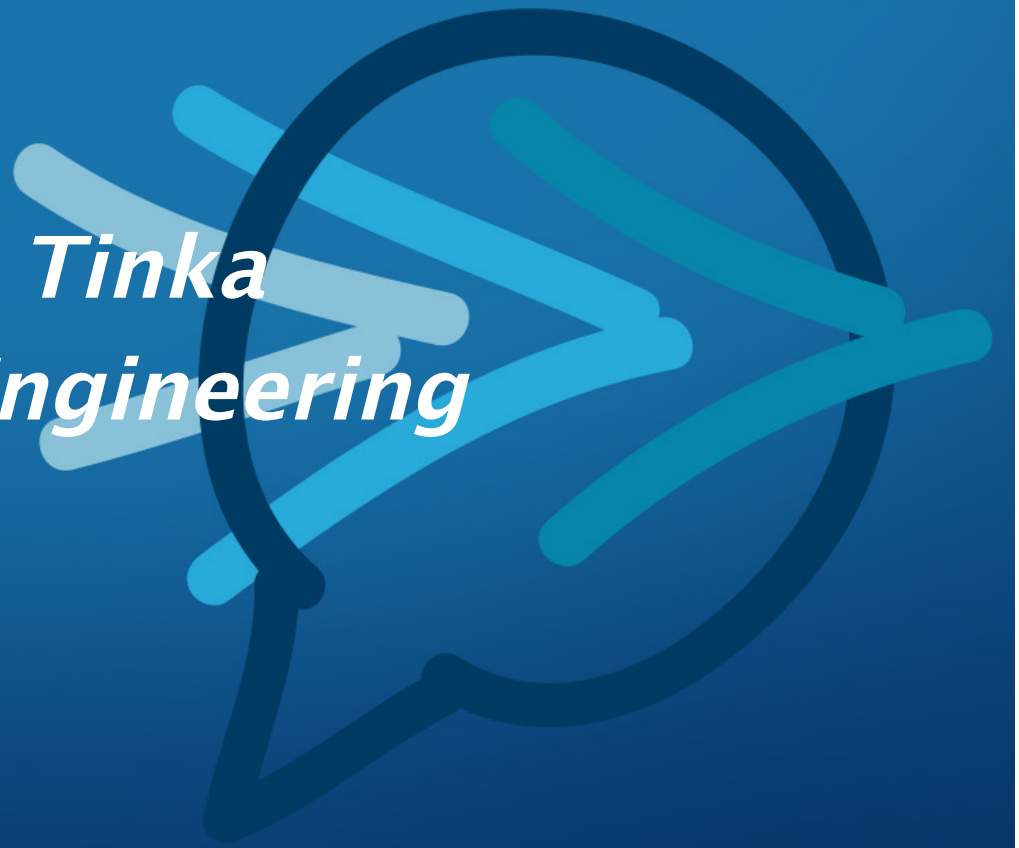
21st Century iBGP Route Reflection

Mark Tinka
Head of Engineering



SEACOM

Disclaimer

This presentation, in no way, endorses or promotes the Cisco solution that it describes.

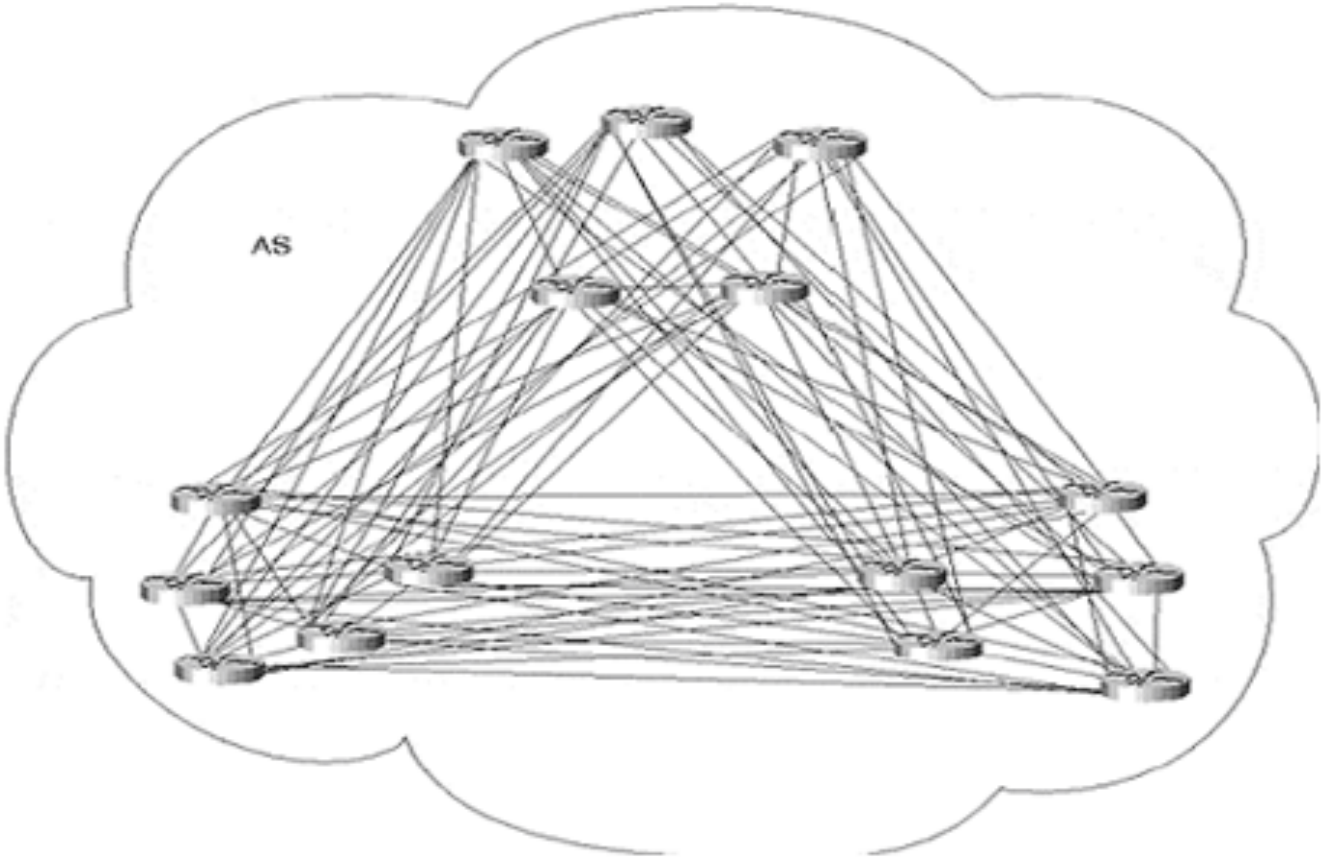
The Cisco infrastructure is merely a tool for the job.

- iBGP Route Reflection Fundamentals
- Previous (And Still Relevant) Route Reflection Models
- The Growing Trend In Route Reflection
- SEACOM's Route Reflection Deployment
- Motivations For SEACOM
- Why Out-Of-Path
- Deployment Nuances
- Operations
- What The Future Holds

iBGP Route Reflection Fundamentals

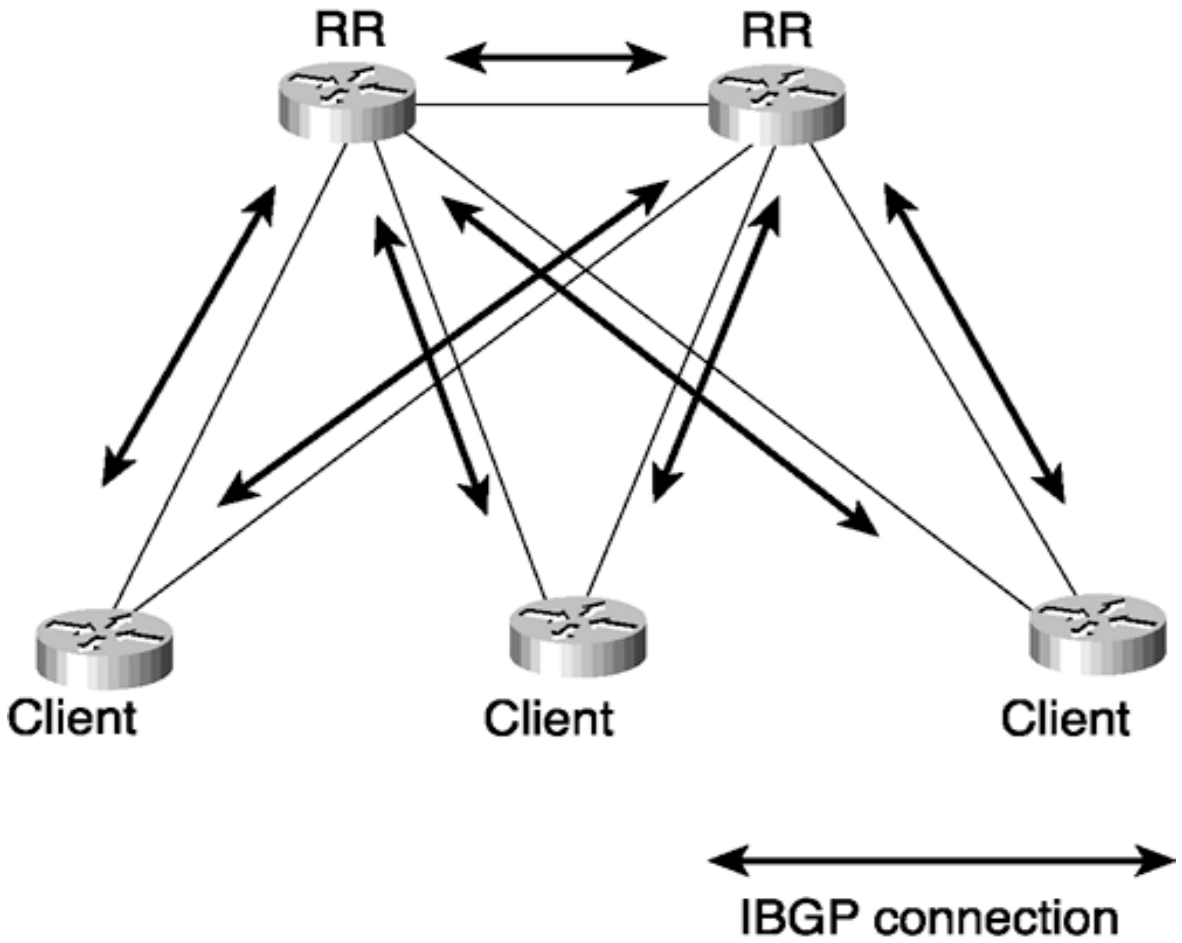
iBGP Route Reflection Fundamentals

- So here is the problem:



iBGP Route Reflection Fundamentals

- And to fix that, here is the most typical solution:



Previous (And Still Relevant) Route Reflection Models

Previous (And Still Relevant) Route Reflection Models

- Use routers for the route reflection.
- Either your existing core routers (in-path).
- Or dedicated routers (out-of-path).

Previous (And Still Relevant) Route Reflection Models

The Growing Trend In Route Reflection

The Growing Trend In Route Reflection

- A number of changes have been taking place in this space:
    - The need for more control plane memory.
    - The need for smaller footprint devices.
    - The need for out-of-path topologies.
    - The need to decouple RIB from FIB (speeds up convergence).
    - The need to leverage commodity hardware.
    - The need for virtualization.
    - The need for innovation through software.
    - The need to nullify hardware limitations.

The Growing Trend In Route Reflection

- Software routers – the REAL ones this time ☺.

- Vendors have been working hard at delivering their code.
- But packaged in software instead of hardware.
- You get all the features, and none of the weight.

- A lot has been inspired by the current industry buzzwords:
    - SDN
    - NFV

- Route reflectors was never really goal. Was just an obvious application.

The Growing Trend In Route Reflection

- Products that I know about:

- Cisco:
  - CSR1000v
  - IOSv
  - XRv
  - NX-OSv

- Juniper:
  - vRR
  - vMX

- Brocade:
  - Vyatta 5600 vRouter

- Alcatel-Lucent:
  - vSR-RR

The Growing Trend In Route Reflection

- Implementation concept is simple:
    - Runs on commodity x86 hardware.
    - Runs as a VM image.
    - VMware ESXi, KVM, Citrix XenServer, Microsoft Hyper-V, e.t.c.

SEACOM's Route Reflection Deployment

SEACOM's Route Reflection Deployment

- SEACOM have implemented Cisco's CSR1000v technology.

- CSR1000v is, essentially, IOS XE as known from the ASR1000 platform.
- Minus all the ASR1000 hardware.
- But with all the IOS XE software features and capabilities.

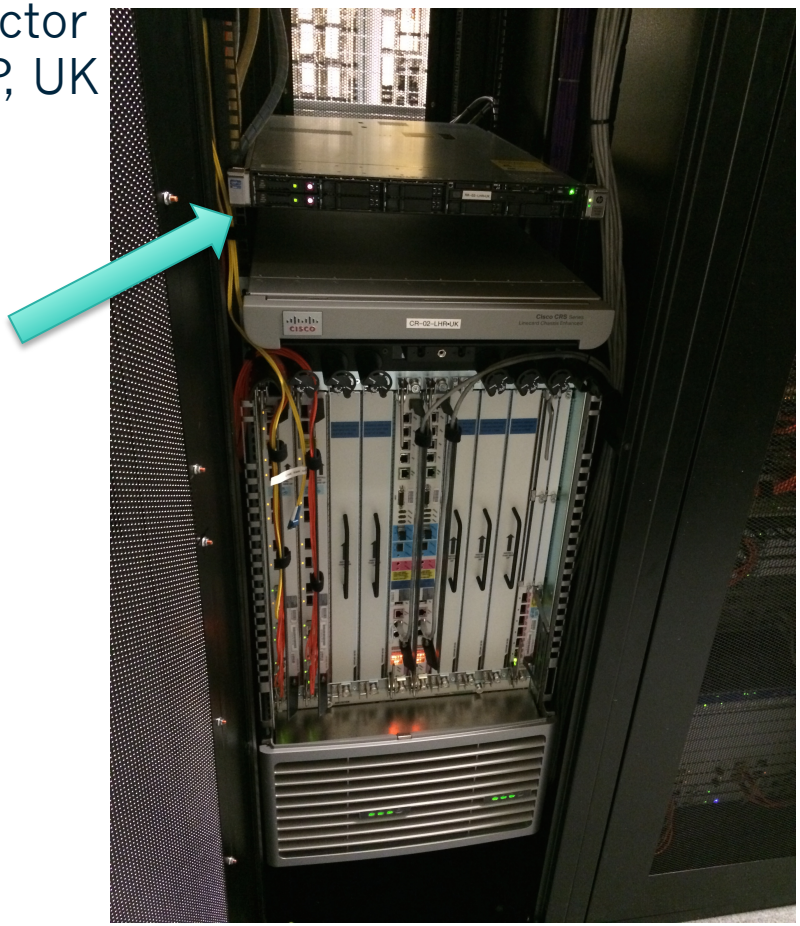
- SEACOM's purpose:
    - Dedicated, out-of-path route reflectors.

SEACOM's Route Reflection Deployment

Route Reflector
Marseille PoP, France

SEACOM's Route Reflection Deployment

Route Reflector
London PoP, UK



So Yes!

They Are Very Real ☺

Motivations For SEACOM

Motivations For SEACOM

- Why did we go with the CSR1000v approach:
    - First and foremost, we wanted scalable hardware.
    - But also, hardware that occupied a very small footprint.
    - We like IOS XE, even though the policy language is “dodgy” ☺.
- The kit:
    - HP ProLiant DL360p Gen8 1U servers.
    - 2x 6-core 2.6GHz E5-2630v2 64-bit CPU’s.
    - 512GB DRAM (not the maximum).
    - 2x 600GB hard drives.
    - 4-port 1Gbps Ethernet card.
    - VMware ESXi 5.5
    - VMware vSphere Client
    - Cisco CSR1000v software image.

Why We Chose Cisco

- We are primarily a dual-vendor house – Cisco & Juniper.
- In 2014, Juniper were talking about vRR, but had no “mature” code.
- vRR eventually morphed into what we know today as vMX.
- Cisco had already started shipping CSR1000v in 2014.

- We gave Juniper a chance to catch up, but they were too slow.
- vMX has only really hit the streets in 2016.

- Other vendors had started shipping some code in 2014.
- Namely, Alcatel-Lucent (now Nokia) and Brocade (Vyatta).
- They were not a stable within our network, so no consideration.

- Why the CSR1000v and not XRv?
    - IOS XR RPL is” verbose” (a lot like Junos).
    - IOS and IOS XE route maps are “cluttered”, but there is order.

Why Out-Of-Path

Why Out-Of-Path

- SEACOM run out-of-path route reflectors for the following reasons:
  - Different vendor equipment between large and small PoP's.
  - Reduction of BGP memory footprint in the MPLS core (BGP-free).
  - Reduction of control plane CPU footprint in the core.
  - Remove impact the core has on routing convergence.
  - Drastically reduce day-to-day operational contact with the core.
  - Remove impact of backbone flaps on route reflector CPU.
  - Run the latest routing features without impacting core forwarding.
  - Keep maintenance to the core at an absolute minimum.
  - Avoid unnecessary "brain" upgrades in the core to support routing.

- The trick with out-of-path route reflectors, however:
  - Route reflectors make routing decisions for clients.
  - So best to have local route reflectors within the PoP…
  - … or as close to the PoP as possible if funds are an issue.
  - … or implement "Add-Paths"/"Diverse-Paths".
  - … or BGP-ORR (draft-ietf-idr-bgp-optimal-route-reflection-10)

Deployment Nuances

Deployment Nuances

- CSR1000v hypervisor support as of IOS XE 3.17S and Denali 16.2:
  - VMware ESXi
  - Citrix XenServer (not supported on Denali 16.2)
  - KVM
  - Microsoft Hyper-V (not supported on Denali 16.2)
  - Amazon Machine Image on AWS (Amazon Web Services)

- We chose VMware ESXi.
- Is flagship support from Cisco, and is best tested.
- We run ESXi v6.0 (began with v5.5 in 2014).

- We run each instance as an independent island.
- Just as one would a real hardware router.

Deployment Nuances

- Cisco offer a multitude of installation options:
    - OVA template (Cisco recommended).
    - ISO image (which we preferred).
    - QCOW2 (QEMU Copy On Write, for KVM).
    - BIN file (used for post-installation upgrades).

- We preferred the ISO image:
    - Allows you to fully customize the installation.
    - But the Cisco-provided instructions are not the greatest.
    - So we did a lot of testing and developed our own guide.

Deployment Nuances

- Network deployment:
    - 2x ports configured for core connectivity.
    - 1x port configured to manage ESXi.
    - 1x port free.
    - iLO port configured for OoB.

    - For security, ESXi and iLO ports are inside the private network.
    - The 2x core-facing ports sit in the public domain.

Deployment Nuances

- ESXi installation:
  - Pretty straightforward.
  - Remember to enable SSH as you will need it to install CSR1000v.
  - Enabling IPv6 support is also recommended.

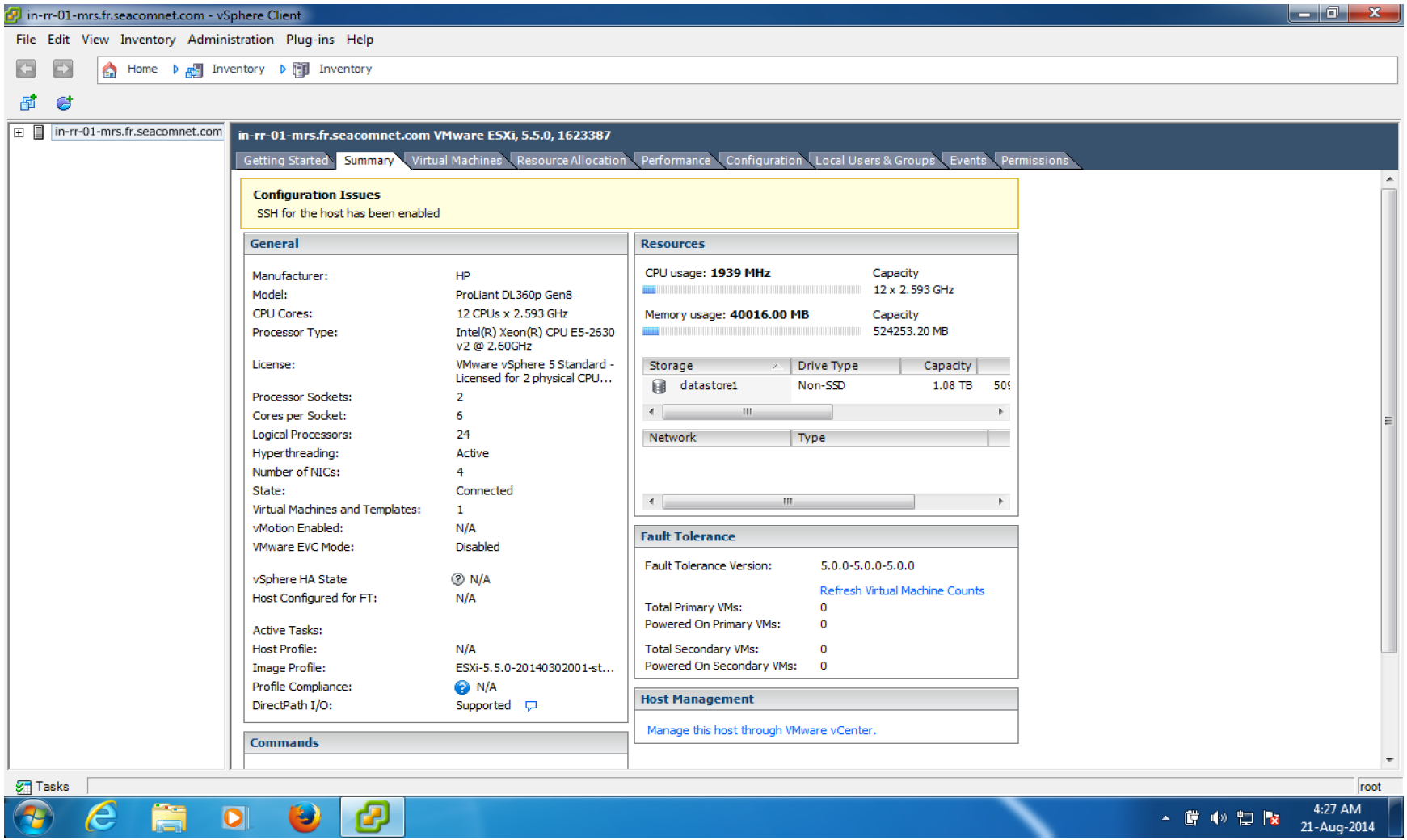
- CSR1000v installation:
  - Done via vSphere client (Windows-only support ☹).
  - Physical memory translates to hard drive space – make enough.
  - Do not allocate 100% of CPU to the VM – we did 75%.
  - Do not allocate 100% of RAM to the VM – we did 50%.
  - Reconfigure the default vSwitch (create new ones for each port).
  - Enable Jumbo frames for the VM – maximum is 9,000 bytes.
  - Disable power management to support high performance.
  - Installation of the VM can take up to 45x minutes – get coffee!

Deployment Nuances

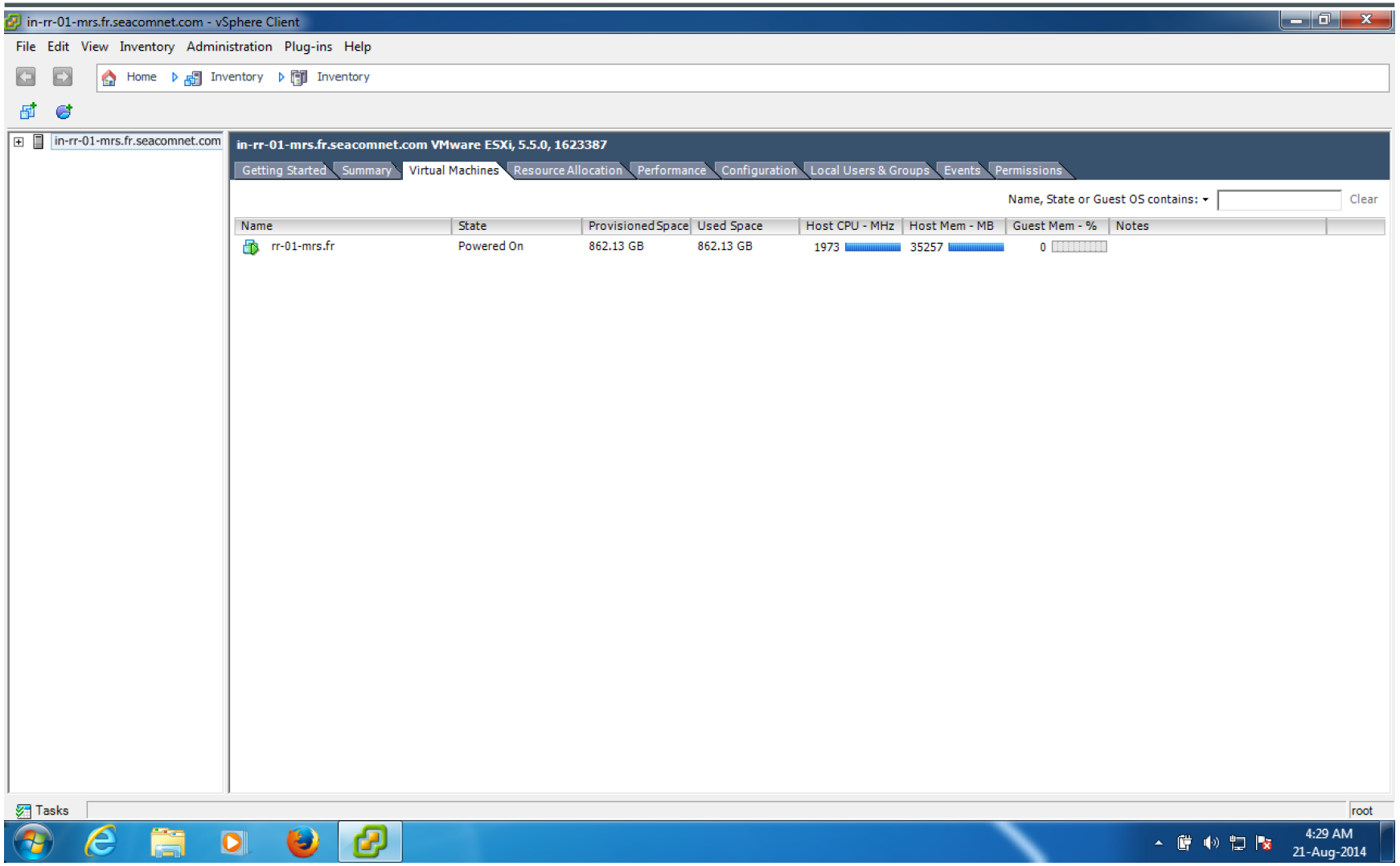
- Initial boot of the CSR1000v image:
    - CSR1000v first-time boot makes various installations.
    - It could take up to 20x minutes for the first boot.
    - Subsequent boots take no more than 5x minutes.

- Do not forget to license your copy of ESXi on each server.
- Or else it will expire after 60x days.

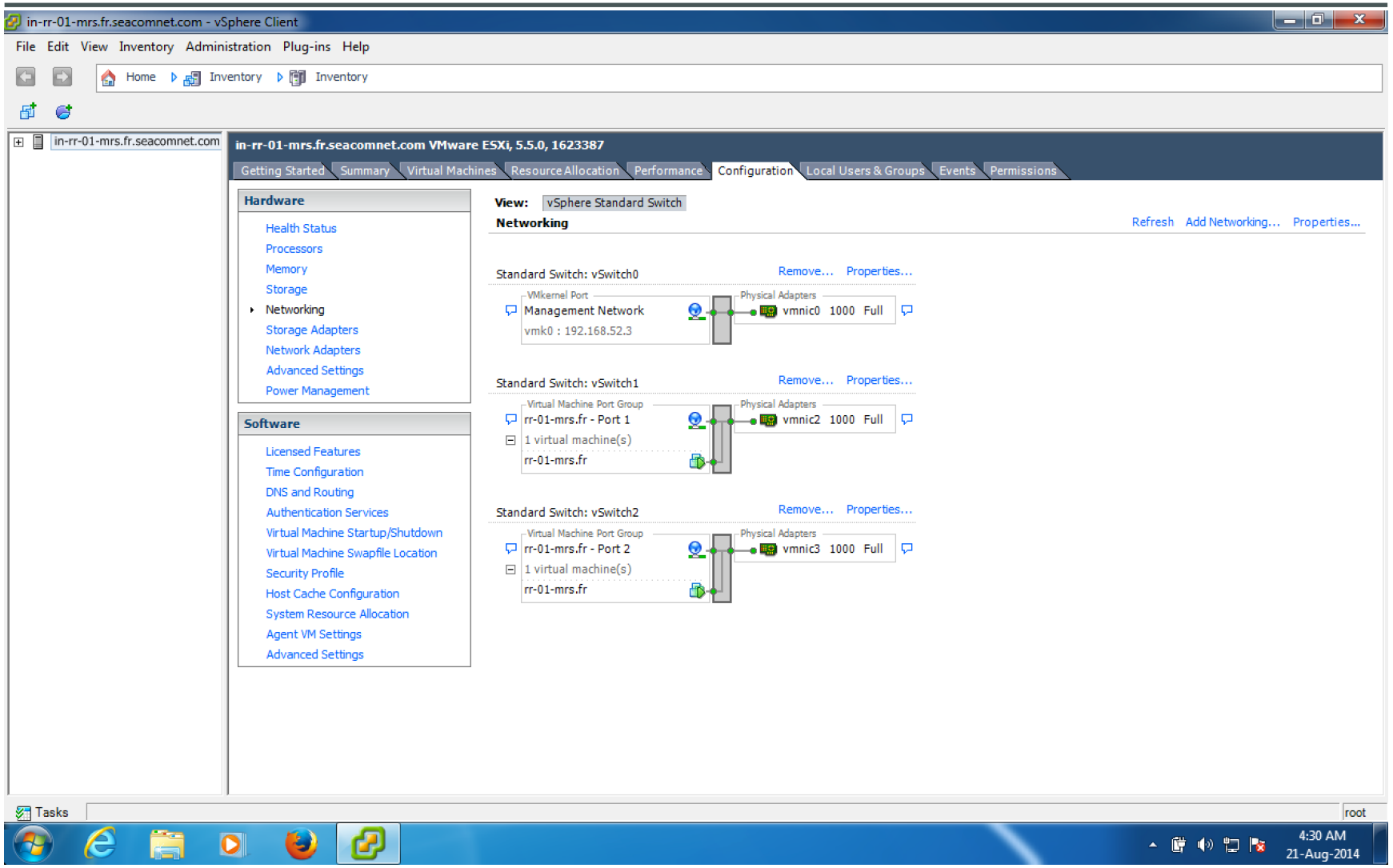
Deployment Nuances



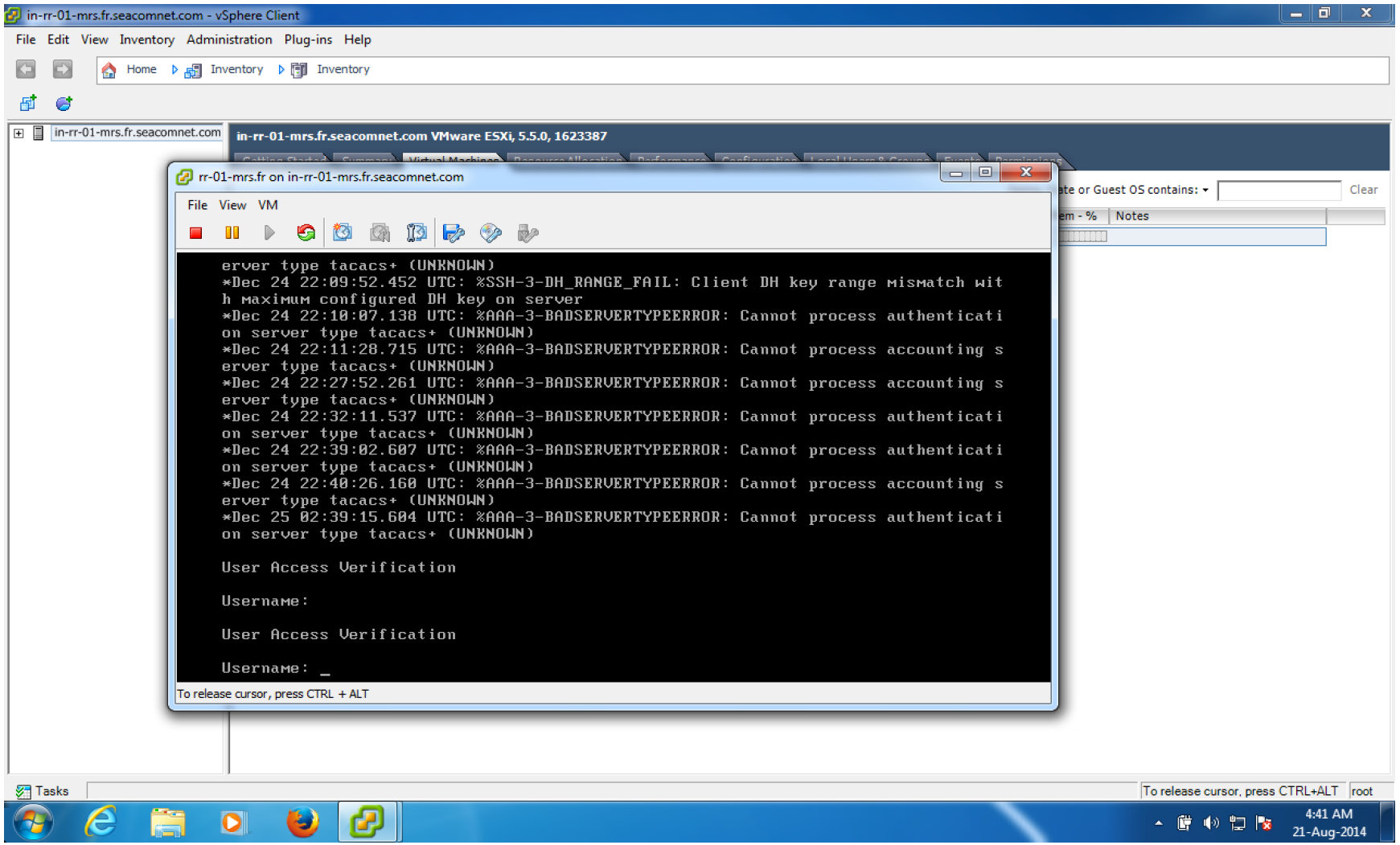
Deployment Nuances



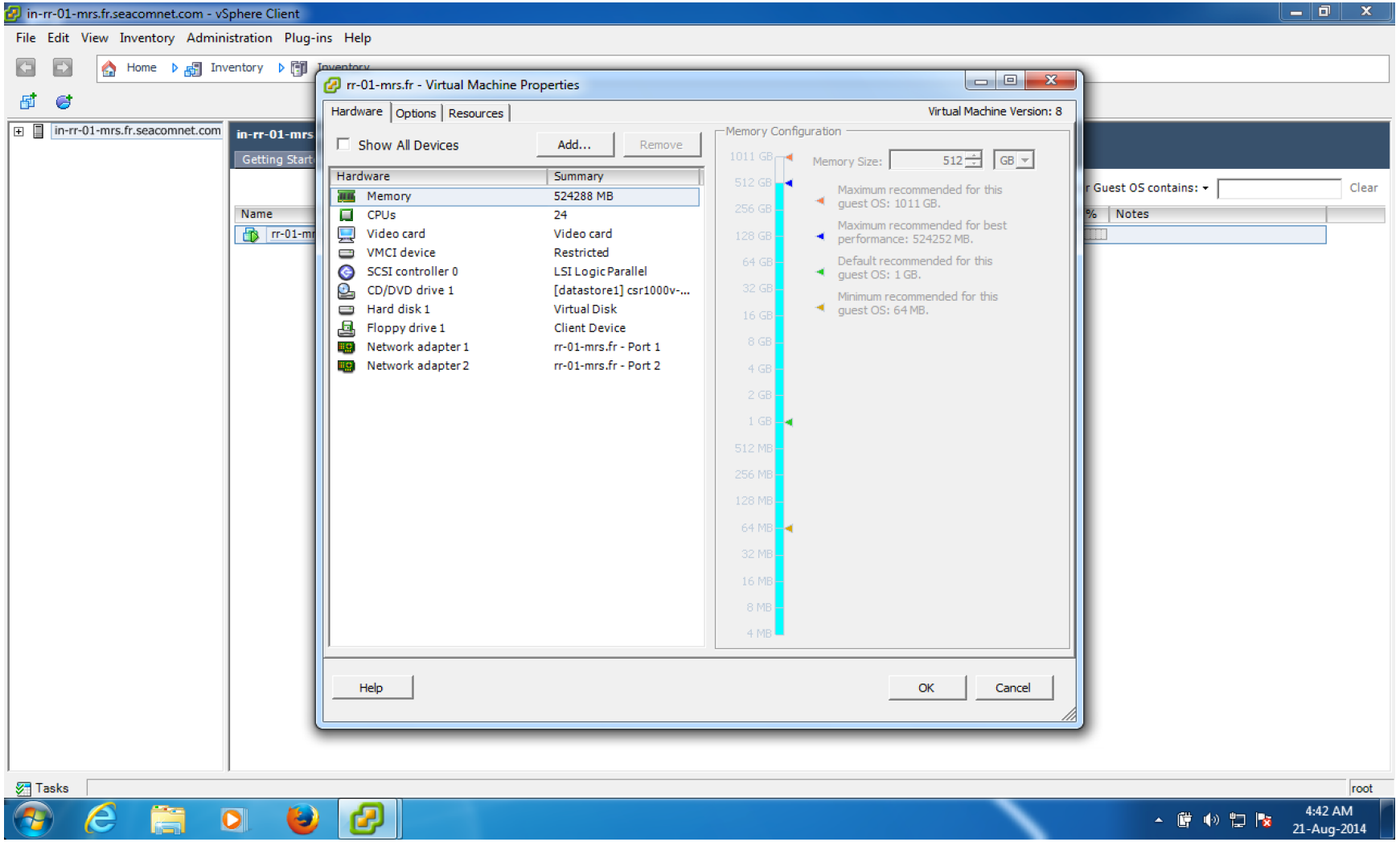
Deployment Nuances



Deployment Nuances



Deployment Nuances



Deployment Nuances

- The different IOS XE releases support different hardware metrics:
  - Number of vCPU's supported.
  - Minimum RAM supported dependent on number of vCPU's.
  - Hard disk drive size required (8GB minimum for all, today).
  - Single "logical" hard drive (for all, today).
  - Minimum ESXi version required.

- Licensing of CSR1000v by Cisco is necessary:
  - Enables software features, e.g., BFD, MPLS, e.t.c.
  - Enables large memory support (16GB as of IOS XE 3.14S).
  - Enables high throughput (2.5Mbps up to 10Gbps).
  - License periods include Evaluation, 1-year, 3-year or Perpetual.
  - License types were Standard, Advanced and Premium (3.12S).
  - Now are IPBase, Security, AX and APPX (3.13S and later).

Operations

Operations

- Basic operational issues to report:
  - It is basically a router, much like a Cisco ASR1000 device.
  - We use BGP-SD to speed up convergence (do not program FIB).
  - Upgrades use ***.bin*** files, so not necessary to re-install.
  - Upgrading ESXi requires different/specific vSphere clients.
  - Installation of a full IPv6 table takes 1x second (32K entries).
  - Installation of a full IPv4 table takes 20x seconds (600K entries).
  - CPU idles at 1% – 3%.
  - CPU can spike to 10% during huge route churn – rare!
  - For now, peak iBGP neighbors are ±300x routers, and growing.
  - IP/MPLS in the Access is driving the route reflector workload.
  - About 80% of iBGP neighbors are exchanging full BGP tables.
  - Take care of MTU if network is larger than 9,000 bytes.
  - Server automatic shutdown due to high data centre temperature.
  - Bug in HP's iLO that crashed ESXi 5.5 from time to time. Patch!
  - All BGP address families are fully supported.
  - IPv4, IPv6, MVPNv4, L2VPN, VPLS, VPNv4, VPNv6.
  - Full multi-vendor support (Cisco & Juniper, in our case).

What The Future Holds

What The Future Holds

- Evaluating performance in terms of failure scenarios.
- Moving parts tend to fail, e.g., fans, hard drives, e.t.c.
- Power supplies are notorious for failing.
- Forgetting to license VMware ESXi will lead to a system outage.

- Vendors are going to release newer VM's with more support.
- Support for higher memory (64GB, 128GB, 256GB, e.t.c.).
- Improvements in leveraging of physical resources (CPU, Network, e.t.c.).

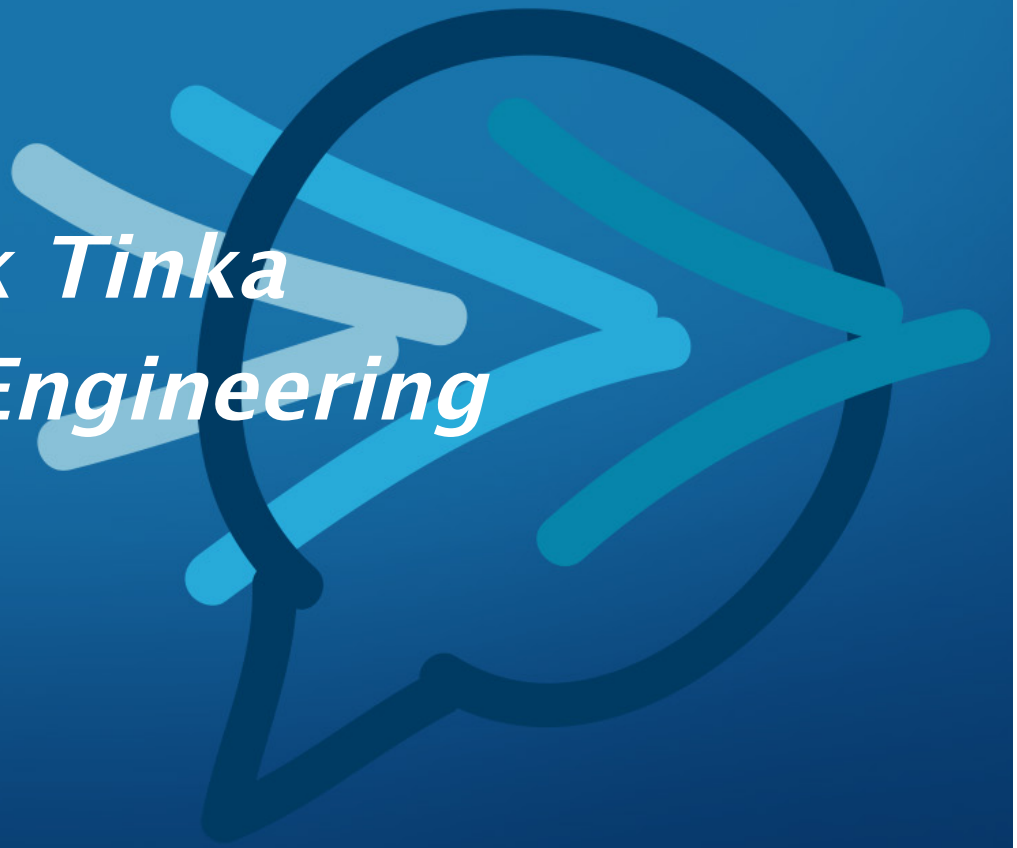
- We shall continue to invest in this technology.
- No reason route reflection should run on purpose-built hardware.
- Use of general-purpose servers with router software in VM's is cool!
- And it works, beautifully!

- Testing on OpenStack/KVM – but for non-route reflection use-cases.

So go out and do it!

Thank You

Q&A



Mark Tinka
Head of Engineering

SEACOM