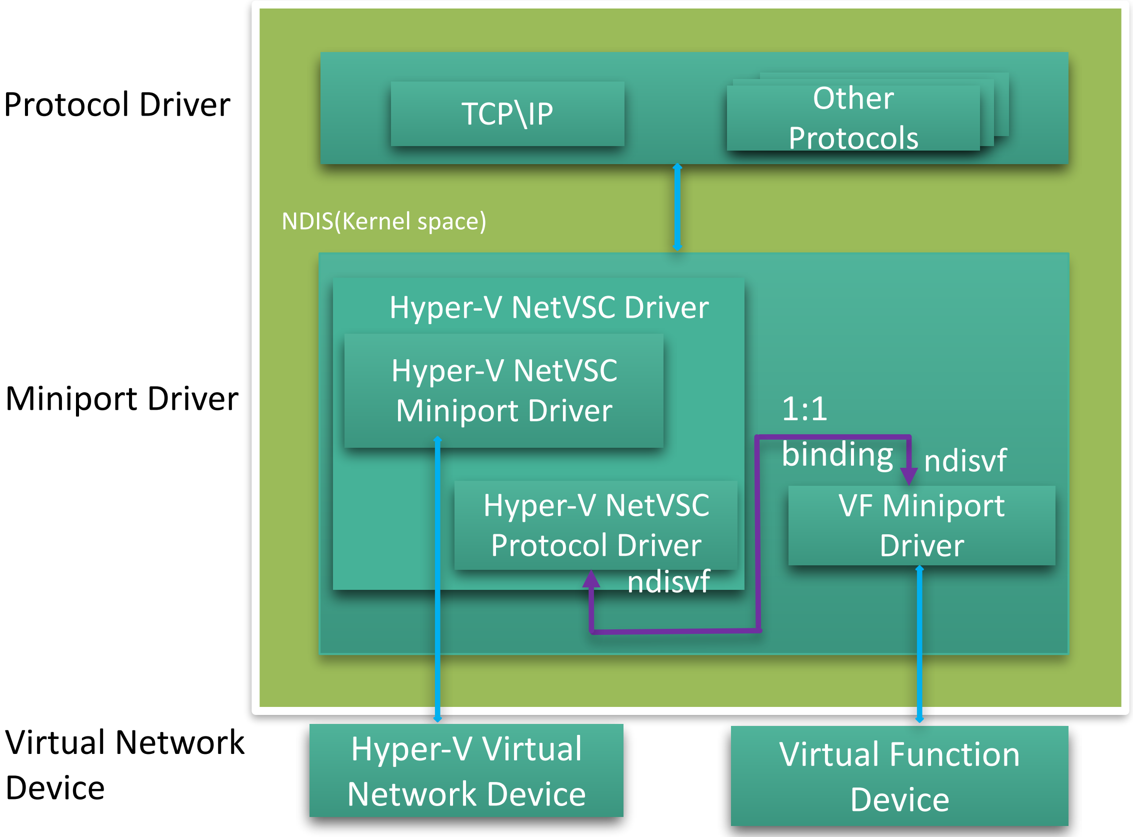
```
PS C:\> Get-NetAdapterBinding -AllBindings

Name                           DisplayName                                         ComponentID          Enabled
----                           -----------                                         -----------          -------
Ethernet 14                    Microsoft LLDP Protocol Driver                      ms_lldp              False
Ethernet 14                    Point to Point Protocol Over Ethernet               ms_pppoe             False
Ethernet 14                    WINS Client(TCP/IP) Protocol                        ms_netbt             False
Ethernet 14                    Microsoft RDMA - NDK                                ms_rdma_ndk          False
Ethernet 14                    Internet Protocol Version 6 (TCP/IPv6)              ms_tcpip6            False
Ethernet 14                    Client for Microsoft Networks                       ms_msclient          False
Ethernet 14                    Microsoft Network Adapter Multiplexor Protocol      ms_implat            True
Ethernet 14                    Link-Layer Topology Discovery Mapper I/O Driver     ms_lltdio            False
Ethernet 14                    Link-Layer Topology Discovery Responder             ms_rspndr            False
Ethernet 14                    NDIS Usermode I/O Protocol                          ms_ndisuio           False
Ethernet 14                    File and Printer Sharing for Microsoft Networks     ms_server            False
Ethernet 14                    NetBIOS Interface                                   ms_netbios           False
Ethernet 14                    WFP Native MAC Layer LightWeight Filter             ms_wfplwf_lower      True
Ethernet 14                    WFP 802.3 MAC Layer LightWeight Filter              ms_wfplwf_upper      False
Ethernet 14                    Microsoft NDIS Capture                              ms_ndiscap           False
Ethernet 14                    QoS Packet Scheduler                                ms_pacer             False
Ethernet 5                     WFP 802.3 MAC Layer LightWeight Filter              ms_wfplwf_upper      False
Ethernet 5                     Microsoft NDIS Capture                              ms_ndiscap           False
Ethernet 5                     Link-Layer Topology Discovery Responder             ms_rspndr            False
Ethernet 5                     Point to Point Protocol Over Ethernet               ms_pppoe             False
Ethernet 5                     Microsoft LLDP Protocol Driver                      ms_lldp              False
Ethernet 5                     Microsoft Network Adapter Multiplexor Protocol      ms_implat            True
Ethernet 5                     Microsoft RDMA - NDK                                ms_rdma_ndk          False
Ethernet 5                     NDIS Usermode I/O Protocol                          ms_ndisuio           False
Ethernet 5                     Internet Protocol Version 6 (TCP/IPv6)              ms_tcpip6            False
Ethernet 5                     Client for Microsoft Networks                       ms_msclient          False
Ethernet 5                     File and Printer Sharing for Microsoft Networks     ms_server            False
Ethernet 5                     NetBIOS Interface                                   ms_netbios           False
Ethernet 5                     WINS Client(TCP/IP) Protocol                        ms_netbt             False
Ethernet 5                     WFP Native MAC Layer LightWeight Filter             ms_wfplwf_lower      True
Ethernet 5                     QoS Packet Scheduler                                ms_pacer             False
sriov                          Internet Protocol Version 4 (TCP/IPv4)              ms_tcpip             True
sriov                          Microsoft Network Adapter Multiplexor Protocol      ms_implat            False
sriov                          Microsoft LLDP Protocol Driver                      ms_lldp              True
sriov                          NDIS Usermode I/O Protocol                          ms_ndisuio           True
sriov                          Internet Protocol Version 6 (TCP/IPv6)              ms_tcpip6            True
sriov                          Link-Layer Topology Discovery Responder             ms_rspndr            True
sriov                          Point to Point Protocol Over Ethernet               ms_pppoe             True
sriov                          Microsoft NDIS Capture                              ms_ndiscap           False
sriov                          Link-Layer Topology Discovery Mapper I/O Driver     ms_lltdio            True
sriov                          Client for Microsoft Networks                       ms_msclient          True
sriov                          NetBIOS Interface                                   ms_netbios           True
sriov                          QoS Packet Scheduler                                ms_pacer             True
sriov                          Microsoft MAC Bridge                                ms_bridge            False
sriov                          WFP Native MAC Layer LightWeight Filter             ms_wfplwf_lower      True
sriov                          WINS Client(TCP/IP) Protocol                        ms_netbt             True
sriov                          Microsoft Load Balancing/Failover Provider          ms_lbfo              True
sriov                          WFP 802.3 MAC Layer LightWeight Filter              ms_wfplwf_upper      True
```

# Hyper-V VM Network

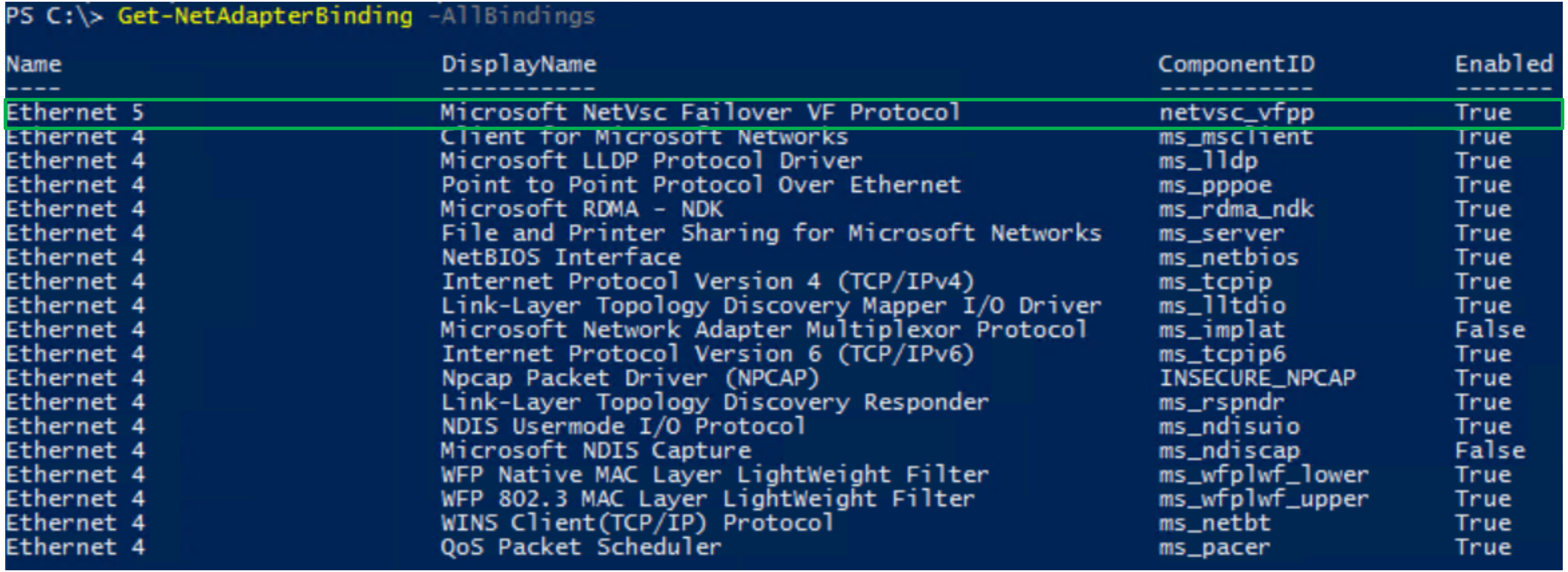
- Network virtual service client(NetVSC )
- Synthetic data path
- SR-IOV data path
- Two Installation files(INF)
- Share same driver binary

# Hyper-V **SR-IOV VF Failover**



- No Bond/Teaming
- No NotifyObject
- No new Virtual Adapter

# Network Binding in Hyper-V



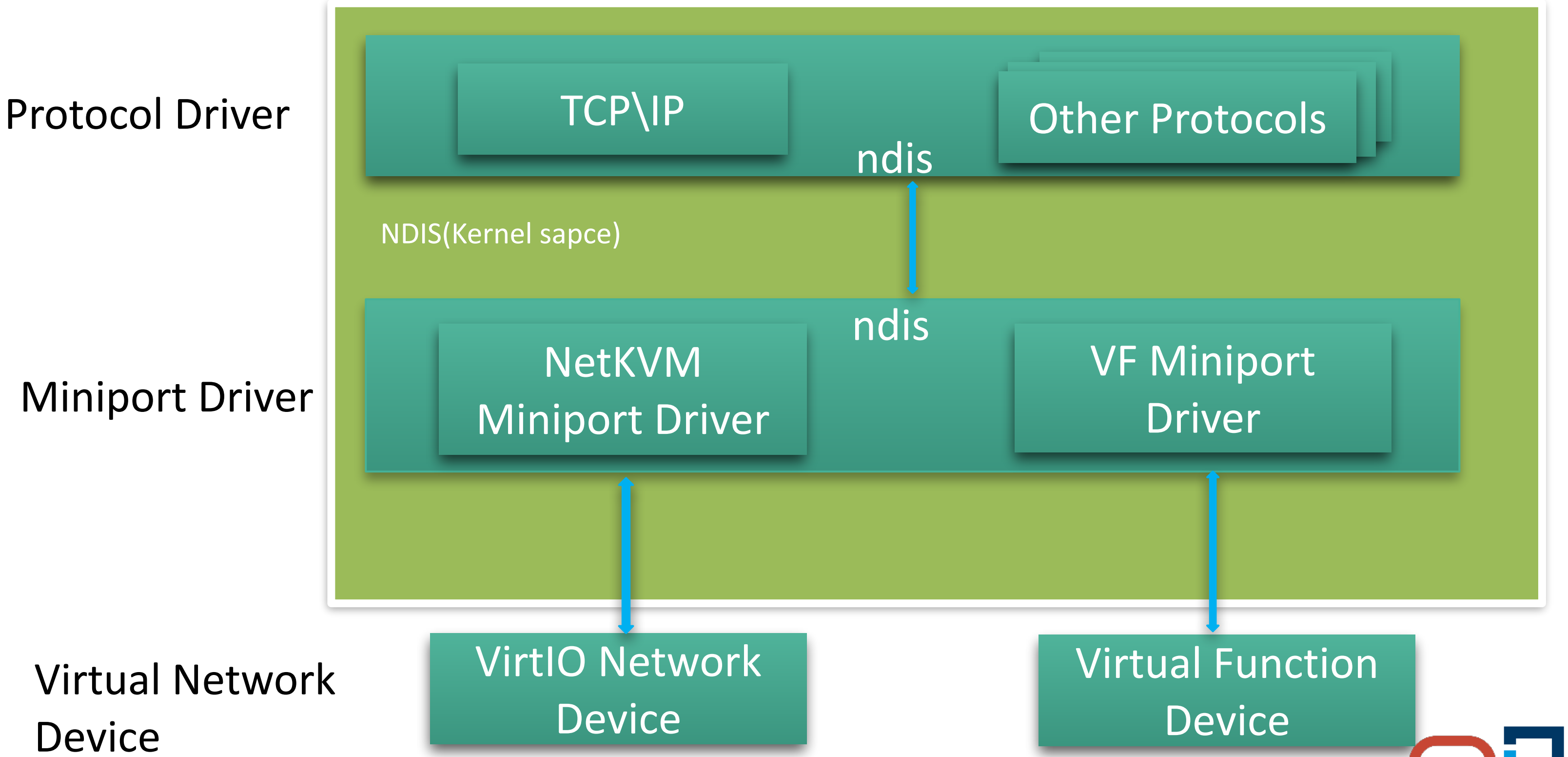
```
PS C:\> Get-NetAdapterBinding -AllBindings

Name                           DisplayName                                        ComponentID           Enabled
----                           -----------                                        -----------           -------
Ethernet 5                     Microsoft NetVsc Failover VF Protocol              netvsc_vfpp           True
Ethernet 4                     Client for Microsoft Networks                      ms_msclient           True
Ethernet 4                     Microsoft LLDP Protocol Driver                     ms_lldp               True
Ethernet 4                     Point to Point Protocol Over Ethernet              ms_pppoe              True
Ethernet 4                     Microsoft RDMA - NDK                               ms_rdma_ndk           True
Ethernet 4                     File and Printer Sharing for Microsoft Networks    ms_server             True
Ethernet 4                     NetBIOS Interface                                  ms_netbios            True
Ethernet 4                     Internet Protocol Version 4 (TCP/IPv4)             ms_tcpip              True
Ethernet 4                     Link-Layer Topology Discovery Mapper I/O Driver    ms_lltdio             True
Ethernet 4                     Microsoft Network Adapter Multiplexor Protocol     ms_implat             False
Ethernet 4                     Internet Protocol Version 6 (TCP/IPv6)             ms_tcpip6             True
Ethernet 4                     Npcap Packet Driver (NPCAP)                        INSECURE_NPCAP        True
Ethernet 4                     Link-Layer Topology Discovery Responder            ms_rspndr             True
Ethernet 4                     NDIS Usermode I/O Protocol                         ms_ndisuio            True
Ethernet 4                     Microsoft NDIS Capture                             ms_ndiscap            False
Ethernet 4                     WFP Native MAC Layer LightWeight Filter            ms_wfplwf_lower       True
Ethernet 4                     WFP 802.3 MAC Layer LightWeight Filter             ms_wfplwf_upper       True
Ethernet 4                     WINS Client(TCP/IP) Protocol                       ms_netbt              True
Ethernet 4                     QoS Packet Scheduler                               ms_pacer              True
```

# Comparison Summary

- MUX Driver model:
  - Complicated, New virtual adapter, Restore offload, NotifyObject.
- Hyper-V model:
  - Simplified, Appropriate for Hyper-V
- 2-netdev model in KVM

# Windows Network of VirtIO and VF



Protocol Driver

TCP\IP

Other Protocols

ndis

NDIS(Kernel sapce)

ndis

Miniport Driver

NetKVM Miniport Driver

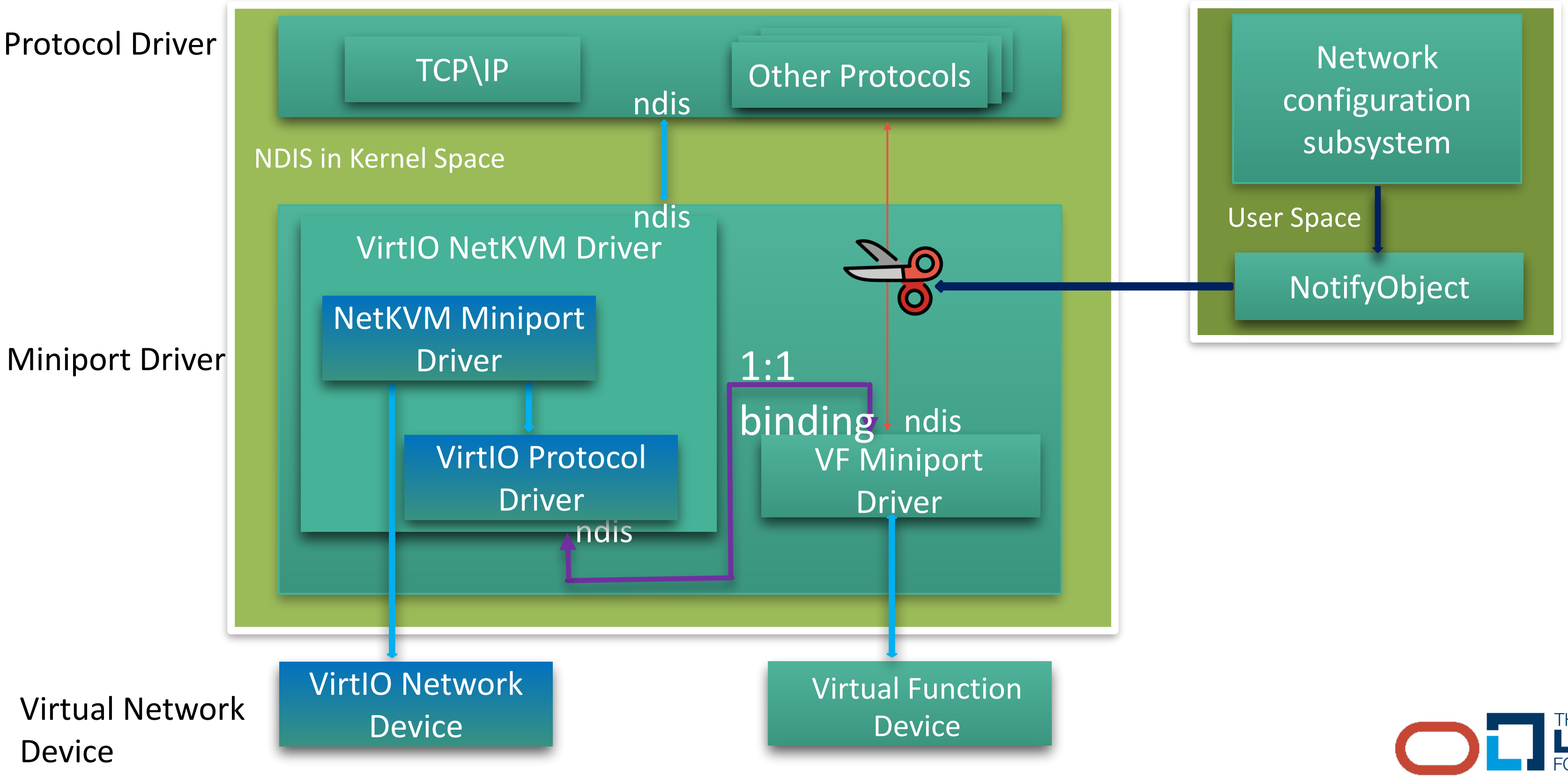
VF Miniport Driver

Virtual Network Device

VirtIO Network Device

Virtual Function Device

# VirtIO SR-IOV VF Failover

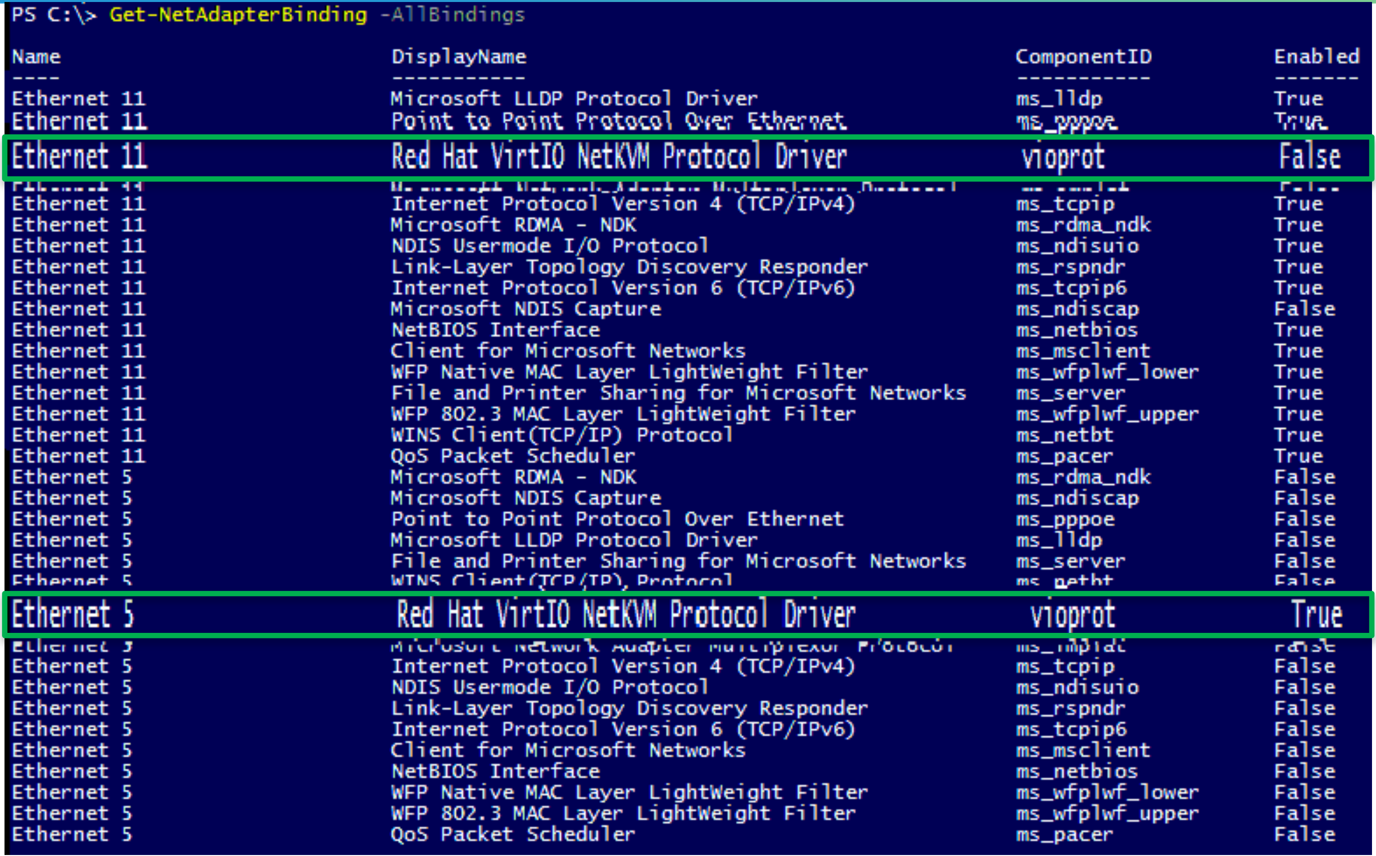


- Netkvm is Open Source driver code for VirtIO network
- VF miniport driver is provided by vendor

# Protocol Driver in 2-netdev Model

- Behaves like a bridge
- VF adapter is coupled to VirtIO adapter with the same MAC address
- Handling TX/RX network data
- Object identifiers(OIDs) are wrapped and forwarded, offloads are propagated

# Network Binding of VirtIO SR-IOV



```
PS C:\> Get-NetAdapterBinding -AllBindings

Name                           DisplayName                                        ComponentID          Enabled
----                           -----------                                        -----------          -------
Ethernet 11                    Microsoft LLDP Protocol Driver                     ms_lldp              True
Ethernet 11                    Point to Point Protocol Over Ethernet              ms_pppoe             True
Ethernet 11                    Red Hat VirtIO NetKVM Protocol Driver              vioprot              False
Ethernet 11                    Microsoft Network Adapter Multiplexor Protocol     ms_implat            False
Ethernet 11                    Internet Protocol Version 4 (TCP/IPv4)             ms_tcpip             True
Ethernet 11                    Microsoft RDMA - NDK                               ms_rdma_ndk          True
Ethernet 11                    NDIS Usermode I/O Protocol                         ms_ndisuio           True
Ethernet 11                    Link-Layer Topology Discovery Responder            ms_rspndr            True
Ethernet 11                    Internet Protocol Version 6 (TCP/IPv6)             ms_tcpip6            True
Ethernet 11                    Microsoft NDIS Capture                             ms_ndiscap           False
Ethernet 11                    NetBIOS Interface                                  ms_netbios           True
Ethernet 11                    Client for Microsoft Networks                      ms_msclient          True
Ethernet 11                    WFP Native MAC Layer LightWeight Filter            ms_wfplwf_lower      True
Ethernet 11                    File and Printer Sharing for Microsoft Networks    ms_server            True
Ethernet 11                    WFP 802.3 MAC Layer LightWeight Filter             ms_wfplwf_upper      True
Ethernet 11                    WINS Client(TCP/IP) Protocol                       ms_netbt             True
Ethernet 11                    QoS Packet Scheduler                               ms_pacer             True
Ethernet 5                     Microsoft RDMA - NDK                               ms_rdma_ndk          False
Ethernet 5                     Microsoft NDIS Capture                             ms_ndiscap           False
Ethernet 5                     Point to Point Protocol Over Ethernet              ms_pppoe             False
Ethernet 5                     Microsoft LLDP Protocol Driver                     ms_lldp              False
Ethernet 5                     File and Printer Sharing for Microsoft Networks    ms_server            False
Ethernet 5                     WINS Client(TCP/IP) Protocol                       ms_netbt             False
Ethernet 5                     Red Hat VirtIO NetKVM Protocol Driver              vioprot              True
Ethernet 5                     Microsoft Network Adapter Multiplexor Protocol     ms_implat            False
Ethernet 5                     Internet Protocol Version 4 (TCP/IPv4)             ms_tcpip             False
Ethernet 5                     NDIS Usermode I/O Protocol                         ms_ndisuio           False
Ethernet 5                     Link-Layer Topology Discovery Responder            ms_rspndr            False
Ethernet 5                     Internet Protocol Version 6 (TCP/IPv6)             ms_tcpip6            False
Ethernet 5                     Client for Microsoft Networks                      ms_msclient          False
Ethernet 5                     NetBIOS Interface                                  ms_netbios           False
Ethernet 5                     WFP Native MAC Layer LightWeight Filter            ms_wfplwf_lower      False
Ethernet 5                     WFP 802.3 MAC Layer LightWeight Filter             ms_wfplwf_upper      False
Ethernet 5                     QoS Packet Scheduler                               ms_pacer             False
```

# Known Issues

- DHCP issue
  - Only happens in specific scenario
- Statistics is missing
  - NetKVM driver needs to keep the statistics of packets sent to or received from VF driver
- Old Windows system
  - Windows Server 2003 and Windows XP
- Need to add VF PNP to Notification Object code or to the registry
- Possible race during boot?

# VirtIO and SR-IOV Failover

- VirtIO specification specifics
  - Feature bit called VIRTIO_NET_F_STANDBY. It is appropriate for 3-netdev model in Linux, but not for Windows 2-netdev model.

# Installation

- Before
  - INF for Miniport

- After
  - INF for Miniport
  - INF for Protocol driver definition and Notify Object

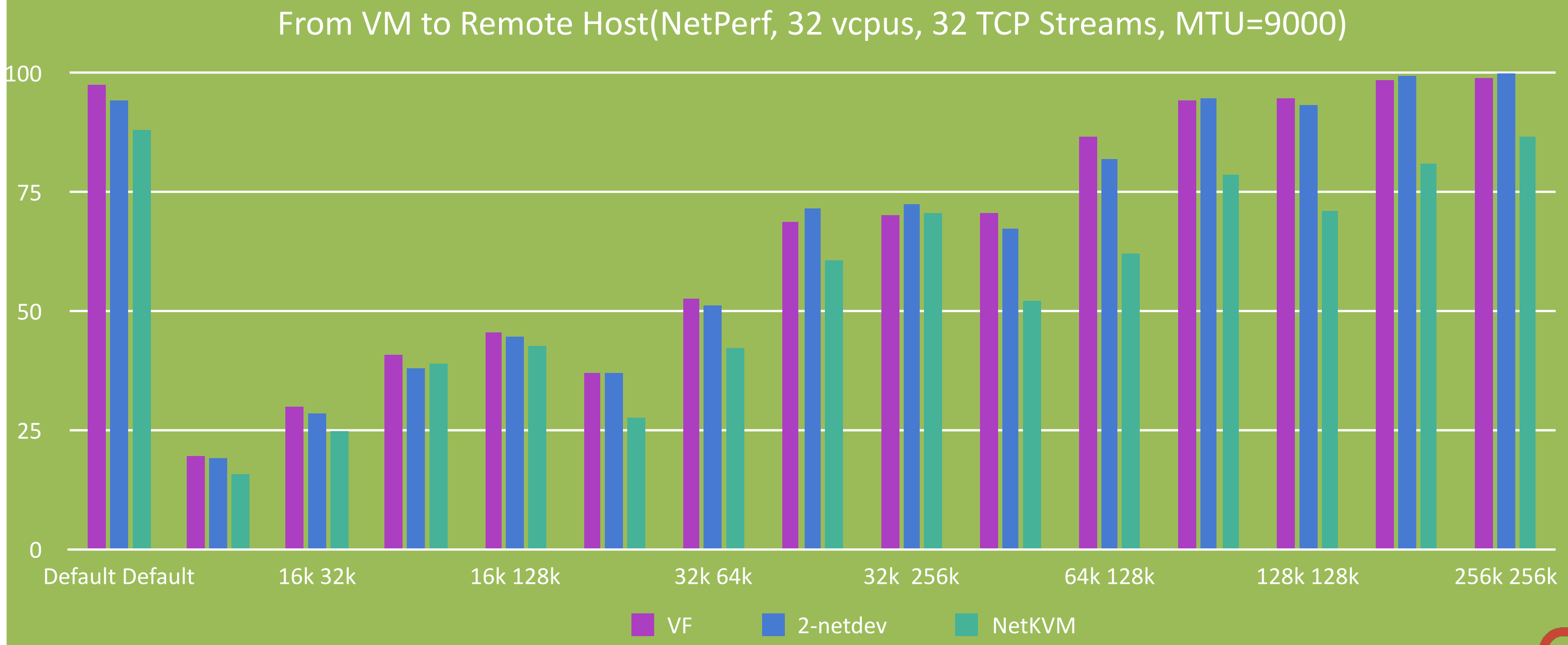
- Not part of the guest tools installer: https://github.com/virtio-win/virtio-win-guest-tools-installer

# WHQL Certification

- Before
  - Certification of miniport
  - Automatic review

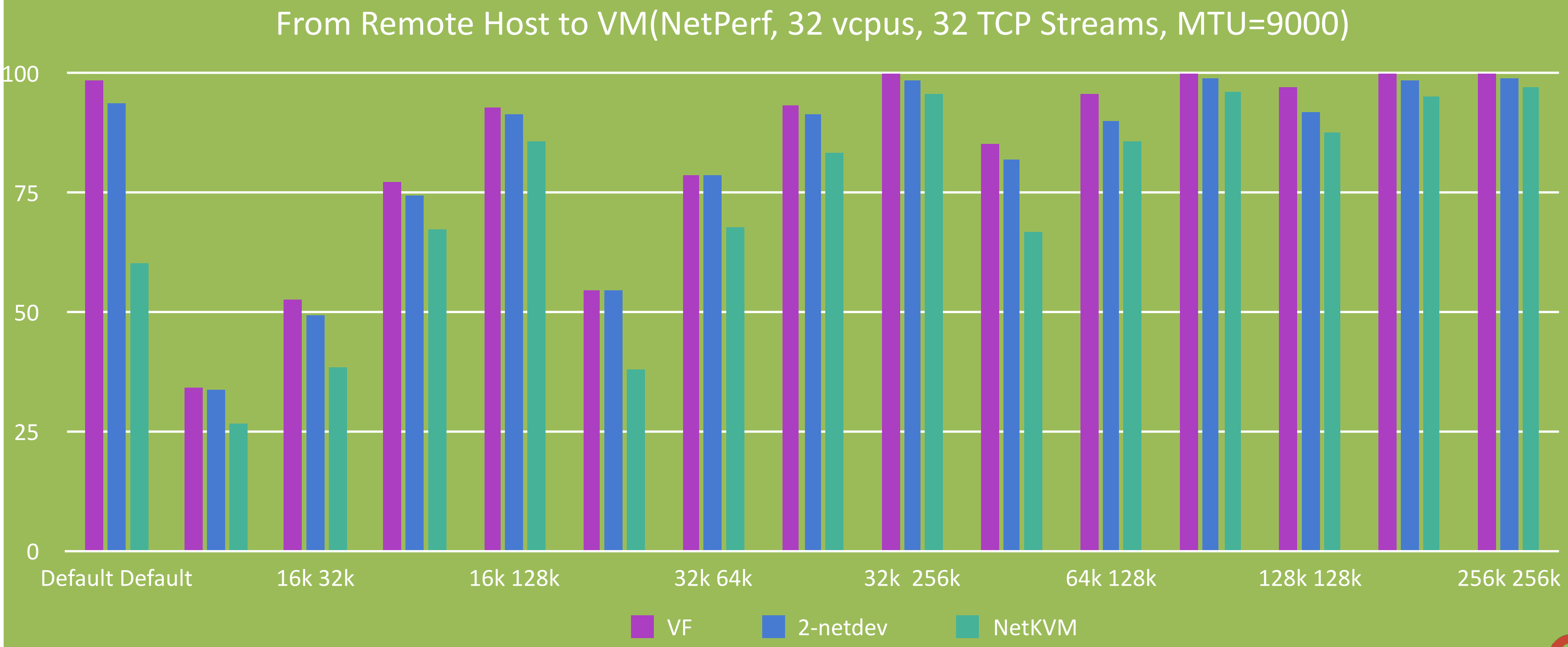
- After
  - Two steps certification
  - Automatic review for miniport
  - Manual review of protocol driver and the final package

# Performance



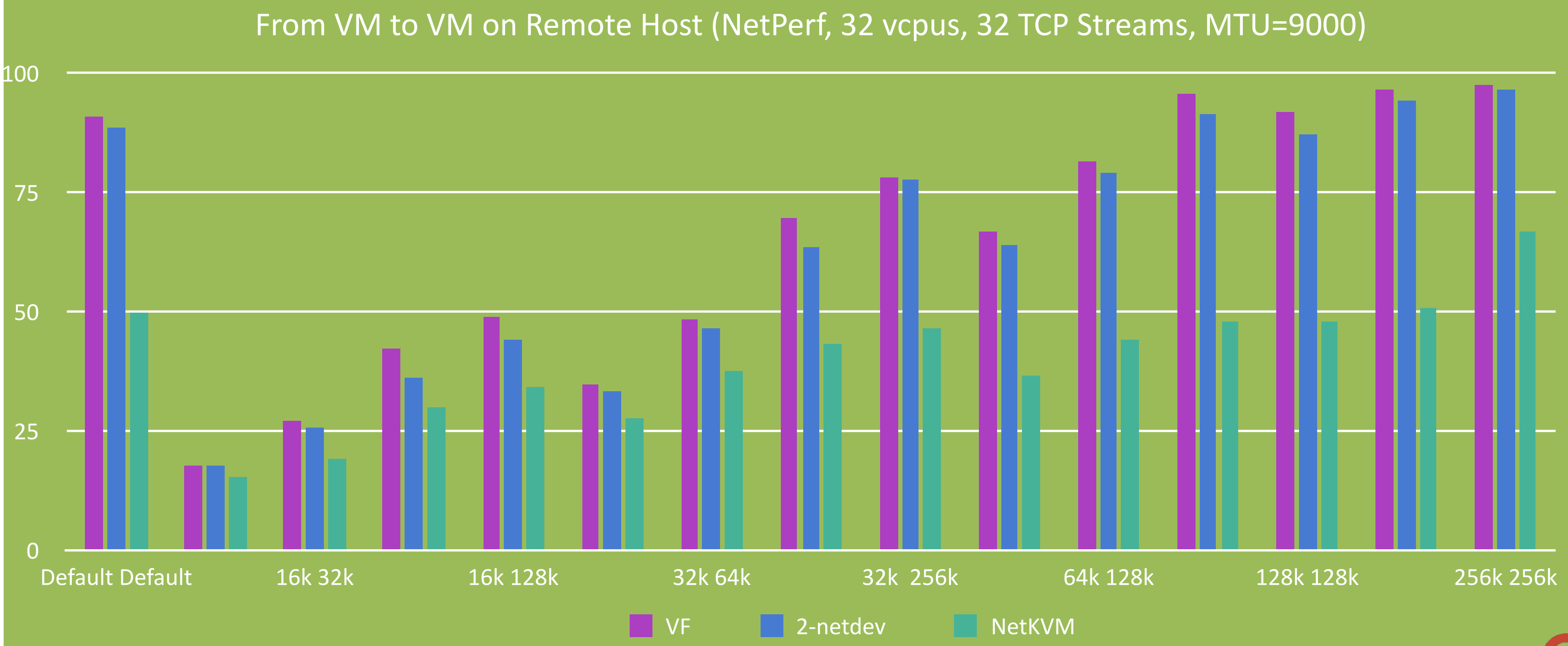
- x axis: “TX/RX Buffer size” and “TX/RX Socket Buffer size”
- y axis: Network throughput from 0 to 100G

# Performance



- x axis: “TX/RX Buffer size” and “TX/RX Socket Buffer size”
- y axis: Network throughput from 0 to 100G

# Performance



- x axis: “TX/RX Buffer size” and “TX/RX Socket Buffer size”
- y axis: Network throughput from 0 to 100G

annie.li@oracle.com

yan@daynix.com

# Links – source code

- virtio-win drivers source code:
  - https://github.com/virtio-win/kvm-guest-drivers-windows
- MiniPort and Protocol driver:
  - https://github.com/virtio-win/kvm-guest-drivers-windows/tree/master/NetKVM
- Notification object:
  - https://github.com/virtio-win/kvm-guest-drivers-windows/tree/master/NetKVM/NotifyObject

# Links – download binary drivers

https://docs.fedoraproject.org/en-US/quick-docs/creating-windows-virtual-machines-using-virtio-drivers/index.html

# Links – related presentations

- KVM Forum 2015 Live Migration with SR-IOV Pass-through - Weidong Han, Huawei
  - https://www.youtube.com/watch?v=vnwEnzVp9Zo
  - https://www.linux-kvm.org/images/9/9a/03x07-Juniper-Weidong_Han-LiveMigrationWithSR-IOVPass-through.pdf
- KVM Forum 2018 - Live Migration Support for GPU with SR-IOV - Zheng Xiao, Alibaba Cloud; Jerry Jiang & Ken Xue, AMD
  - https://events19.linuxfoundation.org/wp-content/uploads/2017/12/Live-Migration-Support-for-GPU-with-SRIOV-Challenges-and-Solution-Zheng-Xiao-Alibaba-Cloud-Jerry-Jiang-Ken-Xue-AMD.pdf
- KVM Forum 2020 (parallel session) - Device Keepalive State for Local Live Migration and VMM Fast Restart - Jason Zeng, Intel
  - https://sched.co/eE3W

KVM

FORUM