



OCP SUMMIT

March 20-21
2018
San Jose, CA

OPEN. FOR BUSINESS.



SONiC - Programmability, Extensibility and Beyond

David A. Maltz

Distinguished Engineer

Microsoft Azure Networking

OPEN. FOR BUSINESS.



Application & Management tools



CANONICAL



ARISTA



Tencent 腾讯

SONiC [Software For Open Networking in the Cloud]



Inventec



Quanta Computer



Celestica™



SAI [Switch Abstraction Interface]



nephos

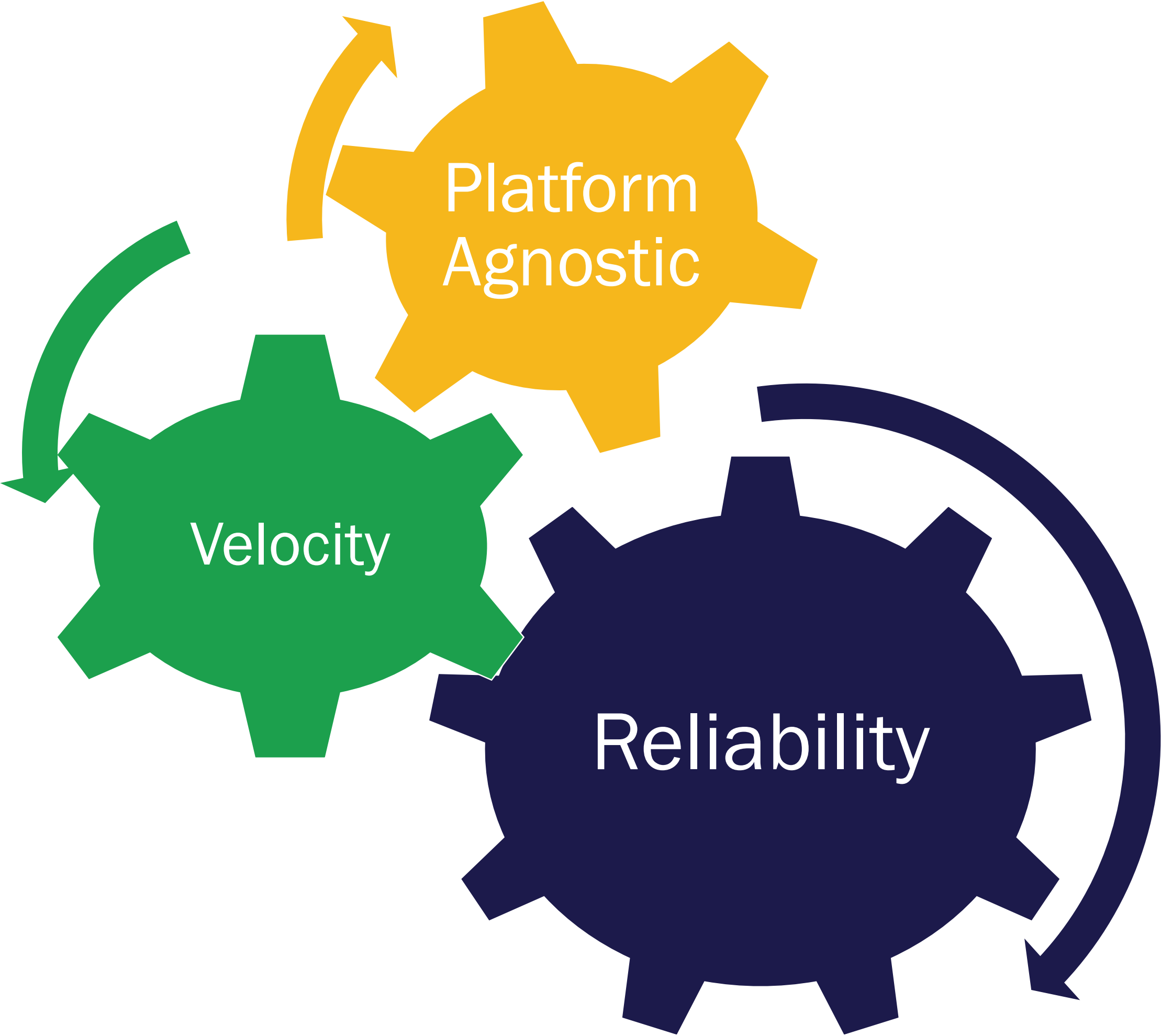
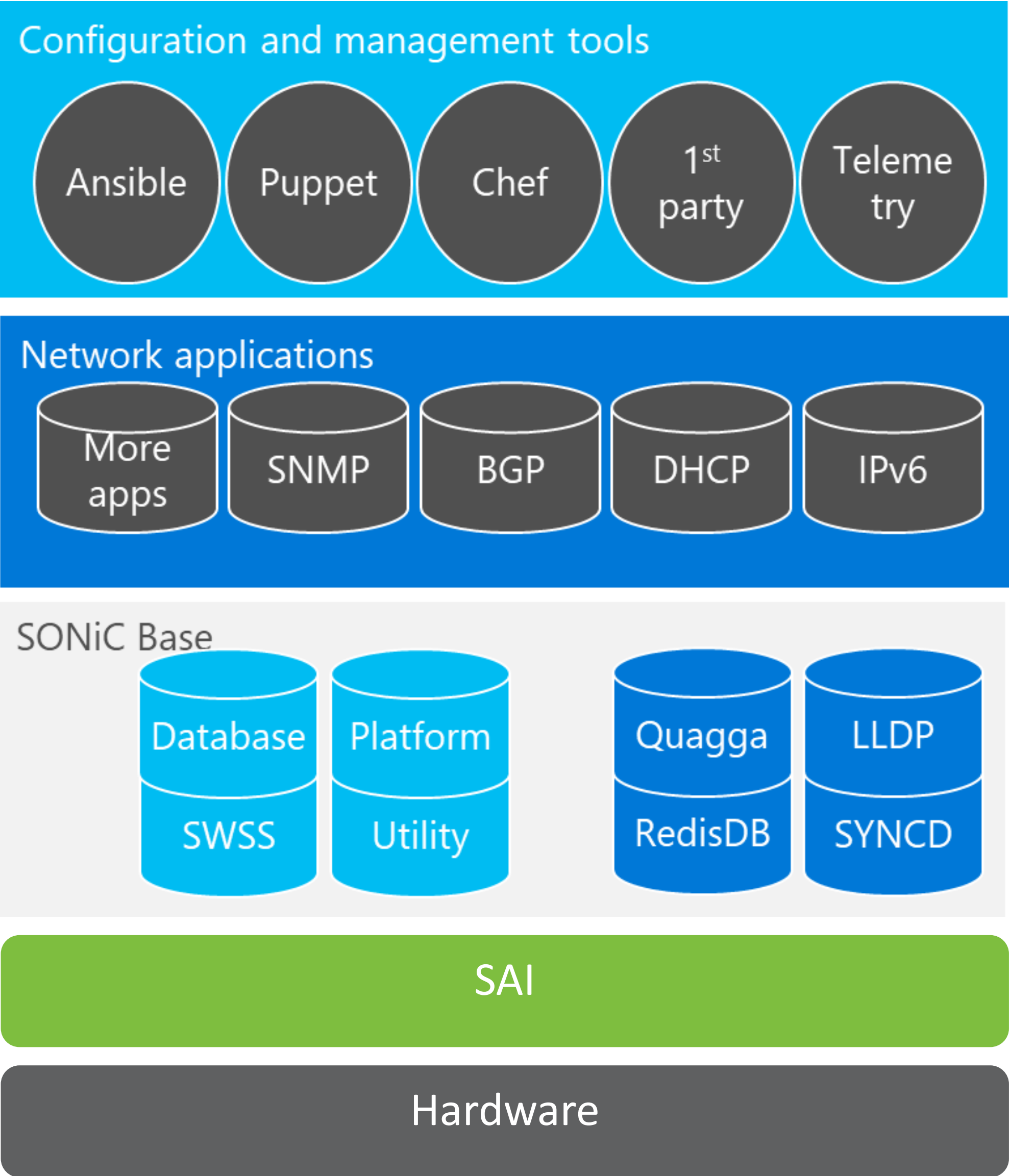


Innovium™

Switch

Silicon/ASIC

SONiC Recap – Containerized Open Source NOS



What Is New -

- SONiC supports Open Optical Monitoring (OOM)
- Richer features, thanks to
 - Alibaba: Vlan Trunk, TACACS, etc. and leading streaming telemetry work
 - LinkedIn: leading FRR integration, BGP convergence and Open 19
 - Tencent: leading VRF work
 - Mellanox: leading RDMA work
- Richer classes of devices
 - Arista: modular chassis
 - Marvell: ARM-based switch
- Richer scenarios via programmability
 - SONiC Network Virtualization

New Challenges from Broad Spectrum of Workload

Microsoft and NetApp Unleash the Power of Data Through the Industry's

Cray is bringing its supercomputers to Microsoft Azure

Fred

Get Ready for VMware Horizon Cloud on Microsoft Azure

Here's

bring

Microsoft Runs SAP HANA Enterprise Cloud on Azure

Microsoft will run SAP HANA Enterprise Cloud on [Microsoft Azure](#). This will allow customers to run SAP S/4HANA in a secure, managed [cloud](#).

Related Articles



Microsoft

That's a lot of eye-crossing product names, so we asked SAP for some clarification. A spokesperson explained:

ement



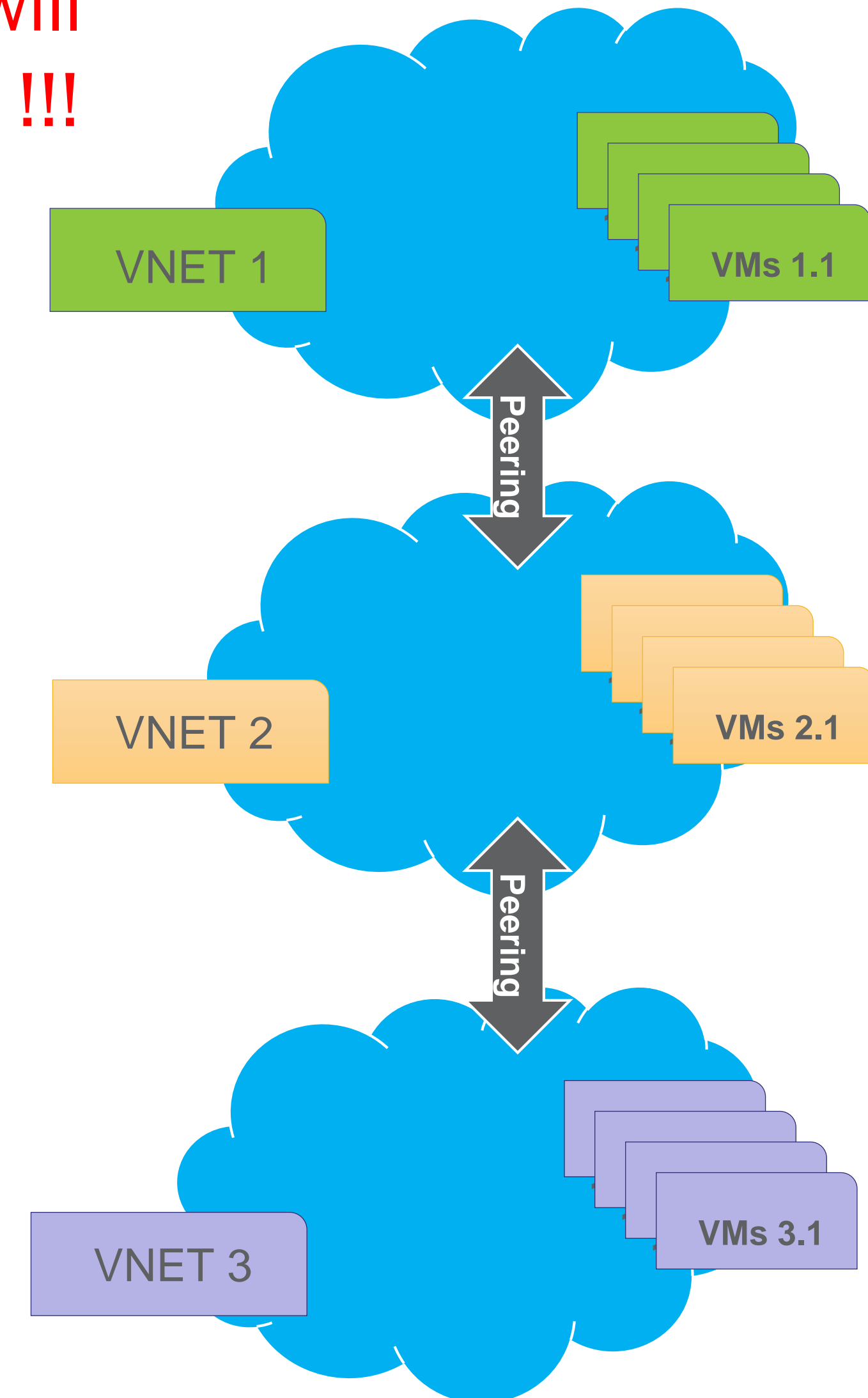
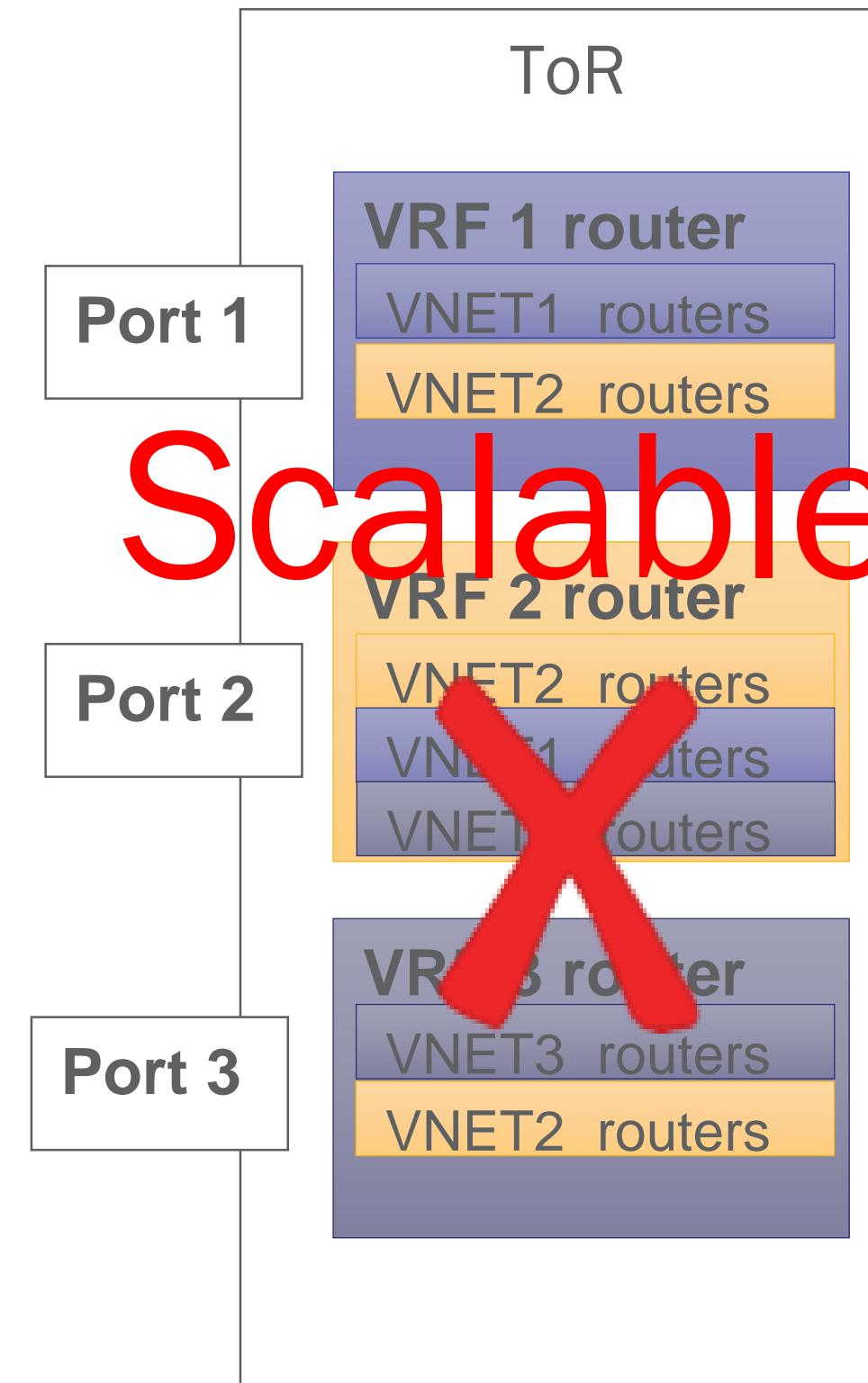
Case Study - VNET Peering in Legacy Network

Traditional Implementation

- VNET represented by VRF
- VNET1 peering with VNET2 implies copy routes from VNET1 to VNET2 and vice versa
- 1K VMs and 100 VNETs will require up to 10M routes !!!



1K VMs and 100 VNETs will require up to 10M routes !!!

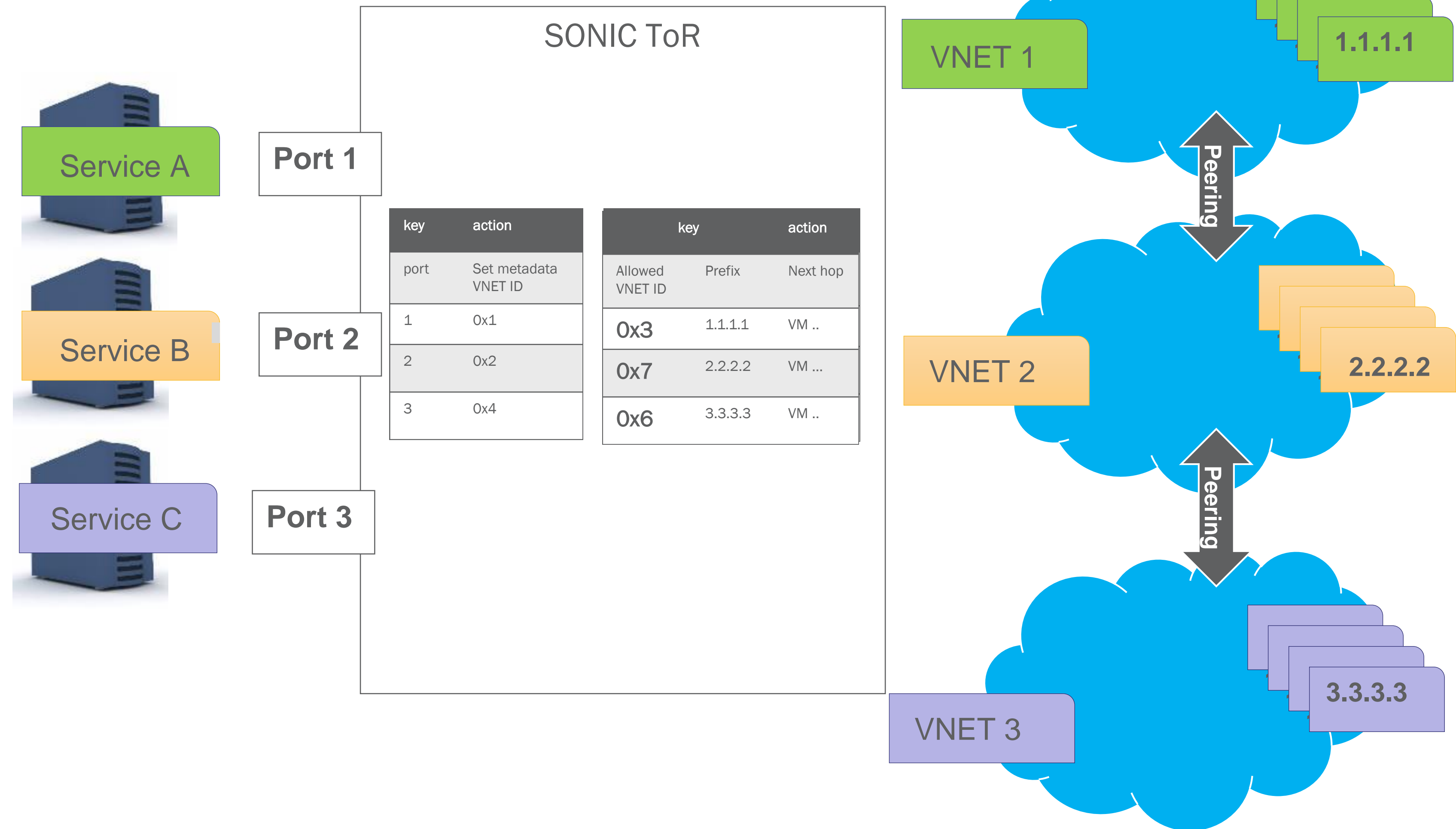


VNET Peering in Programmable Network

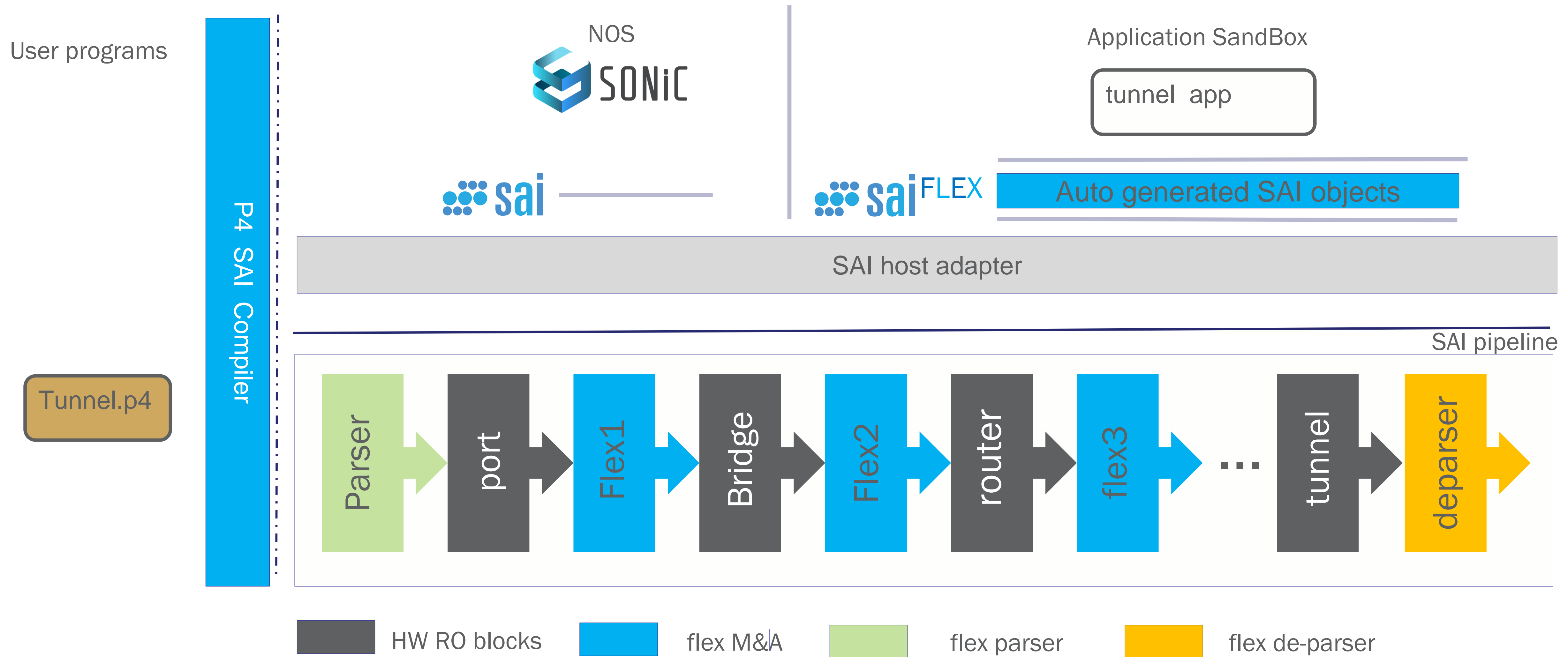
SONiC Implementation

- Two match action tables
- Port to VNET
 - Key: Port
 - Action Set metadata
 - Where metadata = VNET ID
- VNET routing
 - Key: metadata, prefix
 - Where metadata vector of VNET peers
 - Action: next hop
- VNET1 peering with VNET2 -> turn on VNET1 VNET ID in VNET routing metadata of all routes originated by VNET2
 - A single route per VM
 - VM update requires a single route update

1K VMs and 100 VNETs need only
100k routes

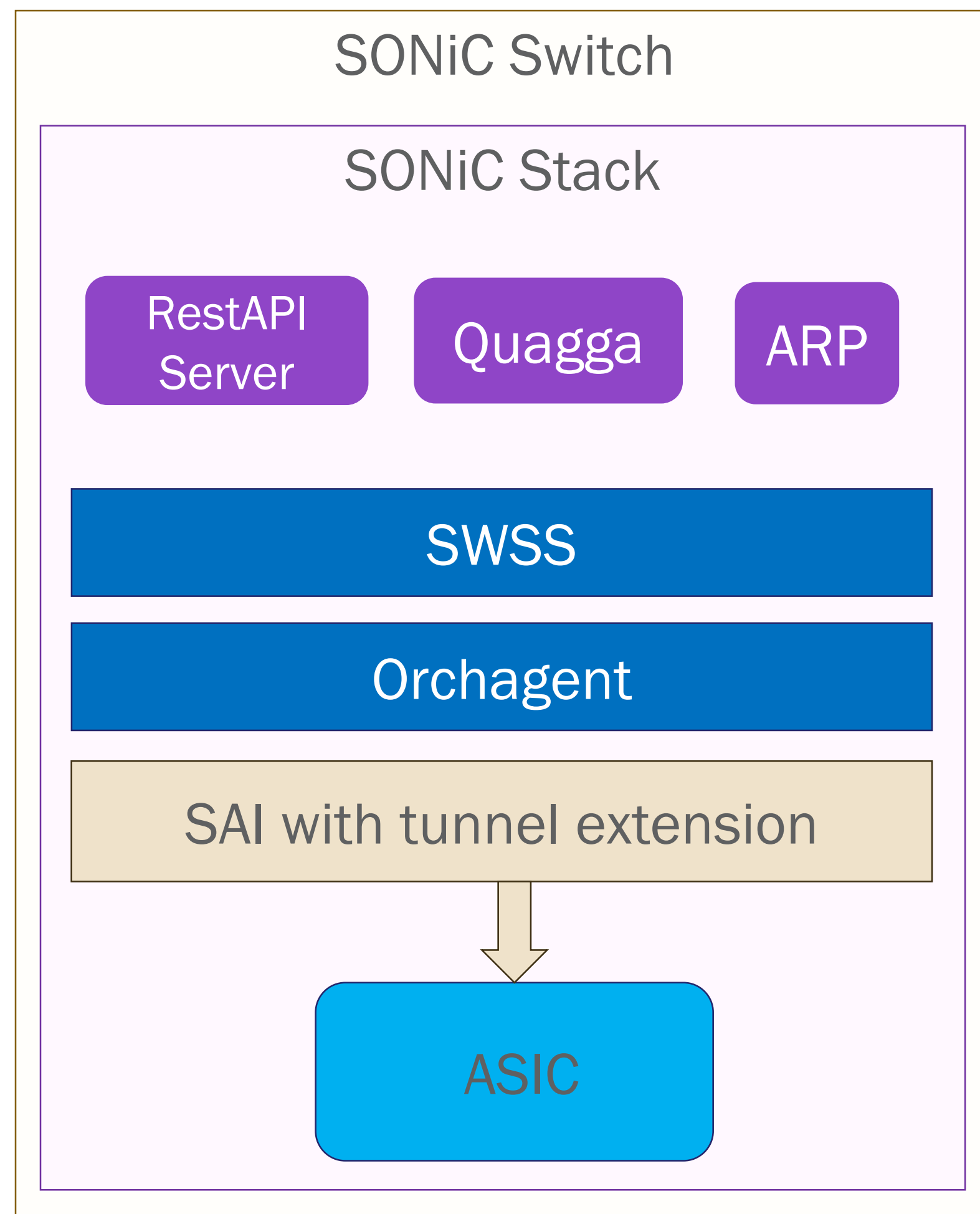


SAI Part: Programmable SAI API



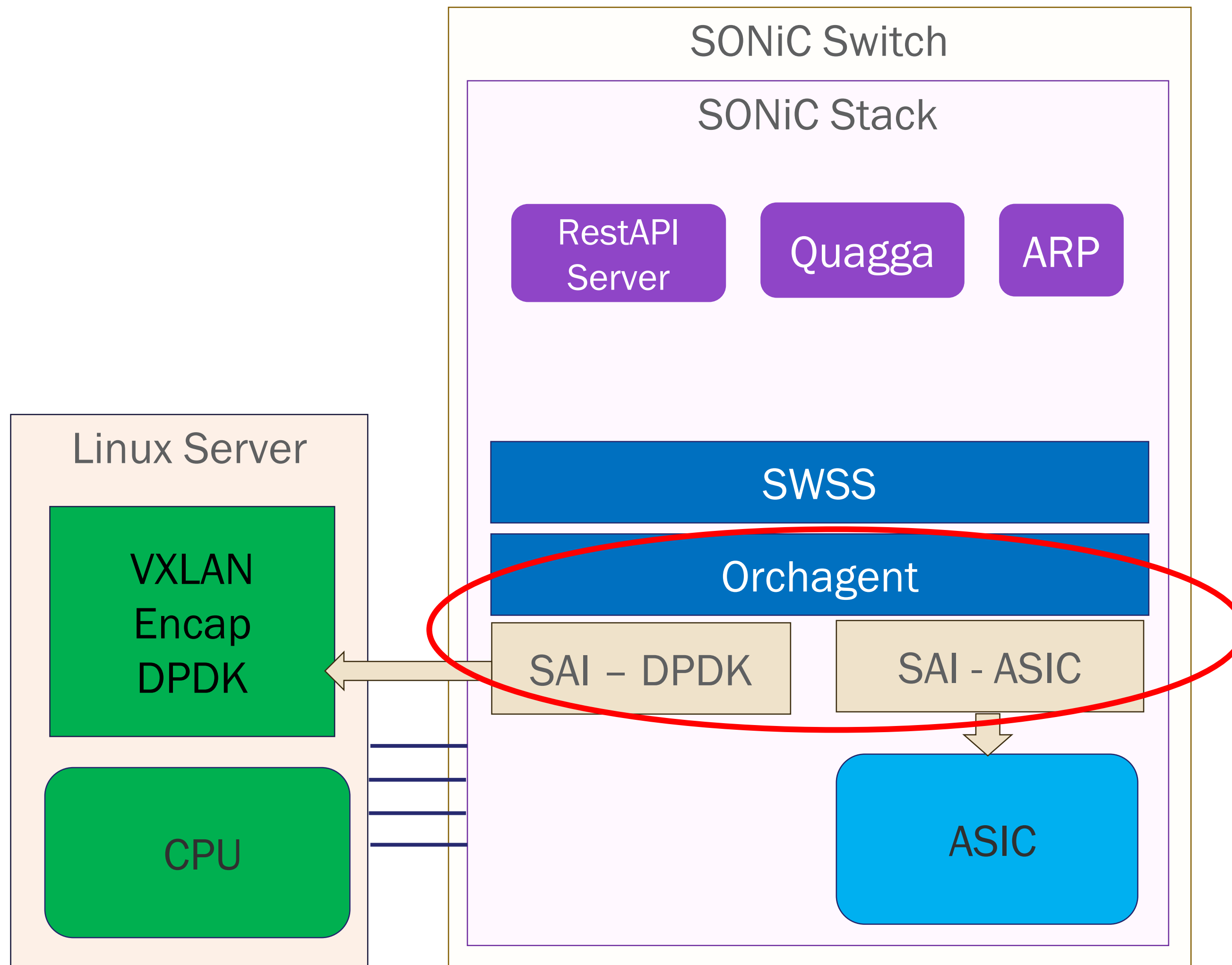
- Multiple switching SW options, develop apps
- SAIFlexAPI – uniform API for all programming language

SONiC Part: Supporting VNET



- RestAPI: Provide RestAPI for external
 - Allow external control to config the switch
 - Provides real-time data path counters and resource monitoring
 - Use OpenAPI specification (swagger)
- SWSS/Orchestration Agent
 - Use SAI with tunnel extension API
 - Provide tunnel support to upper applications

SONiC Part: Achieving Scalability via DPDK



- SWSS/Orchestration agent: manage multiple SAI instances
 - Manage tunnel entry cache between DPDK and ASIC
- Server: data plane scalability and programmability
 - 16M tunnels
 - 40G/100G line rate
 - 25 ~ 30 us forwarding latency
- ASIC: High port density and rich data plane functionality
 - Tunnel entry cache
 - Underlay routing
 - Traffic policing/shaping
 - ACL
 - Mirroring

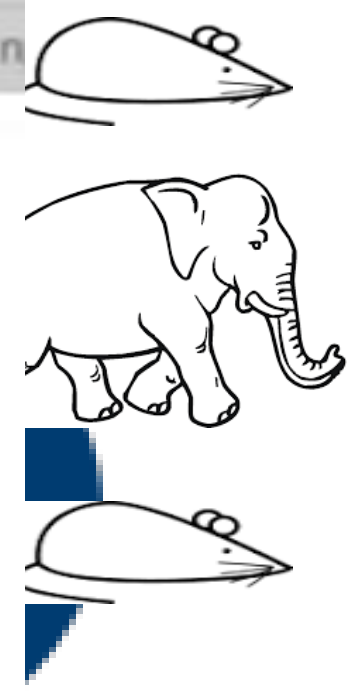
Demo: SONiC for Network Virtualization

```
chine View Input Devices Help
Updating... - Mozilla Firefox ~/sonic/SAI/inc/saibmtor.h ... Terminal - yonatanp@yonat...
Terminal - yonatanp@yonatanp-VirtualBox: ~/Desktop
Terminal - yonatanp@yonatanp-V... x admin@r-arch-sw07: ~ x admin@r-arch-sw07: ~ x yonatanp@yonatanp-V... x yonatanp@yonatanp-V... x yonatanp@yonatanp-V... x yonatanp@yonatanp-Vi... x yonatanp@yonatanp-V... x
9.00-10.00 sec 128 KBytes 1.05 Mbits/sec 0 53.2 KBytes
10.00-11.00 sec 128 KBytes 1.05 Mbits/sec 0 54.6 KBytes
11.00-12.00 sec 128 KBytes 1.05 Mbits/sec 0 56.0 KBytes
12.00-13.00 sec 128 KBytes 1.05 Mbits/sec 0 57.3 KBytes
13.00-14.00 sec 128 KBytes 1.05 Mbits/sec 0 58.7 KBytes
14.00-15.00 sec 128 KBytes 1.05 Mbits/sec 0 58.7 KBytes
15.00-16.00 sec 128 KBytes 1.05 Mbits/sec 0 60.1 KBytes
16.00-17.00 sec 128 KBytes 1.05 Mbits/sec 0 61.4 KBytes
17.00-18.00 sec 128 KBytes 1.05 Mbits/sec 0 91.5 KBytes
18.00-19.00 sec 128 KBytes 1.05 Mbits/sec 0 68.3 KBytes
19.00-20.00 sec 128 KBytes 1.05 Mbits/sec 0 68.3 KBytes
20.00-21.00 sec 128 KBytes 1.05 Mbits/sec 0 68.3 KBytes
-----
Interval Transfer Bitrate Retr
0.00-21.09 sec 2.62 MBytes 1.04 Mbits/sec 0 sender
0.00-21.09 sec 0.00 Bytes 0.00 bits/sec receiver
interrupt - the client has terminated
v-r-vrt-234-009 ~]#
v-r-vrt-234-009 ~]# sleep 45; iperf3 -c 192.168.3.2 -b1m -t45

192.168.3.2 ping statistics ---
64 packets transmitted, 1 received, 0% packet loss, time 0ms
avg/max/mdev = 0.700/0.700/0.700/0.000 ms
v-r-vrt-234-009 ~]# watch ./ping_demo.sh
v-r-vrt-234-009 ~]#
v-r-vrt-234-009 ~]# watch ./ping_demo.sh
v-r-vrt-234-009 ~]#
v-r-vrt-234-009 ~]#
v-r-vrt-234-009 ~]# sleep 5; watch -n0.1 ./ping_demo.sh
v-r-vrt-234-009 ~]#
v-r-vrt-234-009 ~]#
v-r-vrt-234-009 ~]#
v-r-vrt-234-009 ~]# sleep 5; watch -n0.1 ./ping_demo.sh
v-r-vrt-234-009 ~]#
v-r-vrt-234-009 ~]#
v-r-vrt-234-009 ~]# sleep 5; watch -n0.1 ./ping_demo.sh
v-r-vrt-234-009 ~]#
v-r-vrt-234-009 ~]#
v-r-vrt-234-009 ~]# watch -n0.1 ./ping_demo.sh g
v-r-vrt-234-009 ~]#

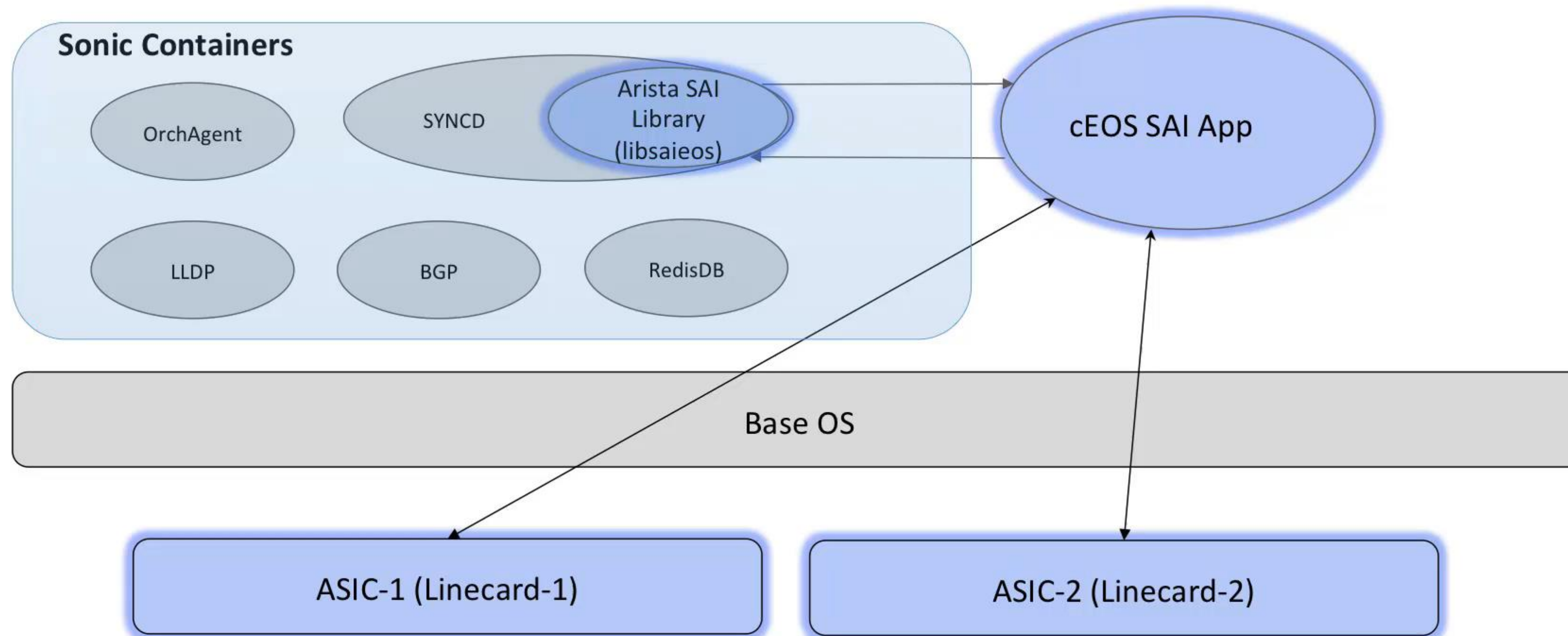
alanlo bshow yp.py dpdkshow.py omers orchagent.adaptive_merge
bshow.py bmt debug marianp orchagent yonatanp
admin@r-arch-sw07:~$ docker cp bmt_debug swss:/bm/.
admin@r-arch-sw07:~$ docker exec -it swss bash
root@r-arch-sw07:/# ls
bin lib proc srv tmp
bm lib64 restapi stam usr
boot libsaimeta.so.0.0.0 root stam2 var
dev media run swss.rec
etc mnt sairedis.rec swss_1.0.0_amd64.deb
home opt sbin swss_debs
host orchagent sonic-rest-api_1.0.0_amd64.deb sys
root@r-arch-sw07:/# cd b
bin/ bm/ boot/
root@r-arch-sw07:/# cd bm/
root@r-arch-sw07:/bm# ls
bmtor sai orch bmt debug orchagent
root@r-arch-sw07:/bm# ./bmt_debug
BMT debug
Debug BMT client connecting on port 50505
BMT>
??? Try - status, flush, pause, resume, window, ithresh, ethresh, evac-stop, insert-stop
BMT> resume
Insertion resumed
BMT> exit
root@r-arch-sw07:/bm#
root@r-arch-sw07:/bm#
root@r-arch-sw07:/bm# exit
exit
admin@r-arch-sw07:~$
admin@r-arch-sw07:~$
admin@r-arch-sw07:~$ ls
alanlo bshow yp.py dpdkshow.py omers orchagent.adaptive_merge
bshow.py bmt debug marianp orchagent yonatanp
admin@r-arch-sw07:~$ watch -n1 ./bmslow_yp.py
admin@r-arch-sw07:~$ watch -n1 ./bmslow_yp.py
admin@r-arch-sw07:~$ watch -n1 ./bmslow_yp.py
admin@r-arch-sw07:~$ watch -n1 ./bmslow_yp.py
admin@r-arch-sw07:~$
```

DE



Demo: SONiC on Chassis Switch

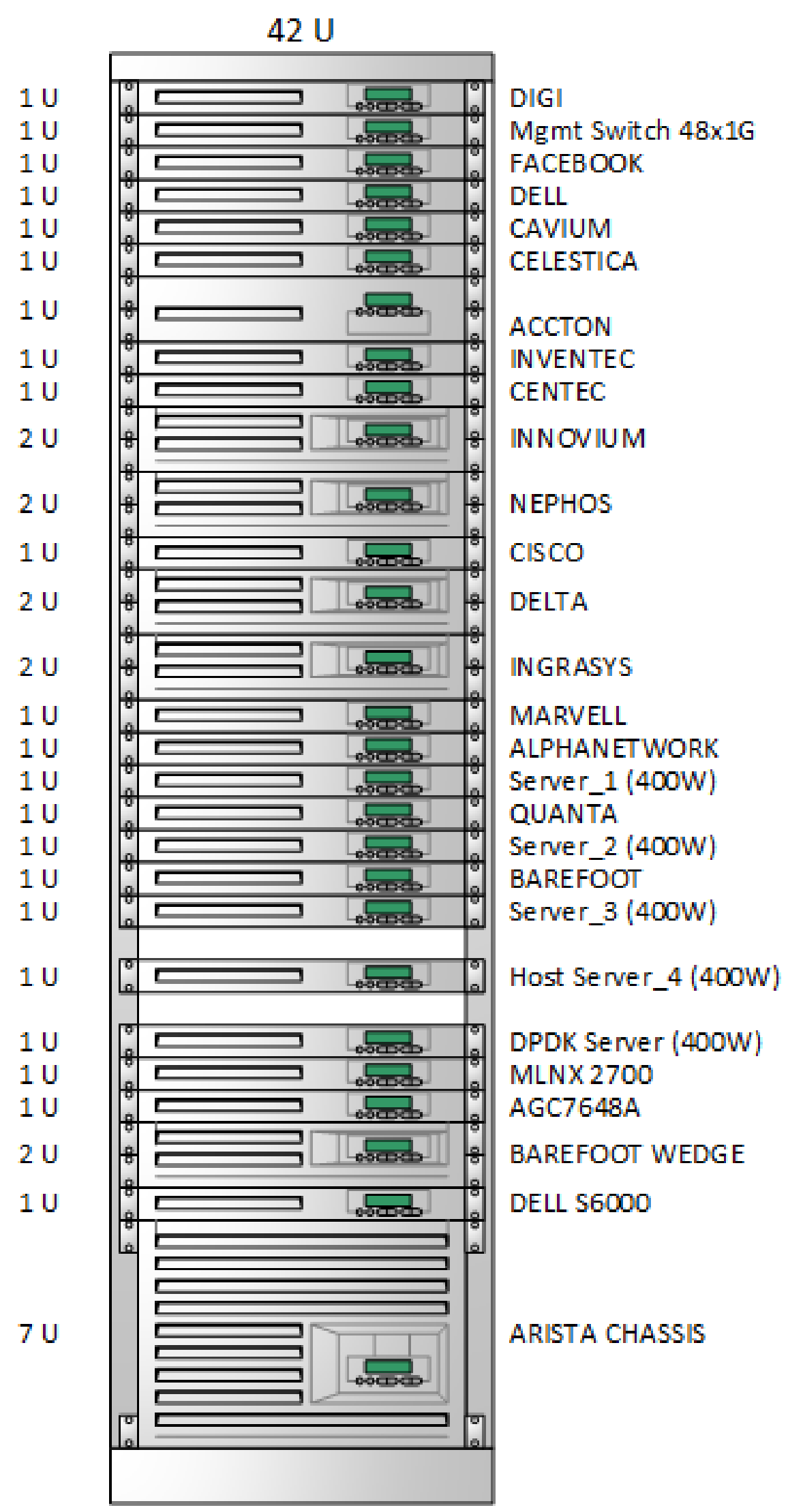
cEOS SAI App for Managing Modular Chassis



Moving Forward: Enabling WAN Scenarios

- Global network is growing exponentially, requires
 - Agility for fast Time to Market feature release and defect remediation
 - To minimize hardware dependencies
 - To scale and grow the WAN efficiently while controlling costs
- Sonic is an integral element of our cloud SDN solutions for intelligent traffic management
- Two major roles
 - Edge Peering Router
 - Backbone Router

More Demos in Microsoft (A11) and Partner Booths



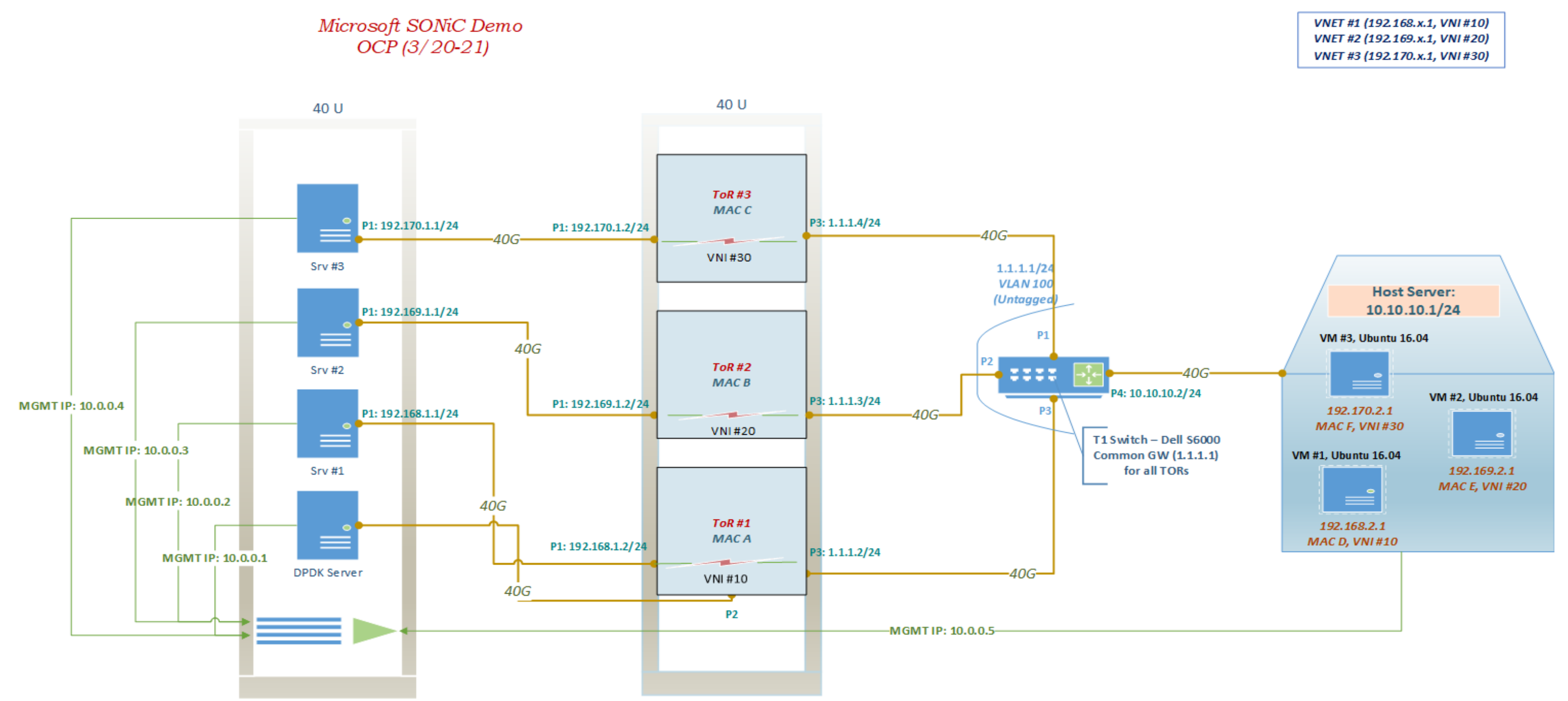
New ASIC Supported

Pizza box and Chassis

Rich Hardware Platform Supported

Virtualization

Programmability



Open Invitation

- [OCP SONiC/SAI workshop on 3/22](#)
- Inviting contributions in all areas
 - SONiC/SAI
 - Hardware platform
 - New features, applications and tools
 - Download it, test it and use it!

Website: <https://azure.github.io/SONiC/>

Mailing list: sonicproject@googlegroups.com

Source code: <https://github.com/Azure/SONiC/blob/gh-pages/sourcecode.md>

Wiki: <https://github.com/Azure/SONiC/wiki/>



OCP SUMMIT