ArchiTechs

MANAGED SERVICES

www.iparchitechs.com

ISP Design – Using a full table RR to improve eBGP performance with MikroTik routers

PRESENTED BY:

KEVIN MYERS,
NETWORK ARCHITECT

Profile: About Kevin Myers

Background:

- 20+ 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

IP ArchiTechs
MANAGED SERVICES

Expert Networking

Whitebox | ISP | Data Center | Enterprise

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

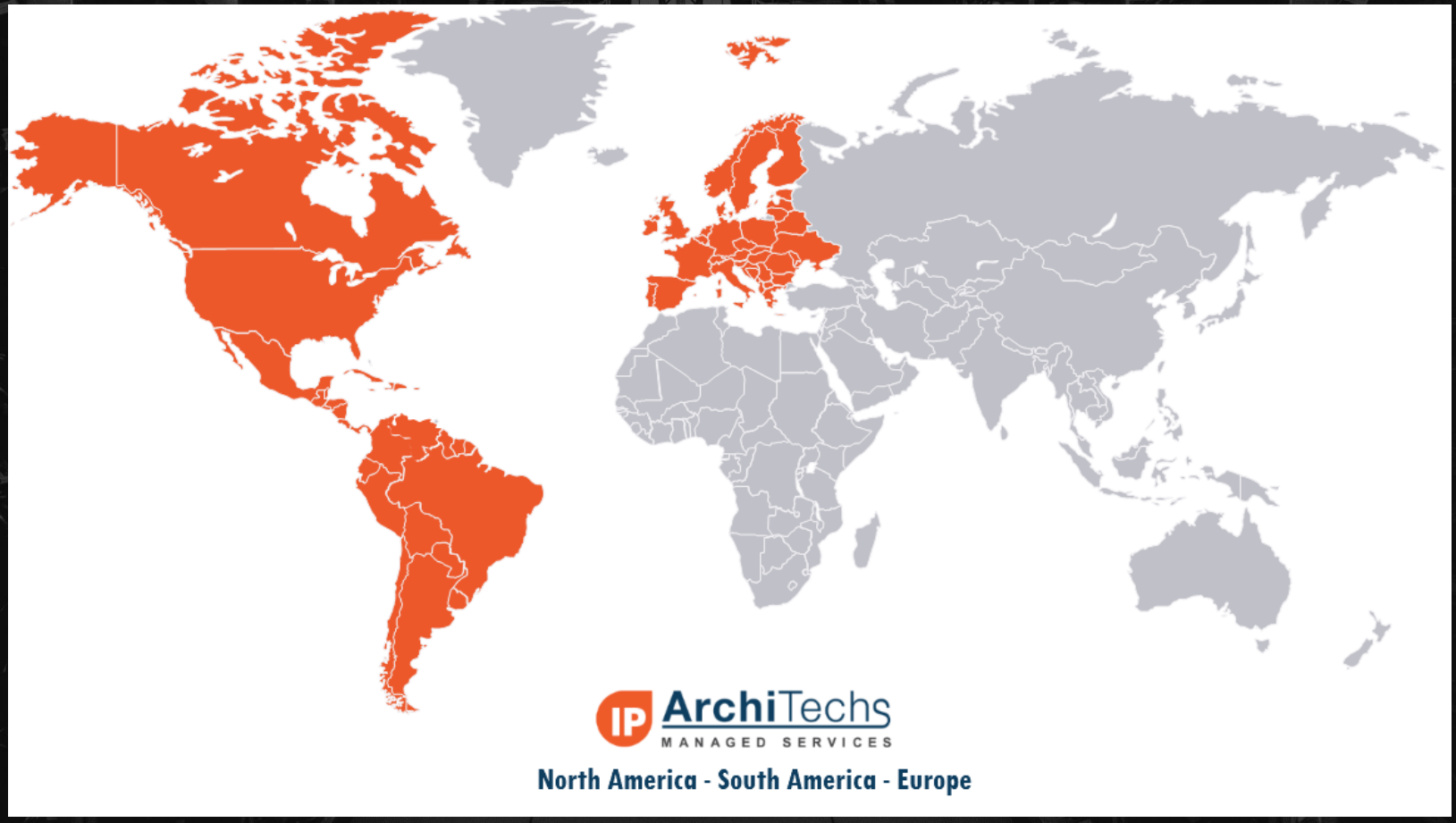
Locations in: US | Canada | South America

Call us at: +1 855-645-7684
E-mail: consulting@iparchitechs.com
Web: www.iparchitechs.com

Profile: About IP ArchiTechs

Now in Europe!

IPA Opened an office in Nis, Serbia in 2018



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

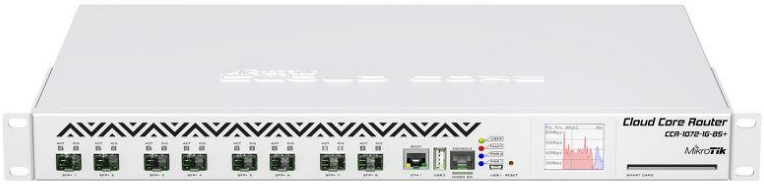
- Performance limitations of using full mesh peering between BGP border routers
- How to leverage open source software to create a high performance BGP RR for MikroTik border routers
- Design benefits of using a BGP full table RR

Design: **CHR vs. Hardware for full tables?**

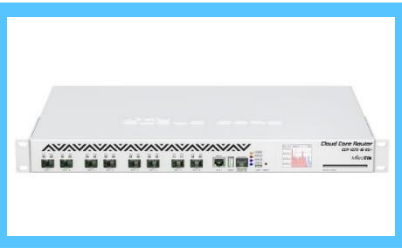
- Which platform is better?
- Throughput capabilities?
- x86 CPU vs. ARM/Tilera?

MikroTik CHR

VS.



Design: CHR vs. Tilera/ARM for BGP Border?

Platform	MikroTik CHR		
CPU MPLS router CPU requirements depend on load and explicit/implicit null	**x86** Better for heavy computational work. Higher power draw.	**Tilera** Optimized for packet transfer. Designed to be low power draw.	**ARM** In between x86 and Tilera for performance.
Throughput At 1530 bytes (L2), and 8970 bytes (L2)	**x86** More CPU and power is required to move data at the same speed as a CCR	**Tilera** Handles throughput at different frame sizes slightly better than x86	**ARM** Handles throughput at different frame sizes similar to Tilera
Performance for full tables	**x86** x86 hardware with HyperV is the best RouterOS based solution for full BGP tables	**Tilera** Lowest performance for full tables	**ARM** Performance is in between Tilera and x86 but the best hardware platform from MikroTik for rapid convergence

Design: **What is Free Range Routing?**

- FRRouting (FRR) is an open source IP routing protocol suite for Linux and Unix platforms
- includes protocol daemons for BGP, IS-IS, LDP, OSPF, PIM, and RIP.Use MikroTik CCRs or CHRs for BGP border routers to handle throughput
- FRR has its roots in the Quagga project. In fact, it was started by many long-time Quagga developers who combined their efforts to improve on Quagga's well-established foundation.

RIP v4 | v6

OSPF v4 | v6

ISIS v4 | v6

BGP4 v4 | v6

BGP4 (4 byte AS) v4 | v6

LDP

Design: **Why use Free Range Routing?**

- FRR is multithreaded for BGP and can improve the performance of MikroTik BGP border routers by acting as a route reflector
- Free software since it is open source – can be compiled from the source or downloaded as VM using CumulusVX.
- Has been traditionally used as an IX route server and can handle a large volume of Full Tables

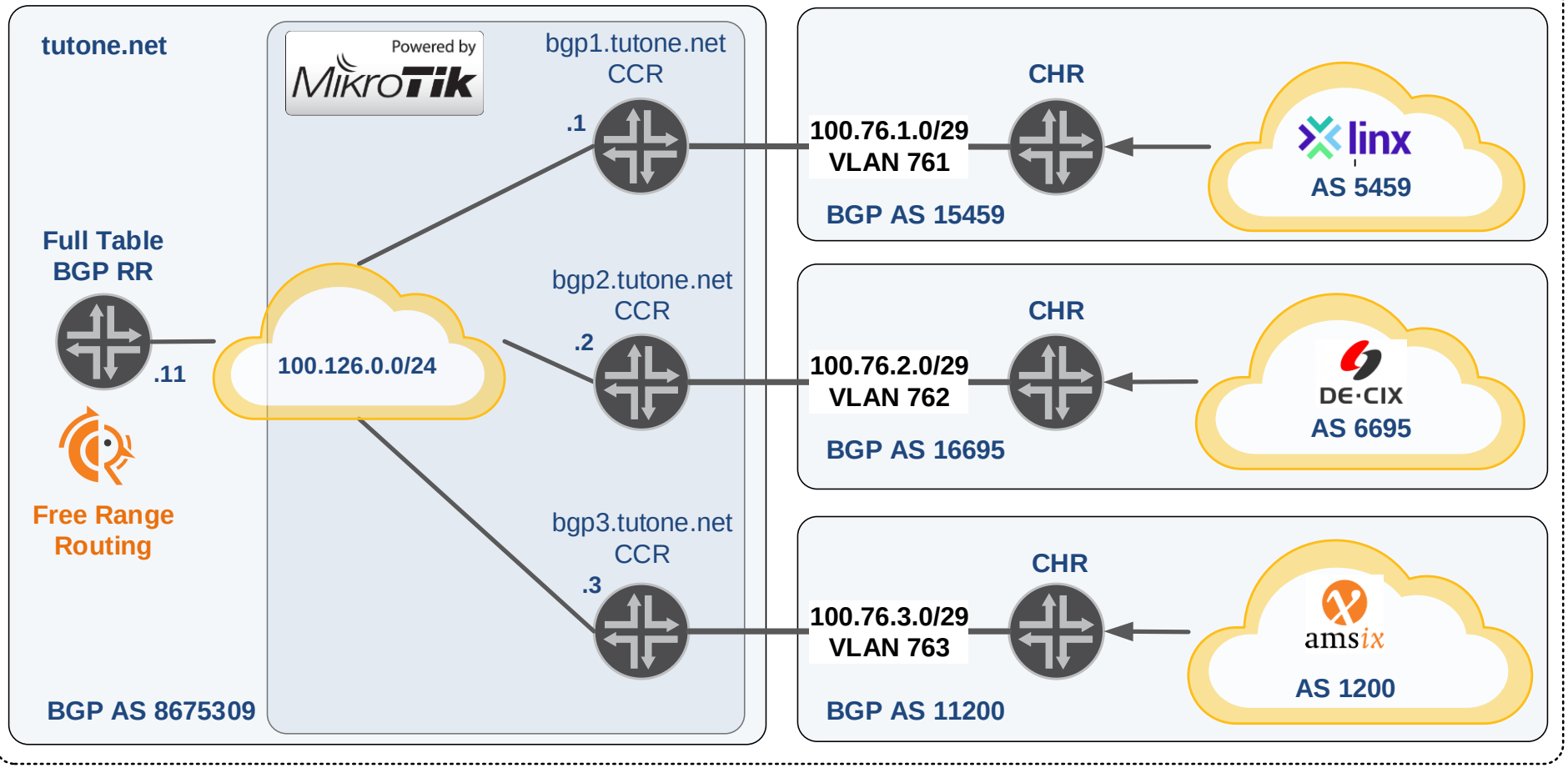
Design: **Which design for full table route reflection?**

- Free Range Routing for BGP RR – Open source routing package that supports BGP across multiple CPUs. Doesn't need throughput – out of path
- Use MikroTik CCRs or CHRs for BGP border routers to handle throughput
- Leverage the strength of Open Source software to increase the performance and scalability of RouterOS as a BGP border router

Free Range Routing

+

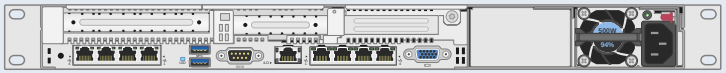
Test Network for full table RR



Design: BGP RR – switch–centric lab setup

Physical Test Network for BGP RR

HP 1U x86 Server



CHR

BGP Full Feeds

Core switch

Baltic Networks Vengeance HV

CHR

CHR Border routers

BGP Border routers

- **Concept of testing**
  - Performance with 3 full tables on CCRs & BGP full table RR
  - Performance with 3 full tables on CCRs & full mesh peerings
  - Performance with 4 full tables on CCRs + CHR & BGP full table RR
  - Performance with 4 full tables on CCRs + CHR & full mesh peerings
  - Performance with 3 full tables on CHR & BGP full table RR
  - Performance with 3 full tables on CHR & full mesh peerings

Design: **Full Table BGP RR performance**

BGP Border	BGP RR	Hypervisor
CCR1036 (2) CCR1072 (1)	Free Range Routing	ESXi 6.0

Total routes: **1,757,631** IP Transit Convergence: **2:55** Full Convergence: **8:51**

```
cumulus@cumulus:~$ net show bgp sum
show bgp ipv4 unicast summary
=============================
BGP router identifier 100.126.0.11, local AS number 8675309 vrf-id 0
BGP table version 1084392
RIB entries 1424597, using 207 MiB of memory
Peers 3, using 58 KiB of memory

Neighbor        V         AS MsgRcvd MsgSent   TblVer  InQ OutQ  Up/Down State/PfxRcd
100.126.0.1     4    8675309  120307  247211        0    0    0 00:08:48       315916
100.126.0.2     4    8675309  134326  247526        0    0    0 00:08:51       539823
100.126.0.3     4    8675309   80099  247525        0    0    0 00:08:49       255494

Total number of neighbors 3
```

Design: **Full Table BGP performance (Full Mesh Peerings)**

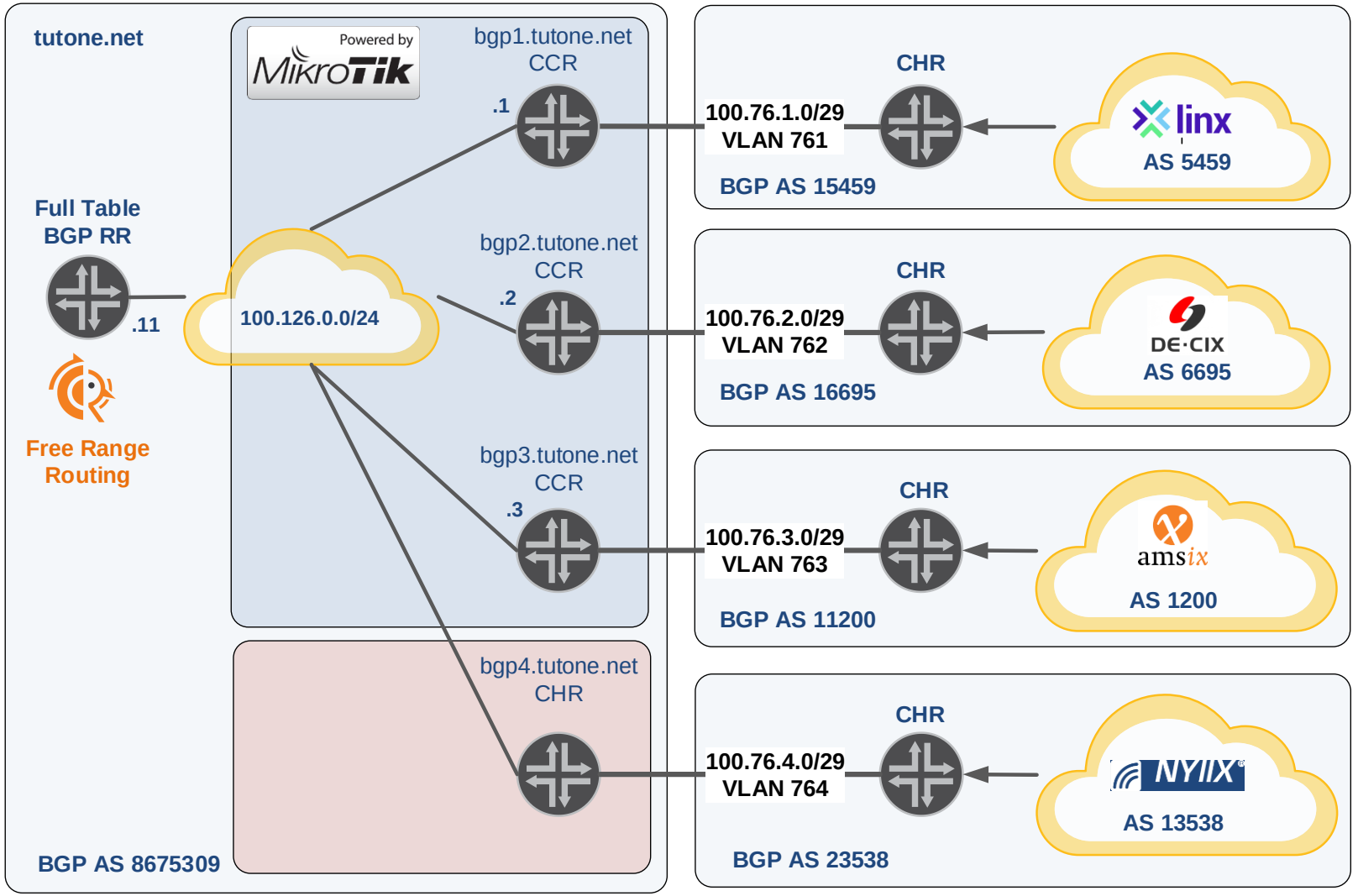
BGP Border	BGP RR	Hypervisor
CCR1036 (2) CCR1072 (1)	None	ESXi 6.0

Total routes: **1,757,631** IP Transit Convergence: **6:58** Full Convergence: **8:59**

```
[admin@US-MUM-2019-BGP-1] > routing bgp peer print status
Flags: X - disabled, E - established
 0 E name="LINX" instance=default remote-address=100.76.1.1 remote-as=15459 tcp-md5-key="" nexthop-choice=default multihop=no
     route-reflect=no hold-time=3m ttl=default in-filter="" out-filter=deny-all address-families=ip default-originate=never
     remove-private-as=no as-override=no passive=no use-bfd=no remote-id=10.255.227.76 local-address=100.76.1.2 uptime=9m
     prefix-count=526315 updates-sent=0 updates-received=526315 withdrawn-sent=0 withdrawn-received=0 remote-hold-time=3m
     used-hold-time=3m used-keepalive-time=1m refresh-capability=yes as4-capability=yes state=established

 1 X name="Full Table RR" instance=default remote-address=100.126.0.11 remote-as=8675309 tcp-md5-key="" nexthop-choice=force-self
     multihop=no route-reflect=no hold-time=3m ttl=default in-filter="" out-filter="" address-families=ip default-originate=never
     remove-private-as=no as-override=no passive=no use-bfd=no

 2 E name="BGP-2" instance=default remote-address=100.126.0.2 remote-as=8675309 tcp-md5-key="" nexthop-choice=default multihop=no
     route-reflect=no hold-time=3m ttl=default in-filter="" out-filter="" address-families=ip default-originate=never
     remove-private-as=no as-override=no passive=no use-bfd=no remote-id=100.127.1.2 local-address=100.126.0.1 uptime=8m59s
     prefix-count=677314 updates-sent=526315 updates-received=677314 withdrawn-sent=0 withdrawn-received=0 remote-hold-time=3m
     used-hold-time=3m used-keepalive-time=1m refresh-capability=yes as4-capability=yes state=established

 3 E name="BGP-3" instance=default remote-address=100.126.0.3 remote-as=8675309 tcp-md5-key="" nexthop-choice=default multihop=no
     route-reflect=no hold-time=3m ttl=default in-filter="" out-filter="" address-families=ip default-originate=never
     remove-private-as=no as-override=no passive=no use-bfd=no remote-id=100.126.0.3 local-address=100.126.0.1 uptime=9m
     prefix-count=246972 updates-sent=526315 updates-received=355270 withdrawn-sent=0 withdrawn-received=108556 remote-hold-time=3m
     used-hold-time=3m used-keepalive-time=1m refresh-capability=yes as4-capability=yes state=established
[admin@US-MUM-2019-BGP-1] >
```

Design: BGP RR - logical lab setup

Test Network for full table RR – 4 peers (CCR+CHR)



Design: Full Table BGP RR performance

BGP Border	BGP RR	Hypervisor
CCR1036 (2) CCR1072 (1) CHR (1) HyperV	Free Range Routing	ESXi 6.0

Total routes: **2,272,474** IP Transit Convergence: **5:33** Full Convergence: **16:51**

```
cumulus@cumulus:~$ net show bgp sum
show bgp ipv4 unicast summary
=============================
BGP router identifier 100.126.0.11, local AS number 8675309 vrf-id 0
BGP table version 7730759
RIB entries 1430606, using 207 MiB of memory
Peers 4, using 77 KiB of memory

Neighbor        V         AS MsgRcvd MsgSent   TblVer  InQ OutQ  Up/Down State/PfxRcd
100.126.0.1     4    8675309  449514  988524        0    0    0 00:16:51       268655
100.126.0.2     4    8675309  468724 1021051        0    0    0 00:16:49       480851
100.126.0.3     4    8675309  329318  992992        0    0    0 00:16:46       209710
100.126.0.4     4    8675309  414039  958768        0    0    0 00:16:45       306614

Total number of neighbors 4
```

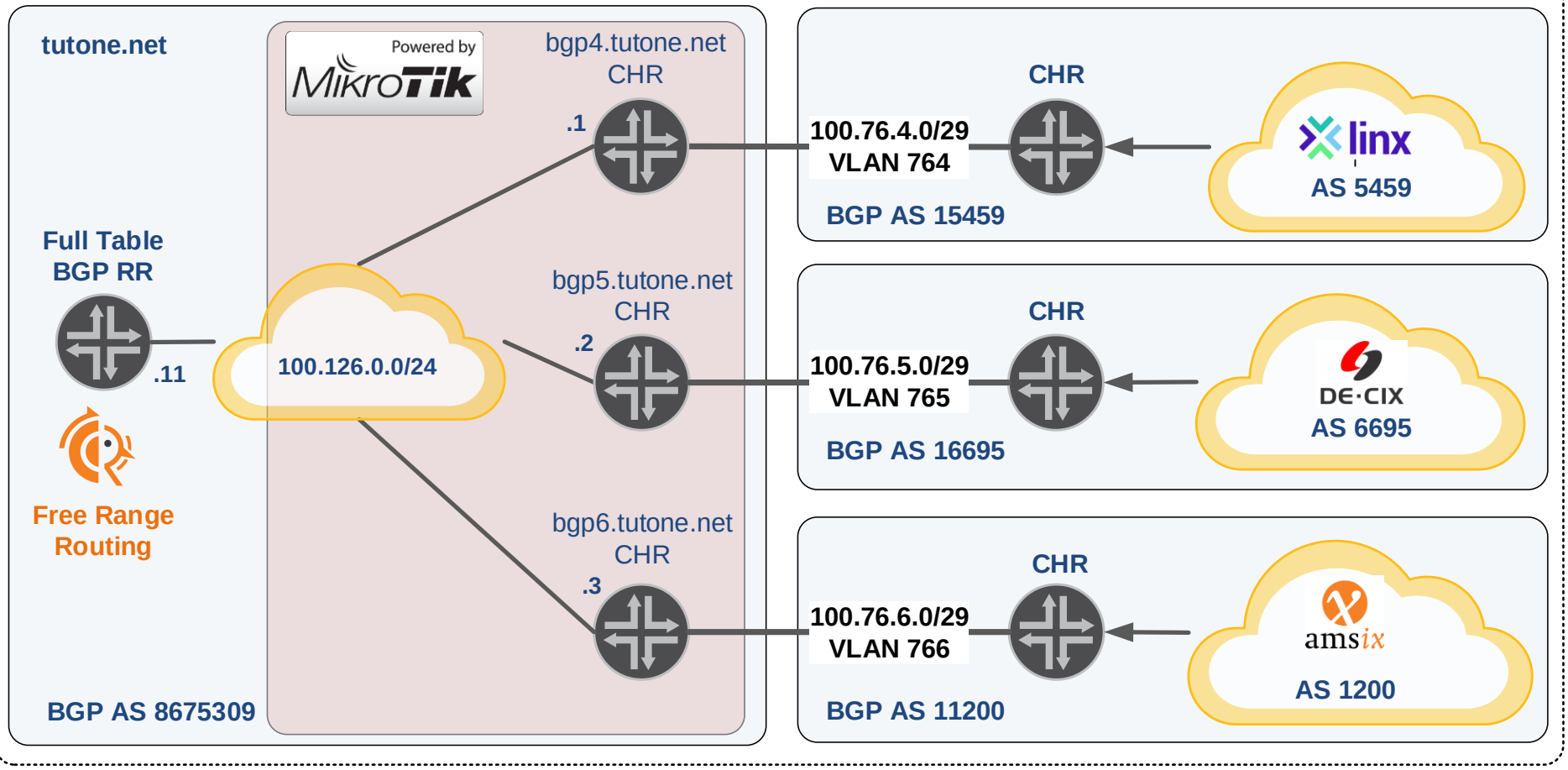
Design: **Full Table BGP performance (Full Mesh Peerings)**

BGP Border	BGP RR	Hypervisor
CCR1036 (2) CCR1072 (1) CHR (1) HyperV	None	ESXi 6.0

Total routes: **2,272,474** IP Transit Convergence: **7:27** Full Convergence: **17:01**

```
[admin@US-MUM-2019-BGP-1] > routing bgp peer print status
Flags: X - disabled, E - established
 0 E name="LINX" instance=default remote-address=100.76.1.1 remote-as=15459 tcp-md5-key="" nexthop-choice=default multihop=no
     route-reflect=no hold-time=3m ttl=default in-filter="" out-filter=deny-all address-families=ip default-originate=never
     remove-private-as=no as-override=no passive=no use-bfd=no remote-id=10.255.227.76 local-address=100.76.1.2 uptime=17m17s
     prefix-count=526315 updates-sent=0 updates-received=526315 withdrawn-sent=0 withdrawn-received=0 remote-hold-time=3m
     used-hold-time=3m used-keepalive-time=1m refresh-capability=yes as4-capability=yes state=established

 1 X name="Full Table RR" instance=default remote-address=100.126.0.11 remote-as=8675309 tcp-md5-key="" nexthop-choice=force-self
     multihop=no route-reflect=no hold-time=3m ttl=default in-filter="" out-filter="" address-families=ip default-originate=never
     remove-private-as=no as-override=no passive=no use-bfd=no

 2 E name="BGP-2" instance=default remote-address=100.126.0.2 remote-as=8675309 tcp-md5-key="" nexthop-choice=default multihop=no
     route-reflect=no hold-time=3m ttl=default in-filter="" out-filter="" address-families=ip default-originate=never
     remove-private-as=no as-override=no passive=no use-bfd=no remote-id=100.127.1.2 local-address=100.126.0.1 uptime=17m13s
     prefix-count=677314 updates-sent=526315 updates-received=677314 withdrawn-sent=0 withdrawn-received=0 remote-hold-time=3m
     used-hold-time=3m used-keepalive-time=1m refresh-capability=yes as4-capability=yes state=established

 3 E name="BGP-3" instance=default remote-address=100.126.0.3 remote-as=8675309 tcp-md5-key="" nexthop-choice=default multihop=no
     route-reflect=no hold-time=3m ttl=default in-filter="" out-filter="" address-families=ip default-originate=never
     remove-private-as=no as-override=no passive=no use-bfd=no remote-id=100.126.0.3 local-address=100.126.0.1 uptime=17m6s
     prefix-count=187173 updates-sent=526315 updates-received=327236 withdrawn-sent=0 withdrawn-received=140194 remote-hold-time=3m
     used-hold-time=3m used-keepalive-time=1m refresh-capability=yes as4-capability=yes state=established

 4 E name="BGP-4" instance=default remote-address=100.126.0.4 remote-as=8675309 tcp-md5-key="" nexthop-choice=default multihop=no
     route-reflect=no hold-time=3m ttl=default in-filter="" out-filter="" address-families=ip default-originate=never
     remove-private-as=no as-override=no passive=no use-bfd=no remote-id=100.76.4.2 local-address=100.126.0.1 uptime=17m4s
     prefix-count=514843 updates-sent=526315 updates-received=514843 withdrawn-sent=0 withdrawn-received=0 remote-hold-time=3m
     used-hold-time=3m used-keepalive-time=1m refresh-capability=yes as4-capability=yes state=established
```

Test Network for full table RR – CHR Border



Design: Full Table BGP RR performance

BGP Border	BGP RR	Hypervisor
CHR (3) HyperV	Free Range Routing	ESXi 6.0

Total routes: **2,272,474** IP Transit Convergence: **00:17** Full Convergence: **00:41**

```
cumulus@cumulus:~$ net show bgp sum
show bgp ipv4 unicast summary
=============================
BGP router identifier 100.126.0.11, local AS number 8675309 vrf-id 0
BGP table version 14490241
RIB entries 1422829, using 206 MiB of memory
Peers 6, using 116 KiB of memory

Neighbor        V         AS MsgRcvd MsgSent   TblVer  InQ OutQ  Up/Down State/PfxRcd
100.126.0.1     4    8675309  449559 1000166        0    0    0 15:06:11       Active
100.126.0.2     4    8675309  468767 1026375        0    0    0 15:06:15       Active
100.126.0.3     4    8675309  329830 1029218        0    0    0 15:05:57       Active
100.126.0.4     4    8675309  823584 1928456        0    0    0 00:00:37       316595
100.126.0.5     4    8675309  487847 1025502        0    0    0 00:00:39       521203
100.126.0.6     4    8675309  289276 1022516        0    0    0 00:00:41       247355

Total number of neighbors 6
```

Design: **Full Table BGP performance (Full Mesh Peerings)**

BGP Border	BGP RR	Hypervisor
CHR (3) HyperV	Free Range Routing	ESXi 6.0

Total routes: **2,272,474** IP Transit Convergence: **00:35** Full Convergence: **2:00**

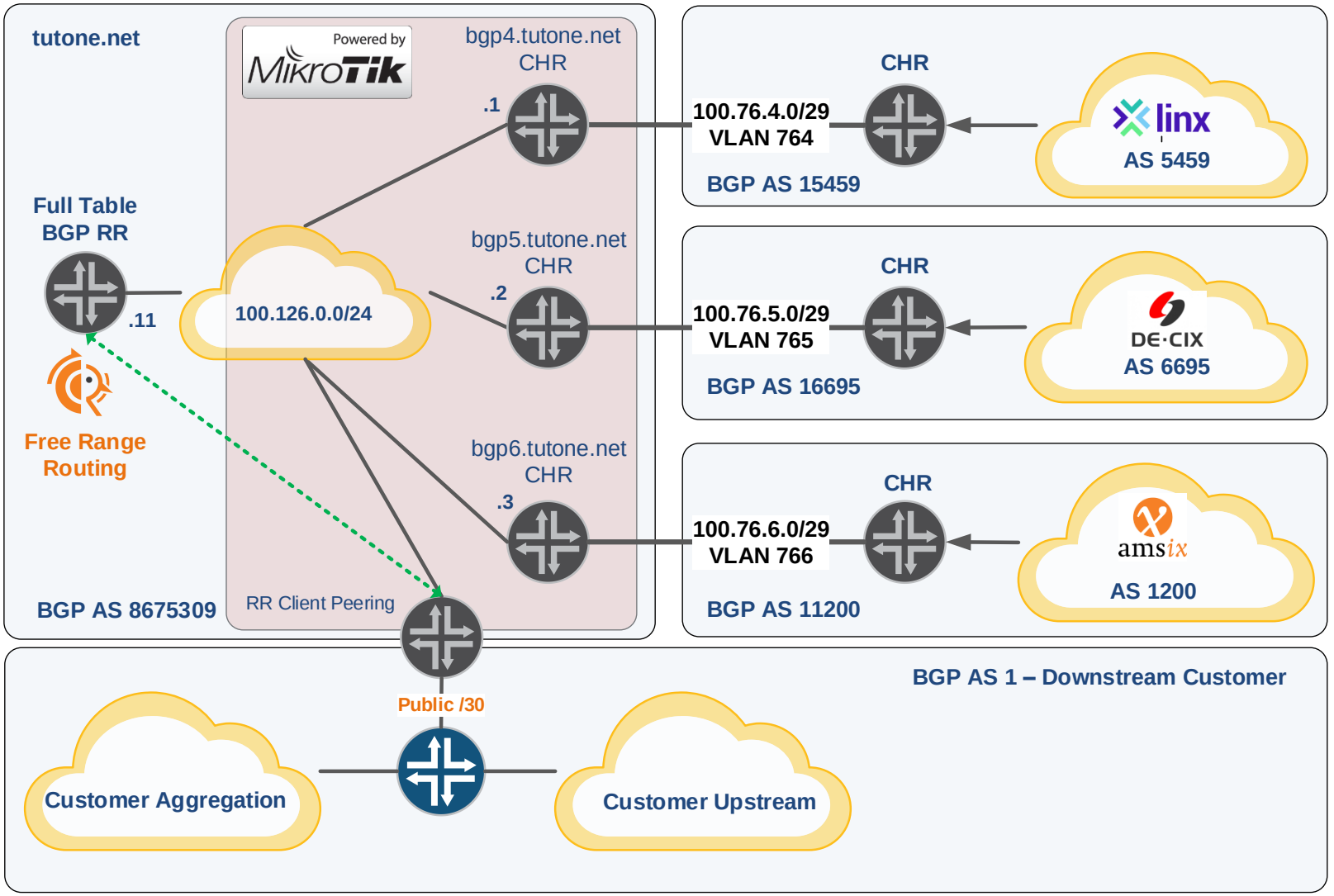
```
[admin@US-MUM-2019-BGP-4] > routing bgp peer print det status
Flags: X - disabled, E - established
 0 X name="Full Table RR" instance=default remote-address=100.126.0.11 remote-as=8675309 tcp-md5-key="" nexthop-choice=force-self
     multihop=no route-reflect=no hold-time=3m ttl=default in-filter="" out-filter="" address-families=ip default-originate=never
     remove-private-as=no as-override=no passive=no use-bfd=no

 1 E name="LINX" instance=default remote-address=100.76.4.1 remote-as=15459 tcp-md5-key="" nexthop-choice=default multihop=no
     route-reflect=no hold-time=3m ttl=default in-filter="" out-filter=deny-all address-families=ip default-originate=never
     remove-private-as=no as-override=no passive=no use-bfd=no remote-id=100.76.1.1 local-address=100.76.4.2 uptime=2m
     prefix-count=526276 updates-sent=0 updates-received=526276 withdrawn-sent=0 withdrawn-received=0 remote-hold-time=3m
     used-hold-time=3m used-keepalive-time=1m refresh-capability=yes as4-capability=yes state=established

 2 E name="BGP-5" instance=default remote-address=100.126.0.5 remote-as=8675309 tcp-md5-key="" nexthop-choice=default multihop=no
     route-reflect=no hold-time=3m ttl=default in-filter="" out-filter="" address-families=ip default-originate=never
     remove-private-as=no as-override=no passive=no use-bfd=no remote-id=100.76.5.1 local-address=100.126.0.4 uptime=1m59s
     prefix-count=674480 updates-sent=526276 updates-received=674480 withdrawn-sent=0 withdrawn-received=0 remote-hold-time=3m
     used-hold-time=3m used-keepalive-time=1m refresh-capability=yes as4-capability=yes state=established

 3 E name="BGP-6" instance=default remote-address=100.126.0.6 remote-as=8675309 tcp-md5-key="" nexthop-choice=default multihop=no
     route-reflect=no hold-time=3m ttl=default in-filter="" out-filter="" address-families=ip default-originate=never
     remove-private-as=no as-override=no passive=no use-bfd=no remote-id=100.76.6.1 local-address=100.126.0.4 uptime=1m59s
     prefix-count=554001 updates-sent=526276 updates-received=554001 withdrawn-sent=0 withdrawn-received=0 remote-hold-time=3m
     used-hold-time=3m used-keepalive-time=1m refresh-capability=yes as4-capability=yes state=established
```

- **Performance Conclusions**
  - BGP RR can help improve the performance of existing CCRs in the load time of the route table from the upstreams
  - Does not significantly impact overall convergence time
  - As the number of IX and IP Transit peers increases, RR becomes more useful for performance and scale
  - CHR + BGP RR achieves the best performance and scale options

Design Advantages of BGP RR

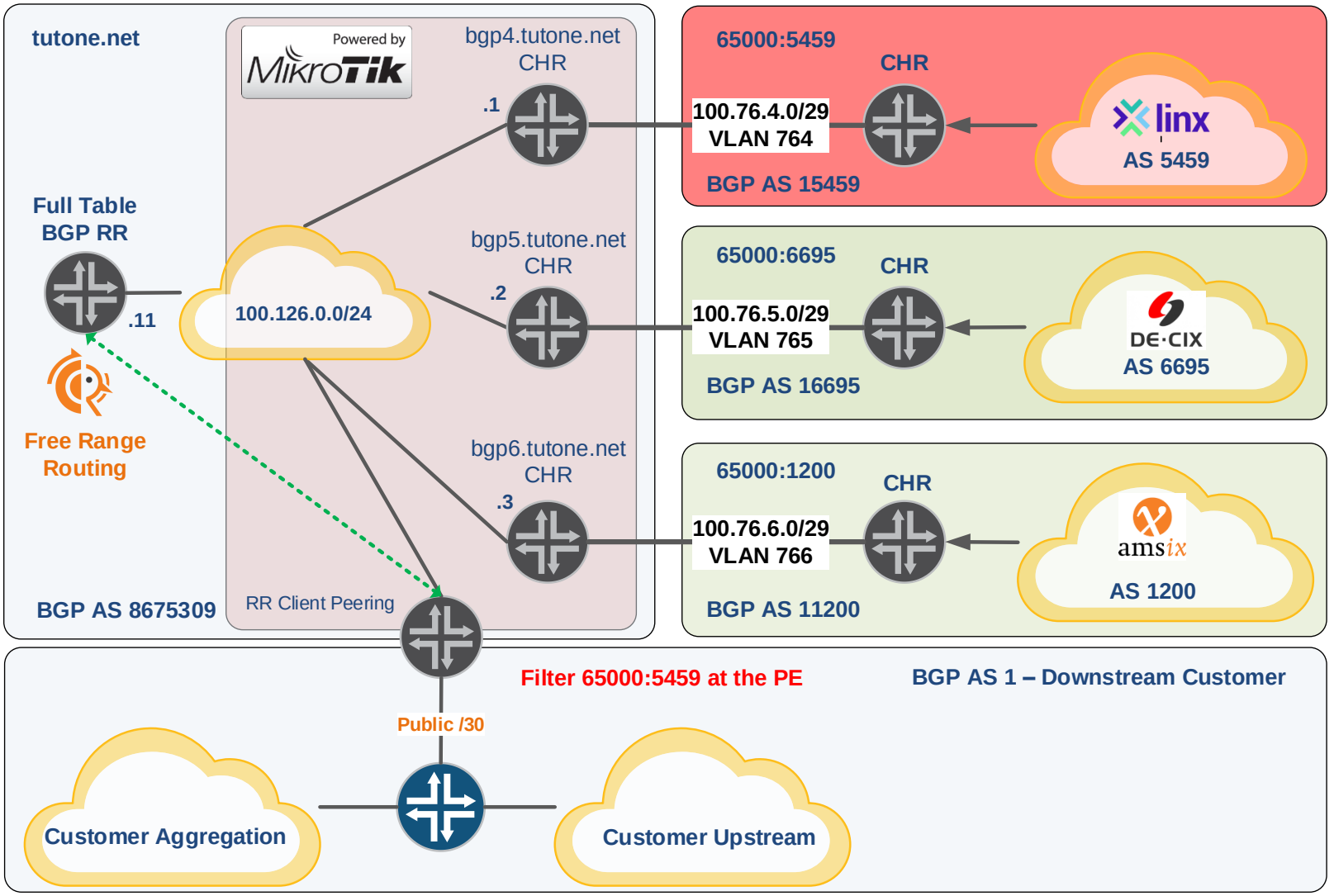
Design: BGP RR design – Customer peering

Full table RR to provide IP Transit



Design: BGP RR design – Customer peering

Use BGP Communities to select upstreams



Questions??