ISP Design – Using MikroTik CHR as a BGP edge router

PRESENTED BY:
KEVIN MYERS,
NETWORK ARCHITECT
Profile: About Kevin Myers

Background:
• 19+ years in Networking
• Designed/Built Networks on 6 continents
• MikroTik Certified Trainer
• MikroTik, Cisco and Microsoft Certified

Community Involvement:

Packet Pushers (Podcast Guest / Blogger)

Group Contributor (RouterOS / WISP Talk and others)

Delegate/Roundtable contributor (NFD14)

MT Forum (Forum Veteran – Member since 2012)

Network Collective (Podcast Guest)
Profile: About IP ArchiTechs

Expert Networking
Whitebox | ISP | Data Center | Enterprise

✓ Global Consulting
✓ Managed Networks
✓ Monitoring
✓ Load Testing
✓ Development

Locations in: US | Canada | South America

Call us at: +1 855-645-7684
E-mail: consulting@iparchitechs.com
Web: www.iparchitechs.com
Design: **Why use the CHR as a BGP edge router?**

**Goal of this presentation:** When the presentation is finished, hopefully you will have walked away with a few key concepts:

- Performance characteristics of the CHR on multiple hypervisors
- Best practices when deploying the CHR
- Benefits of using the CHR vs CCR as a BGP edge router
Design: **CHR vs. CCR for full BGP tables?**

- Which platform is better?
- Throughput capabilities?
- x86 CPU vs. Tilera?
- BGP Updates, Set BGP Attributes (Community, Localpref)
**Design: CHR vs. CCR for full BGP tables?**

<table>
<thead>
<tr>
<th>Platform</th>
<th>CHR (MicroTik)</th>
<th>CCR (Tilera)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CPU</strong></td>
<td>x86</td>
<td>Tilera</td>
</tr>
<tr>
<td>BGP routers with full tables have a high computational requirement.</td>
<td>Better for heavy computational work. Higher power draw.</td>
<td>Optimized for packet transfer. Designed to be low power draw.</td>
</tr>
<tr>
<td><strong>Throughput</strong></td>
<td>x86</td>
<td>Tilera</td>
</tr>
<tr>
<td>At 1500 bytes, 512 bytes and 64 bytes</td>
<td>More CPU and power is required to move data at the same speed as a CCR</td>
<td></td>
</tr>
<tr>
<td><strong>Routing table size</strong></td>
<td>x86</td>
<td>Tilera</td>
</tr>
<tr>
<td>Impact of multiple tables</td>
<td>Not limited by ASIC, x86 can process multiple tables more quickly than CCR if the HV host has a high clock speed CPU.</td>
<td></td>
</tr>
</tbody>
</table>
Design: CHR testing lab setup

Test Network for CHR

- **tutone.net**
  - bgp1.tutone.net
    - Baltic Networks Vengeance HV
    - BGP AS 65001
    - VLAN 763
    - 100.76.1.0/29
    - bgp2.tutone.net
    - CHR
    - VLAN 761
    - 100.76.2.0/29

- **ipperf3**
  - 100.77.1.0/29
  - VLAN 777

- **BGP AS 8675309**

- **BGP AS 6695**
  - AS 5459
  - eBGP peer

- **BGP AS 65001 (15459,16696)**
  - Lab 7606
  - VLAN 762
  - 100.76.2.0/29
  - VLAN 1400
  - eBGP peer
  - bgp1.tutone.net
    - CHR
    - VLAN 763
    - 100.76.3.0/29
  - bgp2.tutone.net
    - CHR
    - VLAN 761
    - 100.76.1.0/29

- **ipperf3**
  - 100.77.1.0/29
  - VLAN 777

- **Test Network for CHR**
  - tutone.net
  - Baltic Networks Vengeance HV
  - bgp1.tutone.net
    - CHR
    - VLAN 763
    - 100.76.3.0/29
Design: CHR testing lab setup

Physical Test Network for CHR

Baltic Networks Vengeance HV

HP 1U x86 Server

10 Gbps DAC

iperf3

BGP AS 8675309

tutone.net

DE-CIX
AS 6695

eBGP peer

DE-CIX
AS 5459

eBGP peer

BGP AS 65001 (15459,16696)

Lab 7606

HP 1U x86 Server

10 Gbps DAC

iperf3

BGP AS 8675309
• Hypervisor details – VM provisioning

• 8 vCPUs Total
  • Performance was better when using more sockets vs. more cores with a single socket
  • With BGP – single core performance is critical. You need to use a CPU with the most powerful single core performance that you can afford for the hypervisor host.

• 4096 MB RAM
• **Concept of testing**
  
  • Performance on 1 BGP full table
  • Performance on 2 BGP full tables
  • Performance at 1500 bytes, 512 bytes and 96 bytes
  • Throughput performance with convergence test

• **Performance settings**
  
  • Sockets vs CPUs – More sockets is better
  • TSO/LSO – Disable for best performance
  • Clock speed – Highest speed possible
<table>
<thead>
<tr>
<th>Platform</th>
<th>Hypervisor</th>
<th>CHR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baltic Vengeance</td>
<td>VM Ware ESXi 6.5</td>
<td>6.41.3</td>
</tr>
</tbody>
</table>

DE-CIX March 2018 Routing Table: Time - 26 seconds
### Design: CHR performance on VMWARE ESXi (2 Full Tables)

<table>
<thead>
<tr>
<th>Platform</th>
<th>Hypervisor</th>
<th>CHR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baltic Vengeance</td>
<td>VM Ware ESXi 6.5</td>
<td>6.41.3</td>
</tr>
</tbody>
</table>

DE-CIX and LINX March 2018 Routing Table: **Time – 4:46 seconds**

---

**BGP**

<table>
<thead>
<tr>
<th>Name</th>
<th>Remote Address</th>
<th>Remote AS</th>
<th>Uptime</th>
<th>Prefix Count</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>LINX (IPA-JXN-7606-01)</td>
<td>100.76.1.1</td>
<td>65001</td>
<td>00:04:46</td>
<td>527926</td>
<td>established</td>
</tr>
<tr>
<td>bgp2.tutone.net</td>
<td>10.255.236.150</td>
<td>8675309</td>
<td>00:04:34</td>
<td>678597</td>
<td>established</td>
</tr>
</tbody>
</table>

50 (bgp2.tutone.net (ESXi)) - WinBox v6.41.3 on CHR (x86_64)

**Session:** 10.255.236.150

**BGP**

<table>
<thead>
<tr>
<th>Name</th>
<th>Remote Address</th>
<th>Remote AS</th>
<th>Uptime</th>
<th>Prefix Count</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>DE-CIX (IPA-JXN-7606-01)</td>
<td>100.76.3.1</td>
<td>16695</td>
<td>00:04:35</td>
<td>680122</td>
<td>established</td>
</tr>
<tr>
<td>bgp1.tutone.net</td>
<td>10.255.236.160</td>
<td>8675309</td>
<td>00:04:35</td>
<td>100150</td>
<td>established</td>
</tr>
</tbody>
</table>
Design: CHR performance on VMWARE ESXi (1500 byte MTU)

<table>
<thead>
<tr>
<th>Platform</th>
<th>Hypervisor</th>
<th>CHR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baltic Vengeance</td>
<td>VM Ware ESXi 6.5</td>
<td>6.41.3</td>
</tr>
</tbody>
</table>

Throughput: **5.3 Gbps**  Peak VM CPU: **28%**
### Design: CHR performance on VMWARE ESXi (512 byte MTU)

<table>
<thead>
<tr>
<th>Platform</th>
<th>Hypervisor</th>
<th>CHR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baltic Vengeance</td>
<td>VM Ware ESXi 6.5</td>
<td>6.41.3</td>
</tr>
</tbody>
</table>

**Throughput:** 3.0 Gbps  
**Peak VM CPU:** 38%
Design: **CHR performance on VMWARE ESXi (96 byte MTU)**

<table>
<thead>
<tr>
<th>Platform</th>
<th>Hypervisor</th>
<th>CHR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baltic Vengeance</td>
<td>VM Ware ESXi 6.5</td>
<td>6.41.3</td>
</tr>
</tbody>
</table>

**Throughput:** **1.1 Gbps**  
**Peak VM CPU:** **43%**
Design: CHR performance on VMWARE ESXi (full table + bw load)

<table>
<thead>
<tr>
<th>Platform</th>
<th>Hypervisor</th>
<th>CHR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baltic Vengeance</td>
<td>VM Ware ESXi 6.5</td>
<td>6.41.3</td>
</tr>
</tbody>
</table>

5 Gbps load plus reset of both upstreams: Convergence Time – 11:05
### Design: CHR performance on Proxmox (KVM) (1 Full Table)

<table>
<thead>
<tr>
<th>Platform</th>
<th>Hypervisor</th>
<th>CHR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baltic Vengeance</td>
<td>ProxMox (KVM) 5.1</td>
<td>6.41.3</td>
</tr>
</tbody>
</table>

DE-CIX March 2018 Routing Table: **Time - 26 seconds**
<table>
<thead>
<tr>
<th>Platform</th>
<th>Hypervisor</th>
<th>CHR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baltic Vengeance</td>
<td>ProxMox (KVM) 5.1</td>
<td>6.41.3</td>
</tr>
</tbody>
</table>

DE-CIX and LINX March 2018 Routing Table: **Time – 1:34**
Design: **CHR performance on Proxmox (KVM) (1500 byte MTU)**

<table>
<thead>
<tr>
<th>Platform</th>
<th>Hypervisor</th>
<th>CHR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baltic Vengeance</td>
<td>ProxMox (KVM) 5.1</td>
<td>6.41.3</td>
</tr>
</tbody>
</table>

Throughput: **4.4 Gbps**  Peak VM CPU: **28%**
Design: CHR performance on Proxmox (KVM) (512 byte MTU)

<table>
<thead>
<tr>
<th>Platform</th>
<th>Hypervisor</th>
<th>CHR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baltic Vengeance</td>
<td>ProxMox (KVM) 5.1</td>
<td>6.41.3</td>
</tr>
</tbody>
</table>

Throughput: **1.8 Gbps**  Peak VM CPU: **29%**
**Design:** CHR performance on Proxmox (KVM) (96 byte MTU)

<table>
<thead>
<tr>
<th>Platform</th>
<th>Hypervisor</th>
<th>CHR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baltic Vengeance</td>
<td>ProxMox (KVM) 5.1</td>
<td>6.41.3</td>
</tr>
</tbody>
</table>

Throughput: **423 Mbps** Peak VM CPU: **28%**
Design: **CHR performance on PROXMOX (KVM)(full table + bw load)**

<table>
<thead>
<tr>
<th>Platform</th>
<th>Hypervisor</th>
<th>CHR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baltic Vengeance</td>
<td>VM Ware ESXi 6.5</td>
<td>6.41.3</td>
</tr>
</tbody>
</table>

5 Gbps load plus reset of both upstreams: **Convergence Time – 9:03**
Design: **CHR performance on Hyper V (Microsoft)**

<table>
<thead>
<tr>
<th>Platform</th>
<th>Hypervisor</th>
<th>CHR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baltic Vengeance</td>
<td>Hyper-V 2012</td>
<td>6.41.3</td>
</tr>
</tbody>
</table>

DE-CIX March 2018 Routing Table: **Time - 12 seconds**

![BGP session details](image-url)
Design: **CHR performance on Hyper V (Microsoft)** (2 full tables)

<table>
<thead>
<tr>
<th>Platform</th>
<th>Hypervisor</th>
<th>CHR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baltic Vengeance</td>
<td>Hyper-V 2012</td>
<td>6.41.3</td>
</tr>
</tbody>
</table>

**DE-CIX and LINX March 2018 Routing Table**: Time - 41 seconds
Design: **CHR performance on Hyper-V (Microsoft)** (1500 byte MTU)

<table>
<thead>
<tr>
<th>Platform</th>
<th>Hypervisor</th>
<th>CHR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baltic Vengeance</td>
<td>Hyper-V 2012</td>
<td>6.41.3</td>
</tr>
</tbody>
</table>

**Throughput:** 4.4 Gbps  
**Peak VM CPU:** 32%
<table>
<thead>
<tr>
<th>Platform</th>
<th>Hypervisor</th>
<th>CHR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baltic Vengeance</td>
<td>Hyper-V 2012</td>
<td>6.41.3</td>
</tr>
</tbody>
</table>

Throughput: **1.6 Gbps**  Peak VM CPU: **33%**

Design: **CHR performance on Hyper-V (Microsoft)(512 byte MTU)**
### Design: CHR performance on Hyper-V (Microsoft) (96 byte MTU)

<table>
<thead>
<tr>
<th>Platform</th>
<th>Hypervisor</th>
<th>CHR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baltic Vengeance</td>
<td>Hyper-V 2012</td>
<td>6.41.3</td>
</tr>
</tbody>
</table>

**Throughput:** 646 Mbps  
**Peak VM CPU:** 37%
<table>
<thead>
<tr>
<th>Platform</th>
<th>Hypervisor</th>
<th>CHR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baltic Vengeance</td>
<td>Hyper-V 2012</td>
<td>6.41.3</td>
</tr>
</tbody>
</table>

5 Gbps load plus reset of both upstreams: **Convergence Time – :43**
Your local internet source for the Pacific Northwest
Design: CHR real world performance example - PogoZone

US WISP - Throughput: **1.6 Gbps** Peak VM CPU: **32%**
• Which Hypervisor is the best?
  • #1 Hyper-V was faster than ESXi and ProxMox all the way around and the most consistent with results in testing. It was by far the clear winner in routing convergence times
  • #2 ProxMox KVM delivered much better convergence speeds than ESXi but not quite as good as Hyper-V
  • #3 VM Ware ESXi was the slowest in everything but raw throughput and it only has a marginal edge over Hyper-V in that area.
Questions??