IP ArchiTechs

MANAGED SERVICES

www.iparchitechs.com 1-855-MIKROTIK

Alta disponibilidad con BGP en Routers MikroTik

PRESENTADO POR:

ABRAHAM RIOS
NETWORK ENGINEER

Perfil: **Acerca de Abraham Rios**

- Puertorriqueño viviendo en Florida
- 8 años en la industria
- Diseño e implementación de redes a nivel core
- Diseño y implementaciónes en redes wifi para hotels, parqueos de RV, areas de acampar y eventos temporales
- Diseño y implementaciónes (carrier grade) de torres

Profile: **Acerca de IP ArchiTechs**

IP ArchiTechs Managed Services

- Expositor en el MUM 2020
- El primer Centro de Asistencia Técnica (TAC) de MikroTik para Carriers de primer nivel.
  - Tres niveles de ingenieros de soporte
  - +1-855-MIKROTIK (inglés / Español) ó support.iparchitechs.com (Inglés / Español)
- Servicios Proactivos de Monitoreo/Tickets/Control de Cambios/IPAM
- Ingeniería de redes para Carries desde el Primer nivel.
- Experiencia en diseños para implementaciones de más de 10,000 nodos
  - Hemos facturado miles de horas de soporte para equipos MikroTik en 6 continentes

Objetivos

- Identificar casos de uso para BGP proveyendo tablas de rutas para tránsito.
- Discutir un diseño práctico con un número pequeño de enlaces con BGP.
- Discutir aplicaciones a gran escala con varios proveedores de enlaces usando BGP
- Uso de CHR en VMware ESXi con 10Gbps o más de tráfico

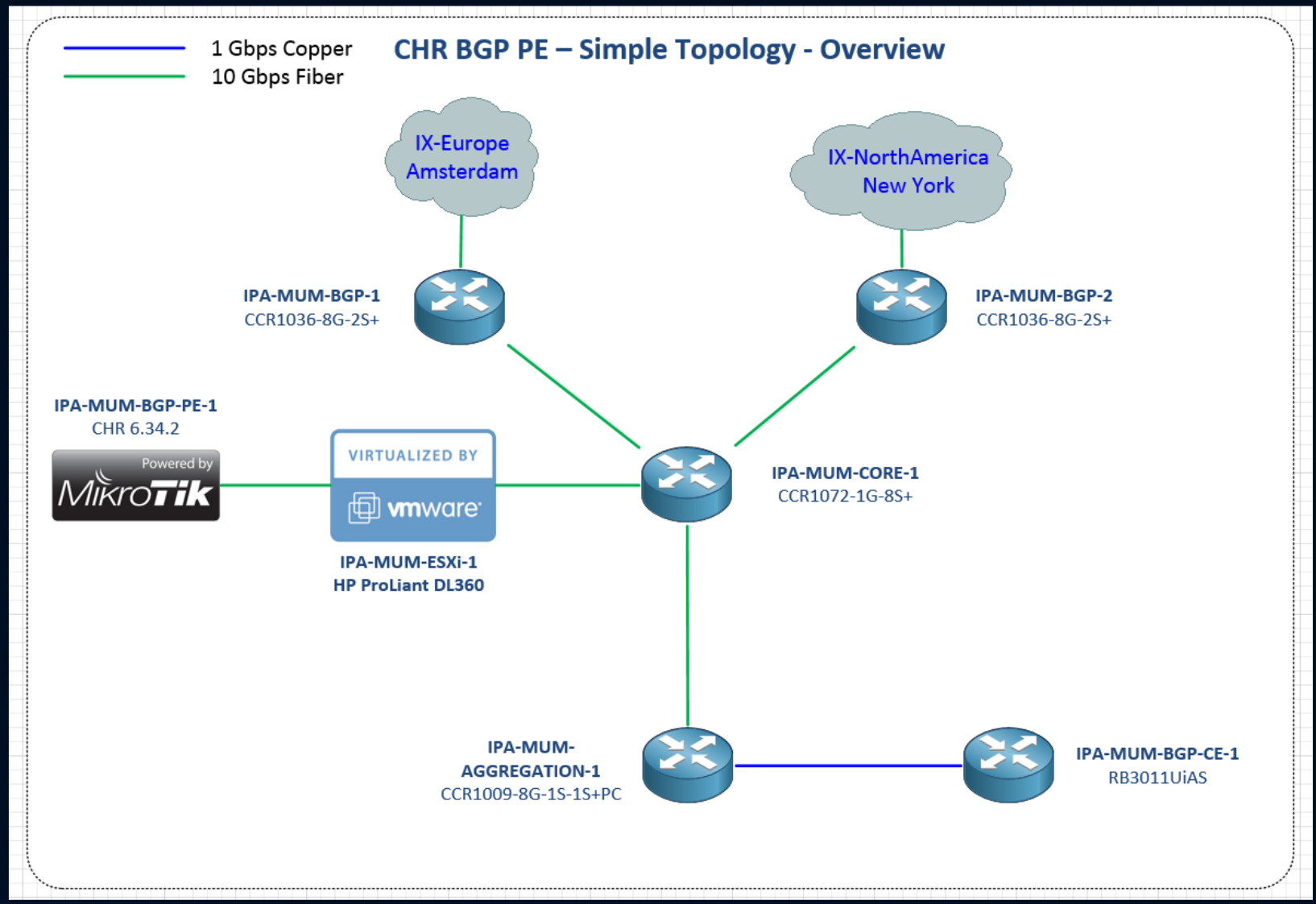
Glosario de Virtualización

- Hipervisor: Es un software, firmware o hardware que permite crear y utilizar máquinas virtuales.
- Tarjeta de red paravirtualizada: Es cuando el hipervisor no necesita emular una tarjeta virtual, quitando el procesamiento extra que conlleva un emulador.
- vSwitch: Es un switch virtual en el hipervisor que maneja las VLANs y la comunicación interna entre máquinas virtuales

Aplicación práctica: ¿Qué problema estamos tratando de resolver?

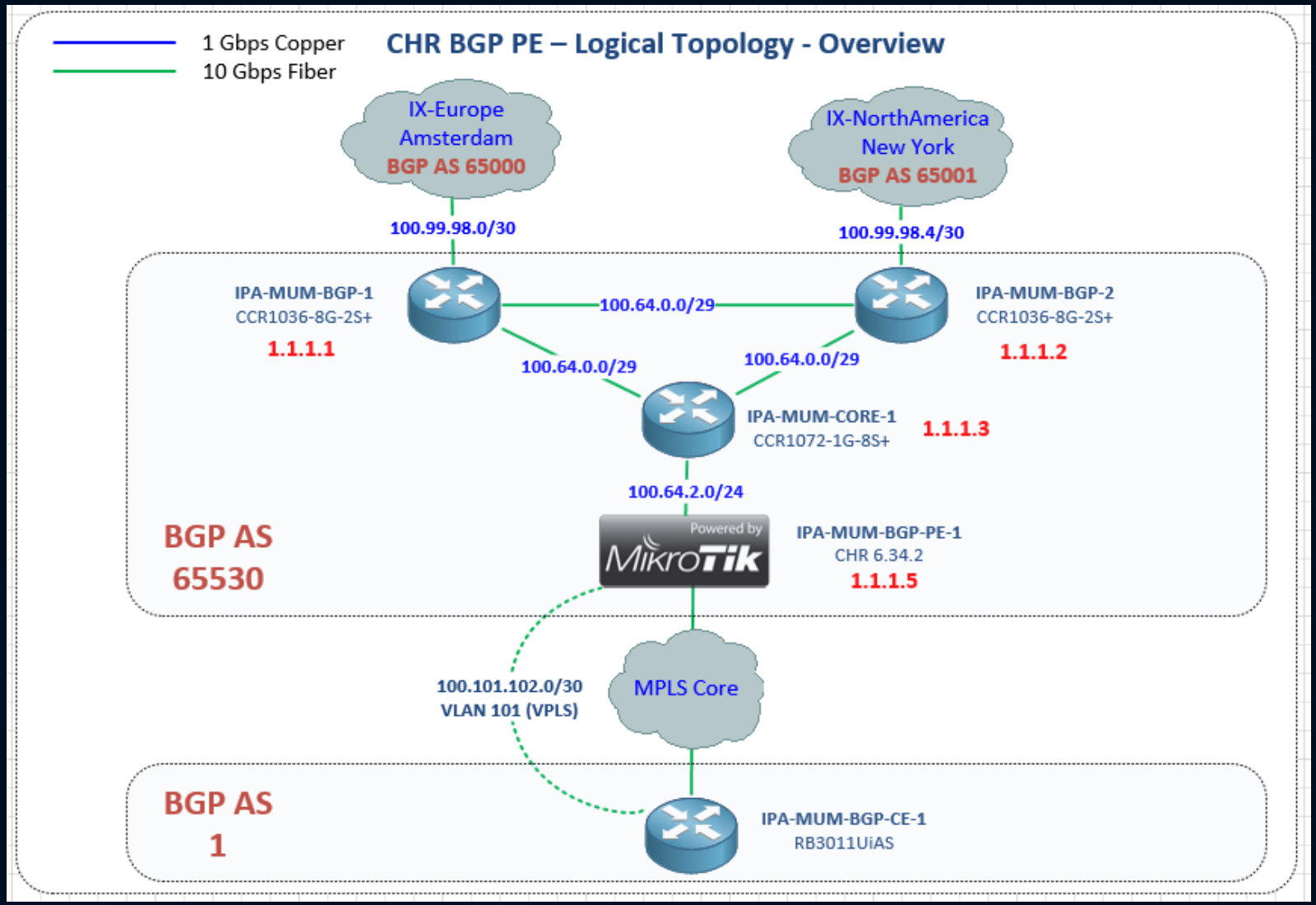
- Utilización de recursos: RouterOS utiliza un solo core para BGP, que puede resultar en un cuello de botella cuando haya muchas conexiones (peering) y rutas. CHR puede correr en un mejor procesador y hacer un uso más eficiente de ese único core.
- Escalabilidad: Si es necesario crecer en capacidad de procesamiento se pueden instalar tantos routers virtuales como sea necesario
- Costo: Al escalar virtualmente la compra de hardware no es tan seguido como escalando por hardware, y si se escala sólo por el BGP se desperdician recursos en un router físico.

Diagrama de red



Diseño Lógico

PE= Provider Edge Router
CE= Customer Edge Router



Usando VPLS para proveer conectividad capa 2 para el tráfico

- ¿Por qué no publicar la tabla completa de BGP?
- Utilización de recursos:
  - La memoria del router afecta el tamaño que la tabla puede tener
  - Rendimiento: Convergencia es mucho más lento que una máquina virtual dedicada
- Seguridad / aislamiento
  - Permite segregar el tráfico completamente a nivel de la red de transporte sin revelar el core MPLS
  - Permite el aislamiento de más clientes utilizando VLANs
- EoIP es una alternativa si MPLS no está disponible

Enlace 1: IX Europa, Amsterdam

- Tabla BGP completa (cortesía de www.stubarea51.com)

```
[admin@IPA-MUM-BGP-1] > routing bgp peer print status
Flags: X - disabled, E - established
 0 E name="IX-Europe-Amsterdam" instance=default remote-address=100.99.98.1 remote-as=65000 tcp-md5-key=""
     nexthop-choice=default multihop=no route-reflect=no hold-time=30m keepalive-time=3m ttl=default in-filter=""
     out-filter=BGP-OUT address-families=ip default-originate=never remove-private-as=no as-override=no passive=no
     use-bfd=no remote-id=100.99.98.1 local-address=100.99.98.2 uptime=2h5m54s prefix-count=293364 updates-sent=0
     updates-received=3696861 withdrawn-sent=0 withdrawn-received=0 remote-hold-time=30m used-hold-time=30m
     used-keepalive-time=3m state=established

 1 E name="IPA-MUM-BGP-2" instance=default remote-address=1.1.1.2 remote-as=65530 tcp-md5-key="" nexthop-choice=default
     multihop=no route-reflect=no hold-time=3m ttl=default in-filter="" out-filter="" address-families=ip
     update-source=Lo1 default-originate=never remove-private-as=no as-override=no passive=no use-bfd=no
     remote-id=100.99.98.6 local-address=1.1.1.1 uptime=22h53m43s prefix-count=175749 updates-sent=3811350
     updates-received=2329918 withdrawn-sent=3568244 withdrawn-received=2139684 remote-hold-time=3m used-hold-time=3m
     used-keepalive-time=1m refresh-capability=yes as4-capability=yes state=established

 2 E name="IPA-MUM-BGP-PE-1" instance=default remote-address=1.1.1.5 remote-as=65530 tcp-md5-key=""
     nexthop-choice=default multihop=no route-reflect=no hold-time=3m ttl=default in-filter="" out-filter=""
     address-families=ip update-source=Lo1 default-originate=never remove-private-as=no as-override=no passive=no
     use-bfd=no remote-id=1.1.1.5 local-address=1.1.1.1 uptime=1d1h43m33s prefix-count=2 updates-sent=3936313
     updates-received=2 withdrawn-sent=3689815 withdrawn-received=0 remote-hold-time=3m used-hold-time=3m
     used-keepalive-time=1m refresh-capability=yes as4-capability=yes state=established
```

Enlace 2: IX Norte América, Nueva York

- Tabla BGP completa (cortesía de www.stubarea51.com)

```
[admin@IPA-MUM-BGP-2] > routing bgp peer print status
Flags: X - disabled, E - established
 0 E name="IPA-MUM-BGP-1" instance=default remote-address=1.1.1.1 remote-as=65530 tcp-md5-key="" nexthop-choice=default multihop=no
     route-reflect=no hold-time=3m ttl=default in-filter="" out-filter="" address-families=ip update-source=Lo1
     default-originate=never remove-private-as=no as-override=no passive=no use-bfd=no remote-id=1.1.1.1 local-address=1.1.1.2
     uptime=23h30m45s prefix-count=334050 updates-sent=2426573 updates-received=4004902 withdrawn-sent=2338312
     withdrawn-received=3622834 remote-hold-time=3m used-hold-time=3m used-keepalive-time=1m refresh-capability=yes as4-capability=yes
     state=established

 1 E name="IX-NorthAmerica-NYC" instance=default remote-address=100.99.98.5 remote-as=65001 tcp-md5-key="" nexthop-choice=default
     multihop=no route-reflect=no hold-time=30m keepalive-time=4m15s ttl=default in-filter="" out-filter=BGP-OUT address-families=ip
     default-originate=never remove-private-as=no as-override=no passive=no use-bfd=no remote-id=100.99.98.5 local-address=100.99.98.6
     uptime=21m57s prefix-count=87566 updates-sent=0 updates-received=676879 withdrawn-sent=0 withdrawn-received=0
     remote-hold-time=30m used-hold-time=30m used-keepalive-time=4m15s state=established

 2 E name="IPA-MUM-BGP-PE-1" instance=default remote-address=1.1.1.5 remote-as=65530 tcp-md5-key="" nexthop-choice=default multihop=no
     route-reflect=no hold-time=3m ttl=default in-filter="" out-filter="" address-families=ip update-source=Lo1
     default-originate=never remove-private-as=no as-override=no passive=no use-bfd=no remote-id=1.1.1.5 local-address=1.1.1.2
     uptime=23h30m44s prefix-count=2 updates-sent=2426569 updates-received=2 withdrawn-sent=2338312 withdrawn-received=0
     remote-hold-time=3m used-hold-time=3m used-keepalive-time=1m refresh-capability=yes as4-capability=yes state=established
[admin@IPA-MUM-BGP-2] >
```

Máquina Virtual PE – Rutas BGP

- El router PE toma la tabla completa de cada enlace y le publica las mejores rutas al router CE

```
Flags: X - disabled, E - established
 0 E name="IPA-MUM-BGP-1" instance=default remote-address=1.1.1.1
     remote-as=65530 tcp-md5-key="" nexthop-choice=default multihop=no
     route-reflect=no hold-time=3m ttl=255 in-filter="" out-filter=""
     address-families=ip update-source=Lo1 default-originate=never
     remove-private-as=no as-override=no passive=no use-bfd=no
     remote-id=1.1.1.1 local-address=1.1.1.5 uptime=1d2h27m6s
     prefix-count=338092 updates-sent=2 updates-received=4144975
     withdrawn-sent=0 withdrawn-received=3754874 remote-hold-time=3m
     used-hold-time=3m used-keepalive-time=1m refresh-capability=yes
     as4-capability=yes state=established

 1 E name="IPA-MUM-BGP-2" instance=default remote-address=1.1.1.2
     remote-as=65530 tcp-md5-key="" nexthop-choice=default multihop=no
     route-reflect=no hold-time=3m ttl=255 in-filter="" out-filter=""
     address-families=ip update-source=Lo1 default-originate=never
     remove-private-as=no as-override=no passive=no use-bfd=no
     remote-id=100.99.98.6 local-address=1.1.1.5 uptime=23h37m15s
     prefix-count=96005 updates-sent=2 updates-received=2449590
     withdrawn-sent=0 withdrawn-received=2338342 remote-hold-time=3m
     used-hold-time=3m used-keepalive-time=1m refresh-capability=yes
     as4-capability=yes state=established

 2 E name="IPA-MUM-BGP-CE" instance=default remote-address=100.101.102.2
-- [Q quit|D dump|down]
```

Router CE

- La tabla entera entre los 2 enlaces se publica al cliente sin cargarla enteramente por la red

```
[admin@IPA-MUM-BGP-CE-1] > routing bgp peer print status
Flags: X - disabled, E - established
 0 E name="IPA-MUM-BGP-PE-1" instance=default remote-address=100.101.102.1
     remote-as=65530 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=1.1.1.5
     local-address=100.101.102.2 uptime=23h41m28s prefix-count=400335
     updates-sent=1 updates-received=6354378 withdrawn-sent=0
     withdrawn-received=4599799 remote-hold-time=3m used-hold-time=3m
     used-keepalive-time=1m refresh-capability=yes as4-capability=yes
     state=established
```

Router Core – Número de rutas

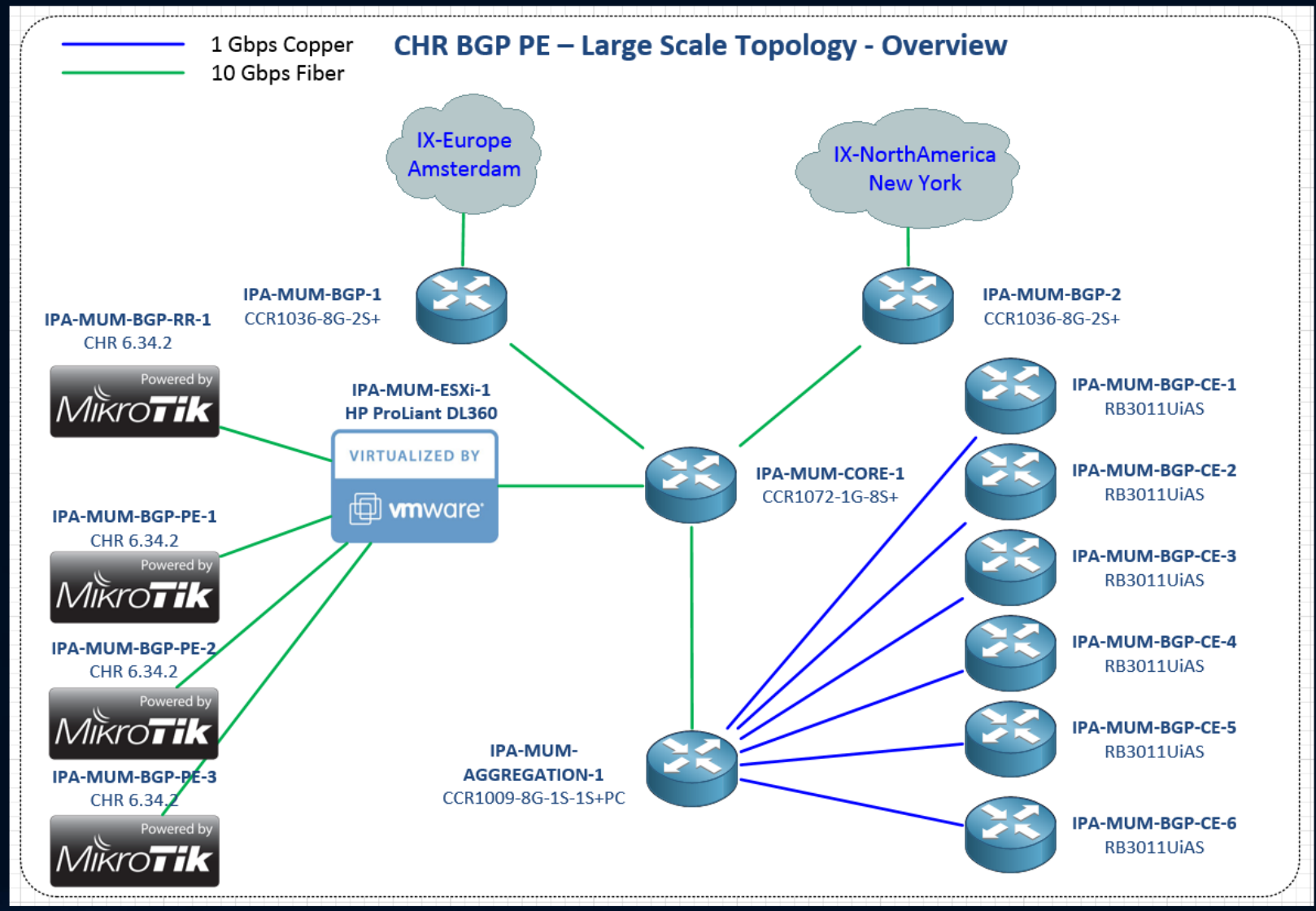
```
[admin@IPA-MUM-CORE-1] > ip route print
Flags: X - disabled, A - active, D - dynamic, C - connect, S - static, r - rip,
 #      DST-ADDRESS        PREF-SRC        GATEWAY            DISTANCE
 0 ADS  0.0.0.0/0                          10.254.253.1              0
 1 ADo  1.1.1.1/32                         100.64.0.10             110
 2 ADo  1.1.1.2/32                         100.64.0.18             110
 3 ADC  1.1.1.3/32         1.1.1.3         Lo1                       0
 4 ADo  1.1.1.4/32                         100.64.0.26             110
 5 ADo  1.1.1.5/32                         100.64.2.11             110
 6 ADC  10.254.253.0/24    10.254.253.105  ether1                    0
 7 ADo  100.64.0.0/29                      100.64.0.10             110
                                           100.64.0.18
 8 ADC  100.64.0.8/29      100.64.0.9      sfp-sfpplus7              0
 9 ADC  100.64.0.16/29     100.64.0.17     sfp-sfpplus4              0
10 ADC  100.64.0.24/29     100.64.0.25     sfp-sfpplus5              0
11 ADC  100.64.2.0/24      100.64.2.1      vlan2000                  0
12 ADo  100.99.98.0/30                    100.64.0.10             110
13 ADo  100.99.98.4/30                    100.64.0.18             110
[admin@IPA-MUM-CORE-1] > ip route print count-only
14
```

- El router CE recive más de 400 mil rutas
- La red core solo recive una tabla pequeña que mejora la velocidad de convergencia y el rendimiento
- ¡La red core tiene 14 rutas para transporter más de 400 mil!

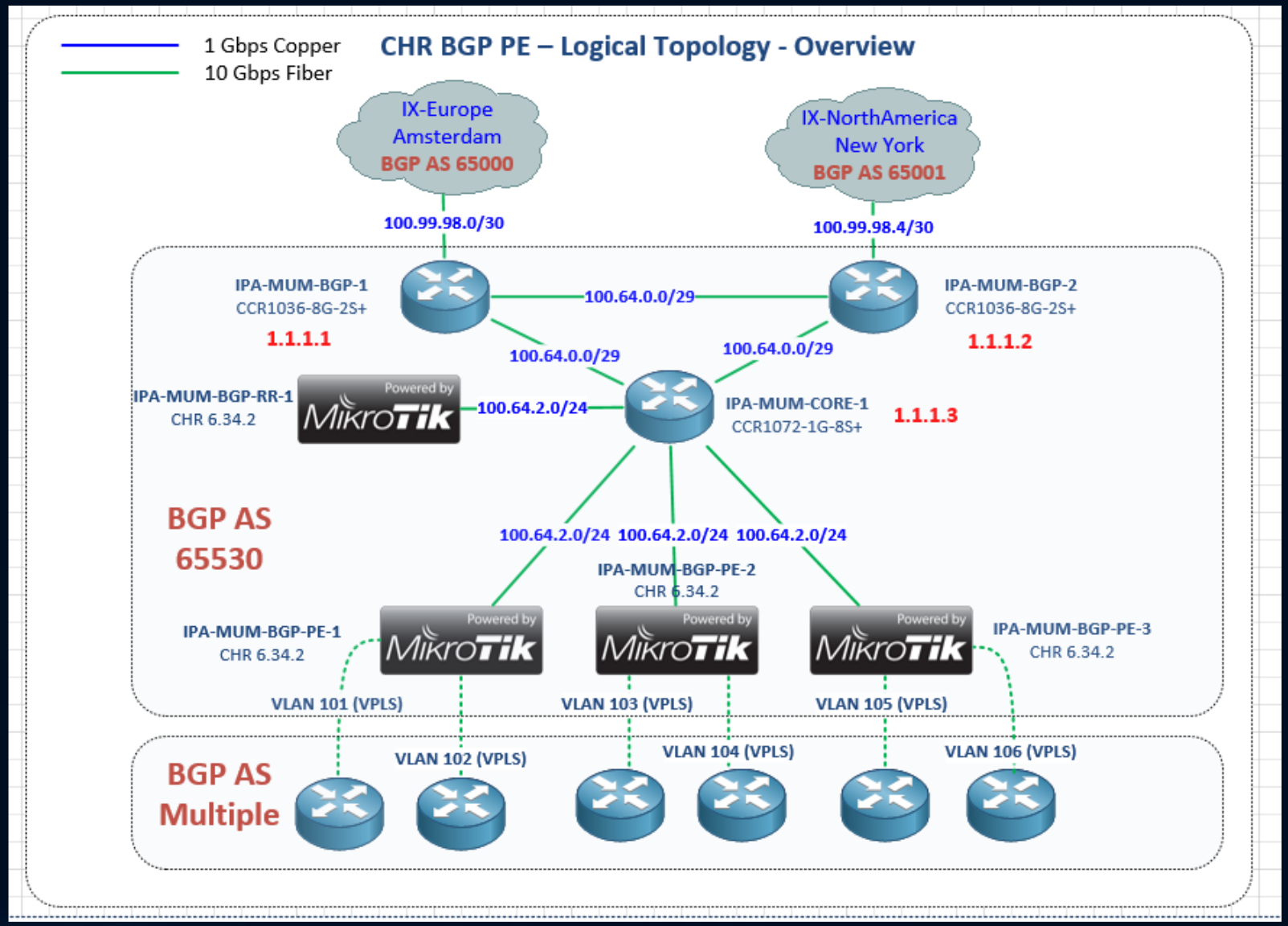
Escalando BGP

- ¿Cómo escalar BGP?
  - Agregando más “peerings”. Dependiendo del hardware, se pueden agregar de 5 a 10 tablas de peering por cada Router virtual PE
  - Agregar más routers PE. Esto permite más peers de tránsito por cliente. Cada router debe conectarse a los 2 edge routers
  - Route Reflection. Se utiliza para alimentar varios routers PE

Diseño de red a gran escala



Diseño lógico



Escalabilidad con Route Reflection

- Reduce el número de peerings en el border router y por lo tanto utilización de recursos
- RR no cambia el siguiente salto aprendido por defecto, así que puede estar fuera de ruta y no necesita llevar tráfico. Típicamente se implementa con OSPF/BGP y loopback peering.
- Varios RR pueden hacer peering con BGP al border router para distribuir la utilización de recursos
- Un CCR puede transmitir un millón de rutas a clientes RR en menos de 2 minutos

Escalabilidad con Route Reflection

- El Cisco ASR1000v tiene todo un menú de licenciamiento, muy caro, para operadores
- MikroTik CHR tiene un potencial mayor con una fracción del costo

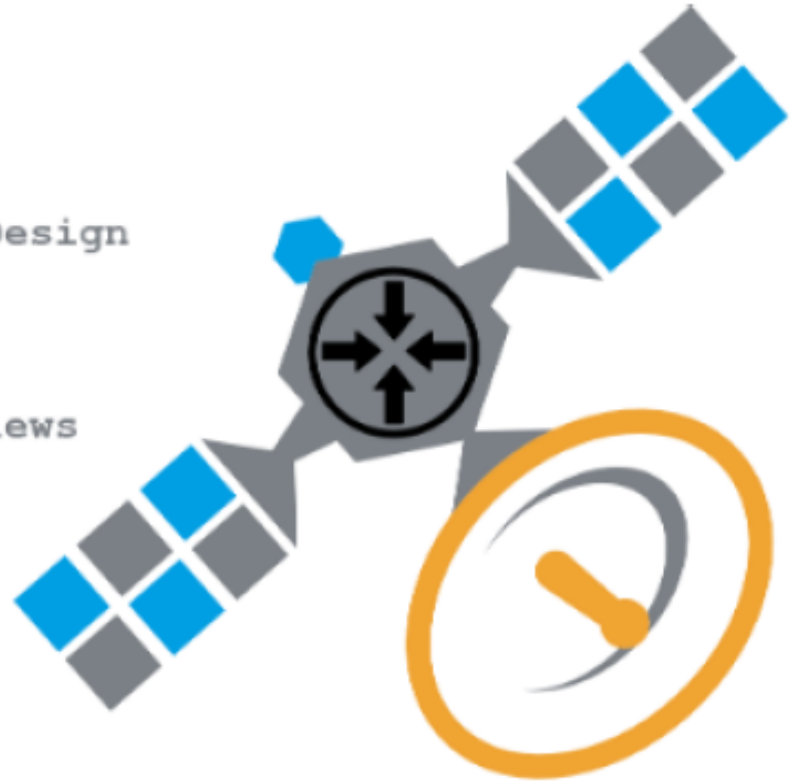
Specifications	MikroTik CHR	Cisco ASR1000V
IP Throughput	10 Gbps +	Limited to 10 Gbps
MPLS Throughput	10 Gbps +	Limited to 5 Gbps
IPSEC license	Included	Separate license ($6500)
Firewall license	Included	Separate license ($3000)
License	1 Gbps, 10 Gbps, Unlimited	1 Gbps, 5 Gbps, 10 Gbps
Cost	$250 for Unlimited	Upwards of $30,000 for up to 10 Gbps with advanced services

/Network Engineering and Design

/MikroTik Integration

/Hardware testing and reviews

/News and Views



```
/routing ospf area add\
area-id=0.0.0.51 type=stub
```

STUBAREA51.NET

IP ArchiTechs
MANAGED SERVICES

1-855-MIKROTIK

www.iparchitechs.com

Ven a nuestro stand por unos Tik Tacs!

- MikroTik Tac = MikroTik Technical Assistance Center

24/7/365 MikroTik TAC | Nationwide Private 4G LTE MPLS | Proactive Network Monitoring | Design / Engineering / Operations

IP ArchiTechs
MANAGED SERVICES

1-855-MIKROTIK

www.iparchitechs.com

24/7/365 MikroTik TAC | Nationwide Private 4G LTE MPLS | Proactive Network Monitoring | Design / Engineering / Operations

- El contenido de esta presentación estará disponible en mum.iparchitechs.com
- Por favor visítanos en nuestro stand
- Email: Abraham.Rios@iparchitechs.com
- Teléfono: +1(303)728-1307 o +1 (303) 590-9946
- Web: www.iparchitechs.com
- GRACIAS POR SU TIEMPO Y QUE DISFRUTEN EL MUM!!!