Michael Takeuchi, MTCINE

BGP, "**Inter-Net**"working Best Path Seekers

15 June 2019, Yangon
MikroTik User Meeting Myanmar

Hello, I am **Michael Takeuchi**

MikroTik Certified Engineer & Consultant from Jakarta, Indonesia
IP Core Engineer in Multiple Countries and Multiple ASN/ISP/Networks

https://www.linkedin.com/in/michael-takeuchi
https://www.facebook.com/mict404
michael@takeuchi.id

Agenda

- What is BGP?
- BGP Best Path Selection Algorithm
- BGP Network Route Propagation Graph
- Challenges as IP Core Network Engineer (Confide :P)
- Study Case
  - Looking for Best Path Study Case
  - Metric Manipulation
- Conclusion
- References

BGP?

- **Border Gateway Protocol** (**BGP**) is a standardized exterior gateway protocol designed to exchange routing and reachability information among autonomous systems (AS) on the Internet.
- The **Border Gateway Protocol** makes routing decisions based on paths, network policies, or rule-sets configured by a network administrator and is involved in making core routing decisions.

- Wikipedia, https://en.wikipedia.org/wiki/Border_Gateway_Protocol

BGP?

- So, basically Border Gateway Protocol (BGP) is a routing protocol that used for **Inter**-**Net**working now
- BGP tell you how to reach everyone else in the Internet
- And Vice Versa, BGP tell everyone else in the Internet how to reach you
- BGP usually configured in Internet Service Provider router
- As a End-User, you doesn’t need to configure BGP
- BGP has two different types:
  - iBGP - BGP Peer within single AS Number
  - eBGP - BGP Peer with different AS Number

BGP Best Path Selection Algorithm

- With the full Internet BGP routing table being upward of 700K+ routes and with a BGP router having the potential to be receiving multiple copies of that routing table from multiple providers, it has to have some way to compare those multiple BGP routing tables and select only the best route to go into the IP routing table on the router. It uses the BGP Best Path Selection Algorithm to do this.
- You should note that MikroTik and Cisco BGP routers have **weight** as the first criteria in the table **where other brands do not.**
- Best path algorithm compares routes received by a **single BGP instance**. Routes installed by different BGP instances are compared by the general algorithm, i.e. route distances are compared and the route with lower distance is preferred.

- Wiki MikroTik, https://wiki.mikrotik.com/wiki/Manual:BGP_Best_Path_Selection_Algorithm

BGP Best Path Selection Algorithm

The most common metric parameter that used to manipulate BGP in RouterOS is:

- Local-Pref
  - Mostly used in iBGP networks, higher Local-Pref is better
- AS-Path
  - Mostly used in eBGP networks, shorter as-path is better

- But all of BGP Best Path Algorithm is not as simple like that ☺ (see on next slide)

BGP Best Path Selection Algorithm (cont'd)

1. Router is ignoring received path if the route is not valid. Route is valid if:
   - NEXT_HOP of the route is valid and reachable
   - AS_PATH received from external peers does not contain the local AS
   - route is not rejected by routing filters
2. The first path received is automatically considered 'best path'. Any further received paths are compared to first received to determine if the new path is better.
3. Prefer the path with the highest WEIGHT.
   WEIGHT parameter is local to the router on which it is configured. A route without assigned WEIGHT have a default value of 0.
4. Prefer the path with the highest LOCAL_PREF. It is used only within an AS.
   A path without LOCAL_PREF attribute have a value of 100 by default.
5. Prefer the path with the shortest AS_PATH. (skipped if ignore-as-path-len set to yes)
   Each AS_SET counts as 1, regardless of the set size. The AS_CONFED_SEQUENCE and AS_CONFED_SET are not included in the AS_PATH length.
6. Prefer the path that was locally originated via aggregate or BGP network

BGP Best Path Selection Algorithm (cont'd)

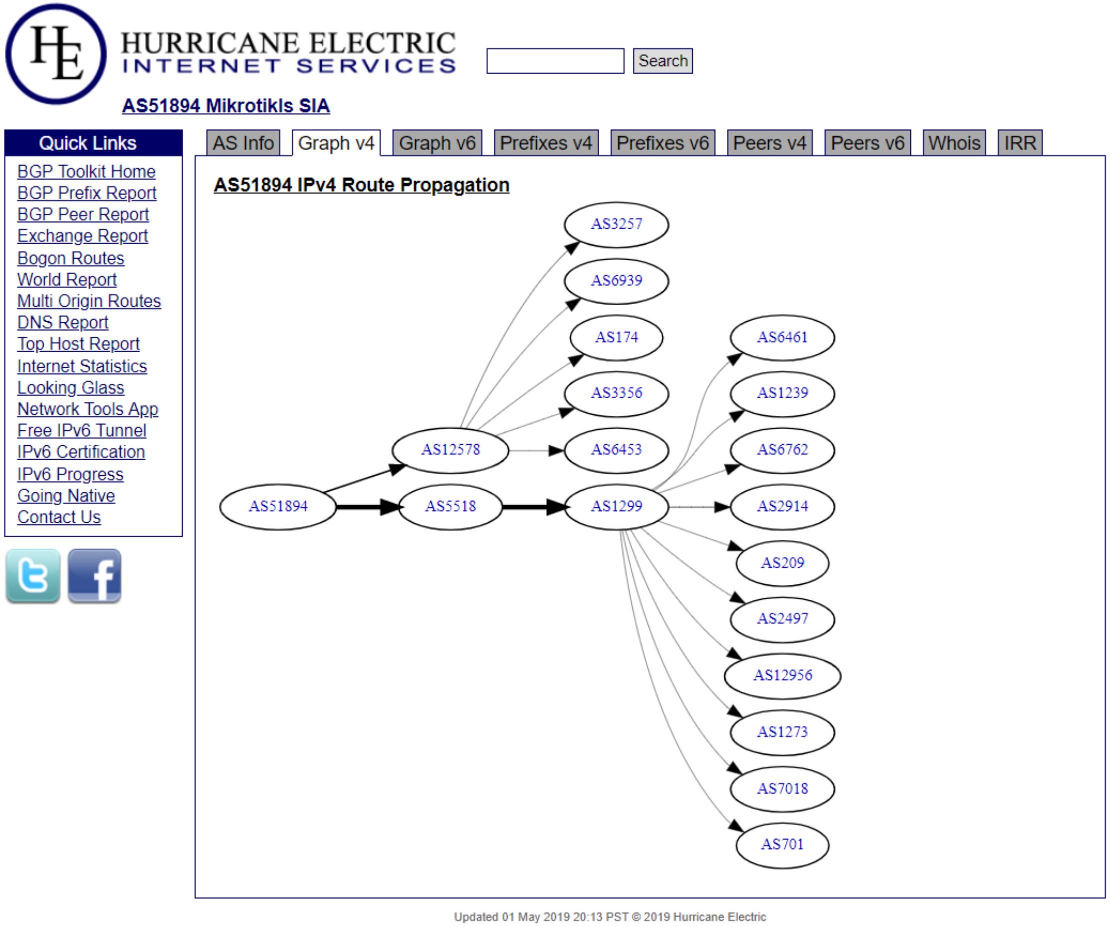
7. Prefer the path with the lowest ORIGIN type.
   Interior Gateway Protocol (IGP) is lower than Exterior Gateway Protocol (EGP), and EGP is lower than INCOMPLETE in other words IGP < EGP < INCOMPLETE
8. Prefer the path with the lowest multi-exit discriminator (MED).
   The router compare MED attribute only for paths that have the same neighboring (leftmost) AS. Paths without explicit MED value are treated as with MED of 0
9. Prefer eBGP over iBGP paths
10. Prefer the route that comes from the BGP router with the lowest router ID. If a route carries the ORIGINATOR_ID attribute, then the ORIGINATOR_ID is used instead of router ID.
11. Prefer the route with the shortest route reflection cluster list. Routes without a cluster list are considered to have a cluster list of length 0.
12. Prefer the path that comes from the lowest neighbor address

BGP Best Path Selection Algorithm (summary)

- Pretty complex
- Not selecting route by Smallest Latency
- Many metric parameter or we can said it **attributes**
- But, BGP has:
  - Community Tagging
    - Usually used for blackholing to drop DOS/DDoS attack
  - Autonomous System
    - Easier identification
  - MP-BGP (Multi-Protocol BGP)
    - Usually used for MPLS VPN

BGP Network Route Propagation Graph

- You can use a tool from https://bgp.he.net/ to take look your ISP BGP Route Propagation graph, so can know that where your packet forwarded is



AS51894 Mikrotikls SIA

AS51894 IPv4 Route Propagation

Updated 01 May 2019 20:13 PST © 2019 Hurricane Electric

Challenges as IP Core Network Engineer

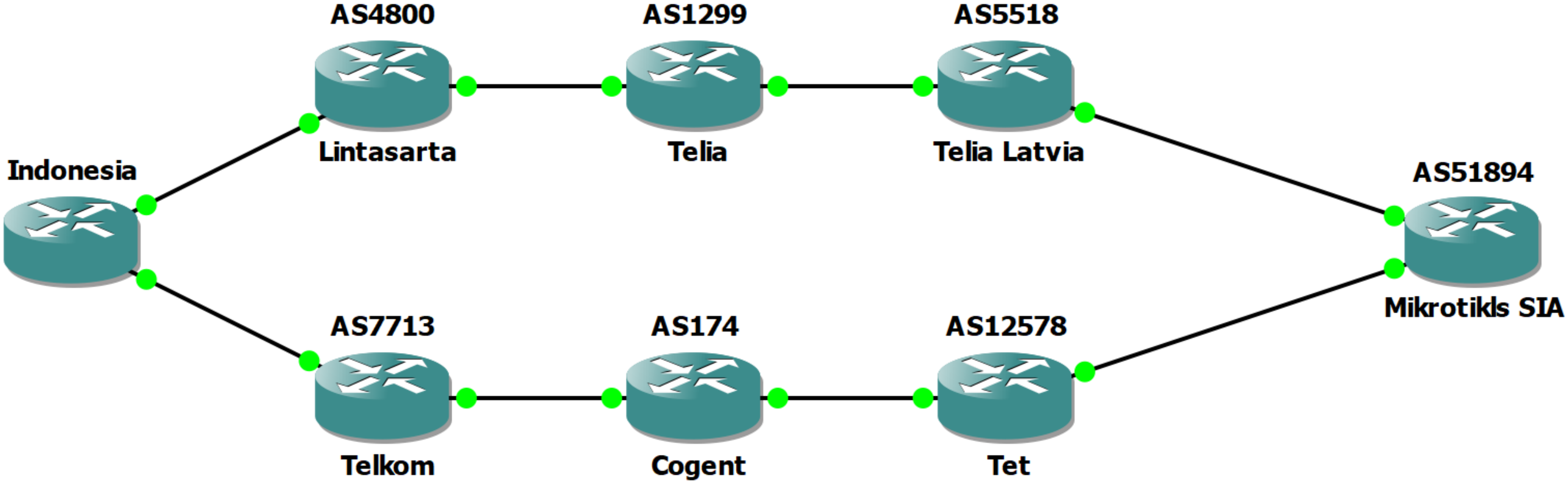
- Route Leak
- BGP Hijacking
- Backbone Issue
- 24 hours standby
- Massive DDoS Attack
- Broadcast on Internet Exchange
- High Latency & Packet Lost Issue
- Misconfiguration Ruins Everything
- Internet has Fickle & Complex Topology
- Coordination with Upstream of Upstream
- Memory Exhaustion and Router Got Hang
- Problem on the Network that We Didn’t Manage it
- And many more… Please comment if I missed something common ☺

Study Case – Looking for Best Path

- One day I want access 159.148.147.0/24 from Indonesia, I have 2 path

Path A via AS4800 – AS1299 – AS5518 – AS51894

Path B via AS7713 – AS174 – AS12578 – AS12578 – AS51894 – AS51894 – AS51894



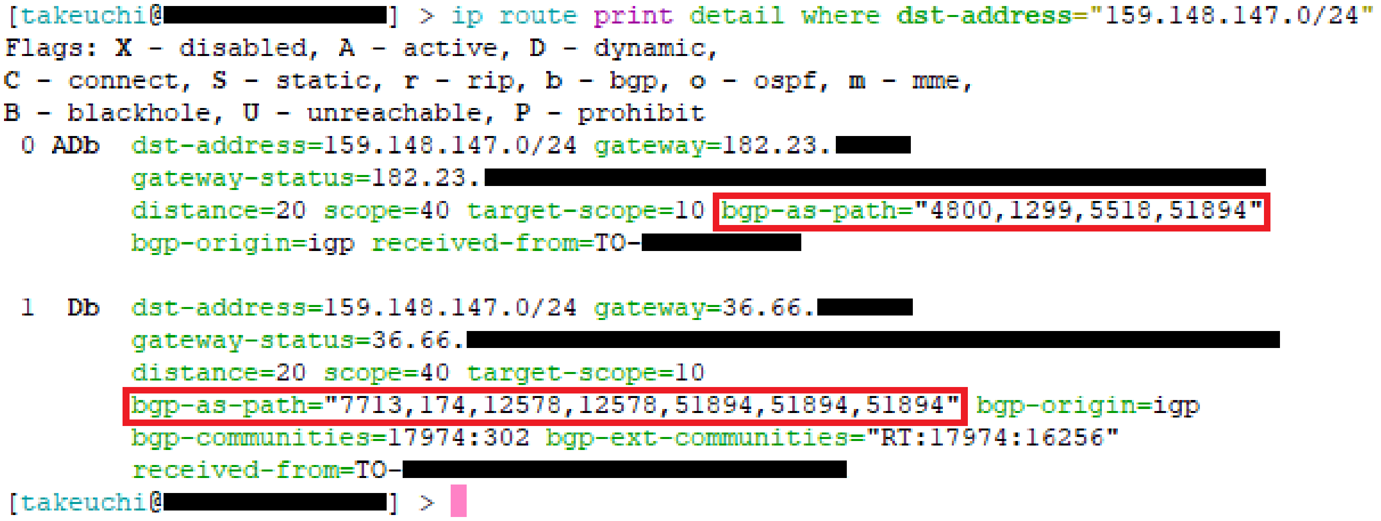
Which one better?

Study Case – Looking for Best Path

- **The Answer is Path A**, because Path A has shorter AS-Path and Path B Metric has been manipulated (BGP Prepend), so the AS-Path of Path B are become longer
- Why AS-Path? Because we talk about eBGP between AS Number and we didn't receive Weight metric or another parameter that we can consider from received IP route that we got from Upstream (AS4800 and AS7713)

Path A via AS4800 – AS1299 – AS5518 – AS51894

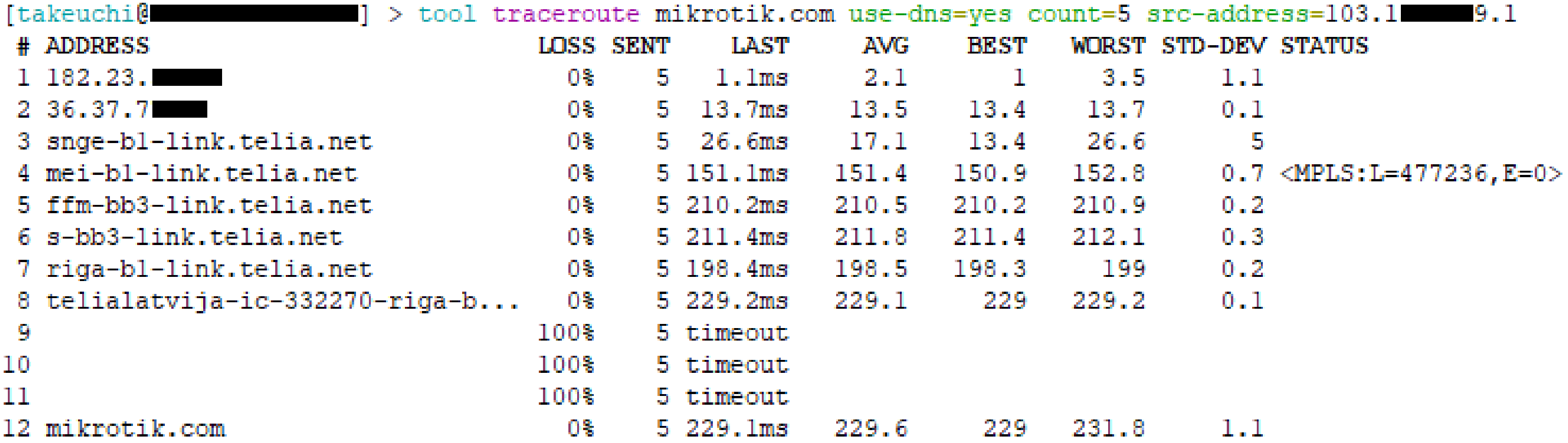
Path B via AS7713 – AS174 – AS12578 – AS12578 – AS51894 – AS51894 – AS51894



```
[takeuchi@[redacted]] > ip route print detail where dst-address="159.148.147.0/24"
Flags: X - disabled, A - active, D - dynamic,
C - connect, S - static, r - rip, b - bgp, o - ospf, m - mme,
B - blackhole, U - unreachable, P - prohibit
 0 ADb  dst-address=159.148.147.0/24 gateway=182.23.[redacted]
        gateway-status=182.23.[redacted]
        distance=20 scope=40 target-scope=10 bgp-as-path="4800,1299,5518,51894"
        bgp-origin=igp received-from=TO-[redacted]

 1  Db  dst-address=159.148.147.0/24 gateway=36.66.[redacted]
        gateway-status=36.66.[redacted]
        distance=20 scope=40 target-scope=10
        bgp-as-path="7713,174,12578,12578,51894,51894,51894" bgp-origin=igp
        bgp-communities=17974:302 bgp-ext-communities="RT:17974:16256"
        received-from=TO-[redacted]
[takeuchi@[redacted]] >
```

Study Case – Looking for Best Path

- Why Path B have few ASN that has been multiplied and make the AS-Path longer?
- Maybe Path A has better quality that AS owner want like latency, bandwidth capacity and any other else because when some ASN got multiplied, it’s must be manipulated with BGP Prepend parameter
- After all Path A only transit with 2 Provider to reach Mikrotikls SIA (Lintasarta & Telia), Telia Global and Telia Latvia is still one provider with different AS Number



```
[takeuchi@[redacted]] > tool traceroute mikrotik.com use-dns=yes count=5 src-address=103.1[redacted]9.1
 # ADDRESS                          LOSS SENT    LAST     AVG    BEST   WORST STD-DEV STATUS
 1 182.23.[redacted]                  0%    5   1.1ms     2.1       1     3.5     1.1
 2 36.37.7[redacted]                  0%    5  13.7ms    13.5    13.4    13.7     0.1
 3 snge-b1-link.telia.net             0%    5  26.6ms    17.1    13.4    26.6       5
 4 mei-b1-link.telia.net              0%    5 151.1ms   151.4   150.9   152.8     0.7 <MPLS:L=477236,E=0>
 5 ffm-bb3-link.telia.net             0%    5 210.2ms   210.5   210.2   210.9     0.2
 6 s-bb3-link.telia.net               0%    5 211.4ms   211.8   211.4   212.1     0.3
 7 riga-b1-link.telia.net             0%    5 198.4ms   198.5   198.3     199     0.2
 8 telialatvija-ic-332270-riga-b...   0%    5 229.2ms   229.1     229   229.2     0.1
 9                                  100%    5 timeout
10                                  100%    5 timeout
11                                  100%    5 timeout
12 mikrotik.com                       0%    5 229.1ms   229.6     229   231.8     1.1
```

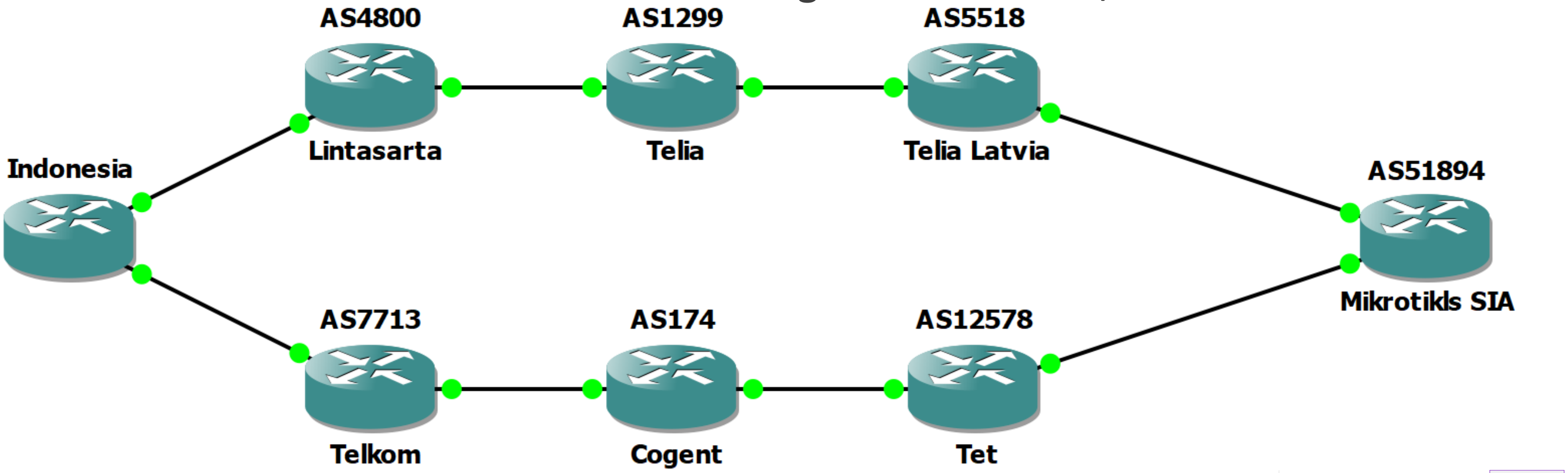
Jakarta – Singapore – Marseille – Frankfurt – Amsterdam – Stockholm – Riga

Study Case – Metric Manipulation

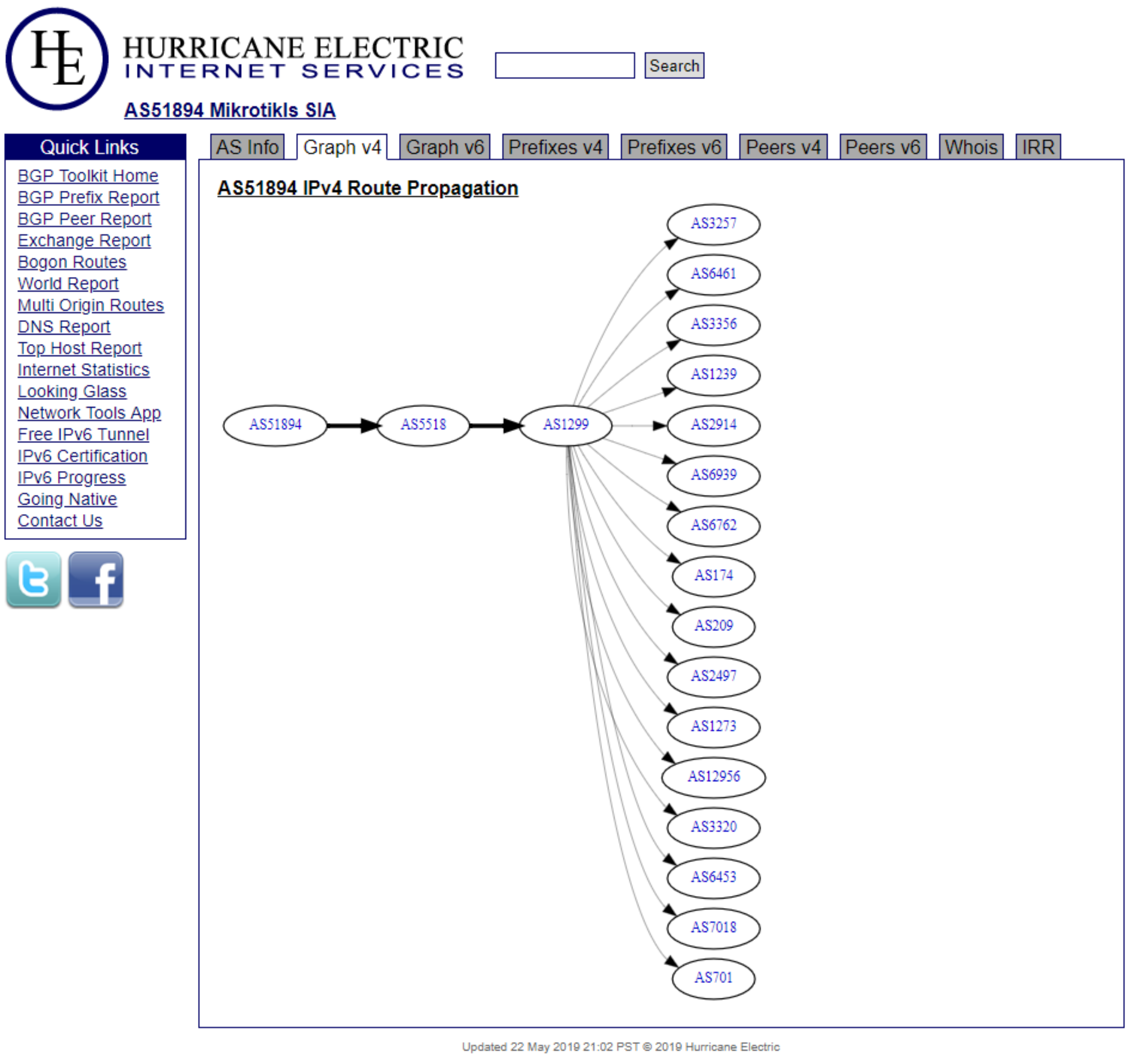
- **Disclaimer** for Metric Manipulation Study Case

1. I will create another virtual environment on GNS3 with the reason that we can see and do traceroute on each hop to make sure the path and metric changes and I will do the demonstration here if we still have enough time
2. The latency and traceroute would be totally different
3. In real case, the configuration is not as simple like what I configured in GNS3

Note for Indonesia ASN is 64510 and coming with 192.0.2.0/24 as the Prefix



Study Case – Metric Manipulation



AS51894 Mikrotikls SIA

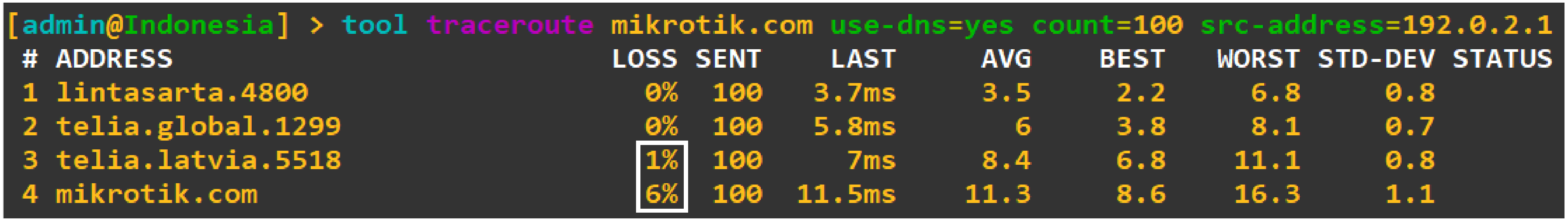
AS51894 IPv4 Route Propagation

Updated 22 May 2019 21:02 PST © 2019 Hurricane Electric

- And the Topology of Mikrotikls already changed
- Tet has dismantled from Mikrotikls upstream
- I can't judge why it can be happened, Internet topology are changed so quick
- That's also why I just create a new virtual environment

Study Case – Metric Manipulation

- One Day Path A has a problem on Telia Latvia to Mikrotikls SIA, we got Packet Lost to Mikrotikls SIA (159.148.147.0/24), we need to re-route the traffic through Path B



```
[admin@Indonesia] > tool traceroute mikrotik.com use-dns=yes count=100 src-address=192.0.2.1
 # ADDRESS                          LOSS SENT    LAST     AVG    BEST   WORST STD-DEV STATUS
 1 lintasarta.4800                    0%  100   3.7ms     3.5     2.2     6.8     0.8
 2 telia.global.1299                  0%  100   5.8ms       6     3.8     8.1     0.7
 3 telia.latvia.5518                  1%  100     7ms     8.4     6.8    11.1     0.8
 4 mikrotik.com                       6%  100  11.5ms    11.3     8.6    16.3     1.1
```

- So we need to manipulate the BGP metric to re-route the traffic through Path B, we can **set-bgp-prepend** through routing filter configuration
- By the way the traceroute was totally different from my previous slide because this is in different environment ☺

Study Case – Metric Manipulation

- in-filter = Manipulate Our Routing Table
- out-filter = Manipulate Neighbor Routing Table

```
/routing filter
add action=accept chain=$in-filter prefix=$destination-prefix set-bgp-prepend=3
add action=accept chain=$out-filter prefix=$source-prefix set-bgp-prepend=6
```

```
/routing bgp peer
set in-filter=$in-filter out-filter=$out-filter [find remote-as="4800"]
```

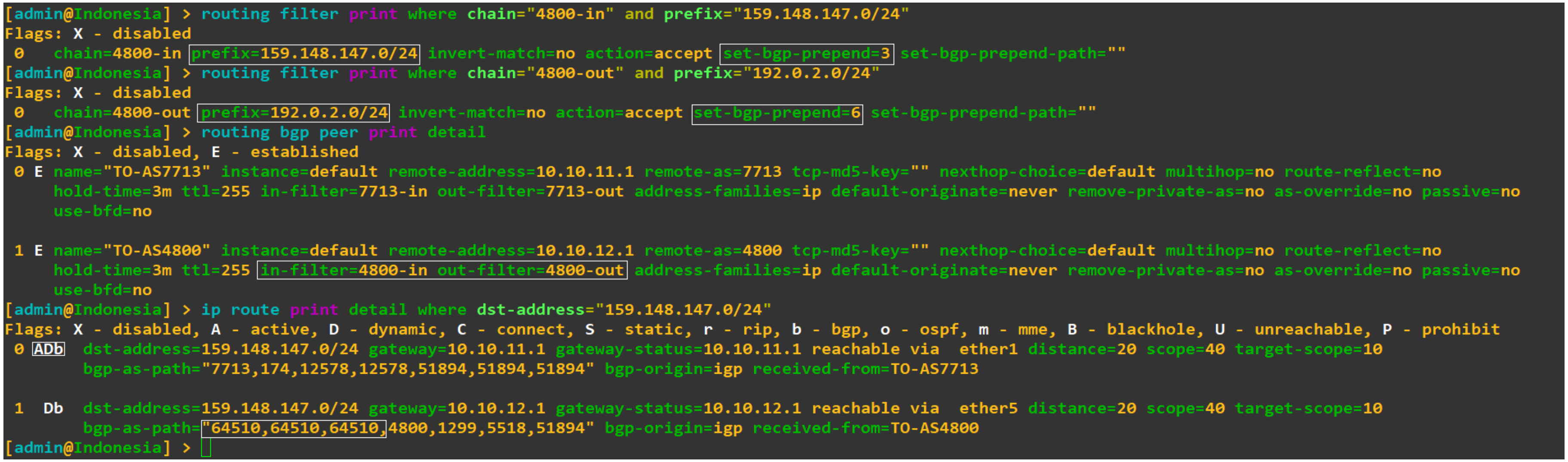
- Remember communication are 2-ways (Transmit and Receive), So:
- We need to manipulate **our routing table** to reach destination via Path B

and

- We need to manipulate **neighbor routing table** to reach us via Path B
- Because Path A have a problem on it, it would be useless if the route back is still use Path A even we go through Path B (remember, 2-ways communication)

Study Case – Metric Manipulation

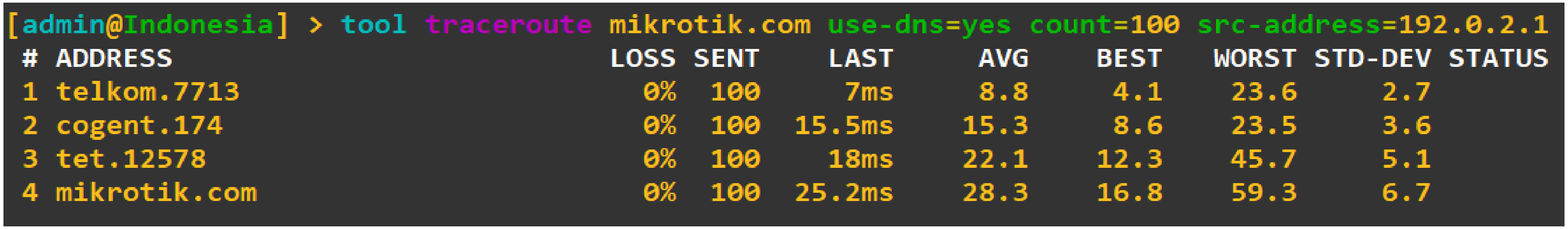
- We do BGP Prepend on Incoming 159.148.147.0/24 through 4800 to make the as-path longer and make our routing table doesn’t choose Path A via AS4800 because the as-path is longer then Path B
- We also do BGP Prepend on Outgoing 192.0.2.0/24 through AS4800 to make the as-path longer, so the everyone knows that if they reach us via AS4800 will have the longer as-path, and will choose Path B rather than Path A to reach us



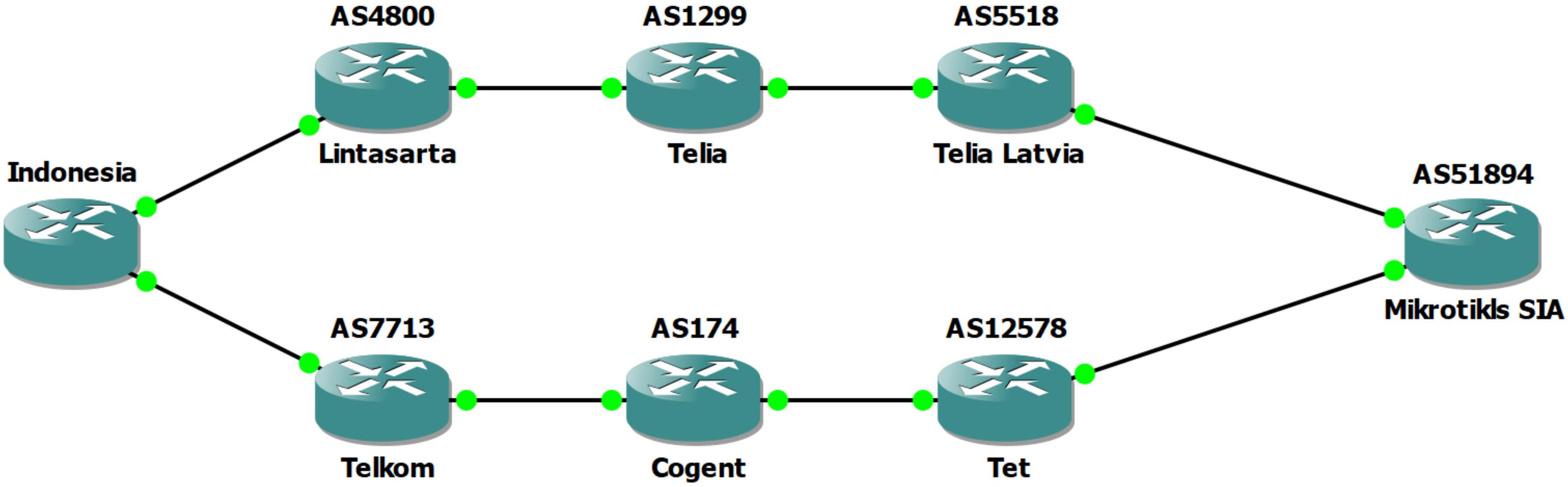
```
[admin@Indonesia] > routing filter print where chain="4800-in" and prefix="159.148.147.0/24"
Flags: X - disabled
 0   chain=4800-in prefix=159.148.147.0/24 invert-match=no action=accept set-bgp-prepend=3 set-bgp-prepend-path=""
[admin@Indonesia] > routing filter print where chain="4800-out" and prefix="192.0.2.0/24"
Flags: X - disabled
 0   chain=4800-out prefix=192.0.2.0/24 invert-match=no action=accept set-bgp-prepend=6 set-bgp-prepend-path=""
[admin@Indonesia] > routing bgp peer print detail
Flags: X - disabled, E - established
 0 E name="TO-AS7713" instance=default remote-address=10.10.11.1 remote-as=7713 tcp-md5-key="" nexthop-choice=default multihop=no route-reflect=no
     hold-time=3m ttl=255 in-filter=7713-in out-filter=7713-out address-families=ip default-originate=never remove-private-as=no as-override=no passive=no
     use-bfd=no

 1 E name="TO-AS4800" instance=default remote-address=10.10.12.1 remote-as=4800 tcp-md5-key="" nexthop-choice=default multihop=no route-reflect=no
     hold-time=3m ttl=255 in-filter=4800-in out-filter=4800-out address-families=ip default-originate=never remove-private-as=no as-override=no passive=no
     use-bfd=no
[admin@Indonesia] > ip route print detail where dst-address="159.148.147.0/24"
Flags: X - disabled, A - active, D - dynamic, C - connect, S - static, r - rip, b - bgp, o - ospf, m - mme, B - blackhole, U - unreachable, P - prohibit
 0 ADb  dst-address=159.148.147.0/24 gateway=10.10.11.1 gateway-status=10.10.11.1 reachable via  ether1 distance=20 scope=40 target-scope=10
        bgp-as-path="7713,174,12578,12578,51894,51894,51894" bgp-origin=igp received-from=TO-AS7713

 1  Db  dst-address=159.148.147.0/24 gateway=10.10.12.1 gateway-status=10.10.12.1 reachable via  ether5 distance=20 scope=40 target-scope=10
        bgp-as-path="64510,64510,64510,4800,1299,5518,51894" bgp-origin=igp received-from=TO-AS4800
[admin@Indonesia] >
```

Study Case – Metric Manipulation

- Now you can see that the route to reach mikrotik.com has changed to Path B and we have no packet lost anymore

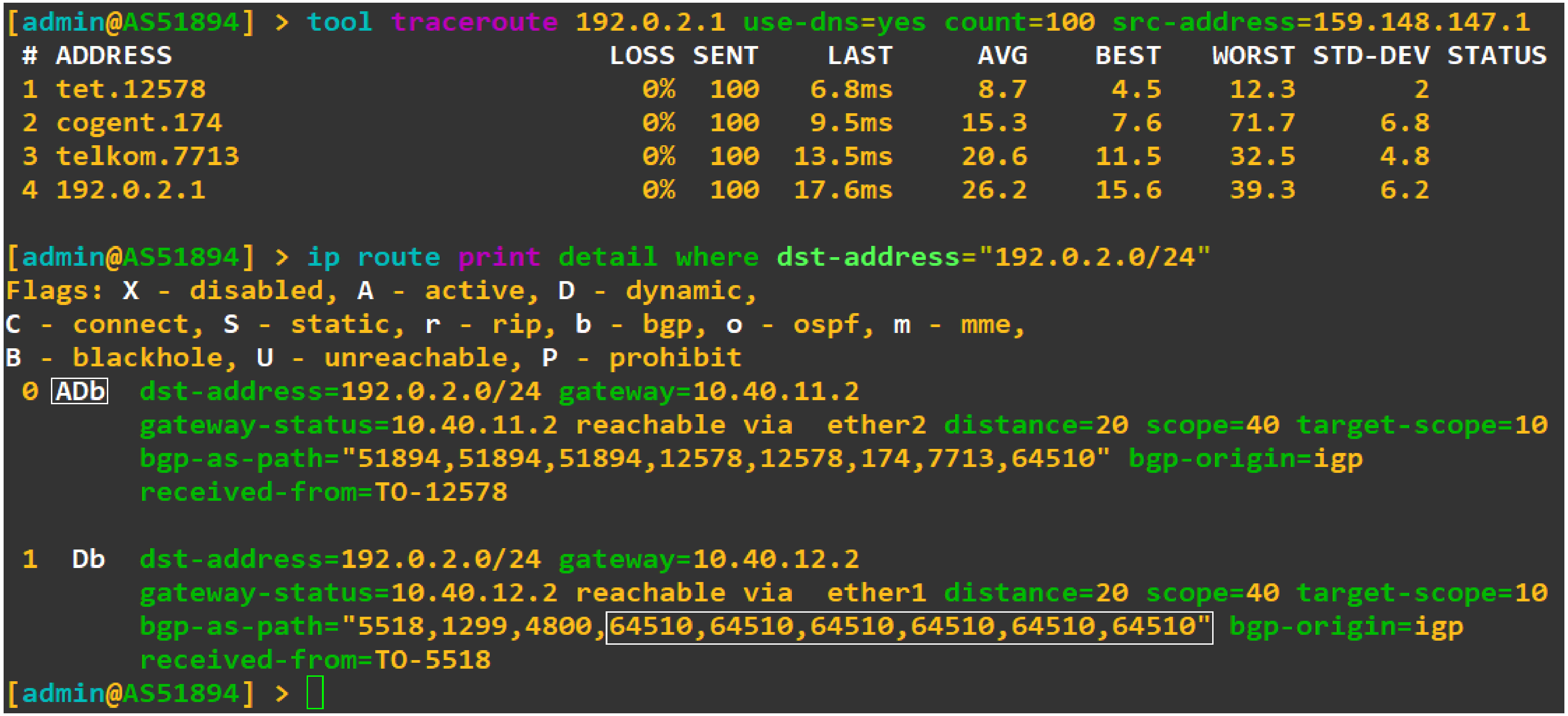


```
[admin@Indonesia] > tool traceroute mikrotik.com use-dns=yes count=100 src-address=192.0.2.1
 # ADDRESS                          LOSS SENT    LAST     AVG    BEST   WORST STD-DEV STATUS
 1 telkom.7713                        0%  100     7ms     8.8     4.1    23.6     2.7
 2 cogent.174                         0%  100  15.5ms    15.3     8.6    23.5     3.6
 3 tet.12578                          0%  100    18ms    22.1    12.3    45.7     5.1
 4 mikrotik.com                       0%  100  25.2ms    28.3    16.8    59.3     6.7
```



Study Case – Metric Manipulation

- From Mikrotikls SIA to Indonesia also routed through Path B, you can see that Path A has longer as-path and make the BGP Best Path Selection Algorithm choose Path B rather than Path A



```
[admin@AS51894] > tool traceroute 192.0.2.1 use-dns=yes count=100 src-address=159.148.147.1
 # ADDRESS                          LOSS SENT    LAST     AVG    BEST   WORST STD-DEV STATUS
 1 tet.12578                          0%  100   6.8ms     8.7     4.5    12.3       2
 2 cogent.174                         0%  100   9.5ms    15.3     7.6    71.7     6.8
 3 telkom.7713                        0%  100  13.5ms    20.6    11.5    32.5     4.8
 4 192.0.2.1                          0%  100  17.6ms    26.2    15.6    39.3     6.2

[admin@AS51894] > ip route print detail where dst-address="192.0.2.0/24"
Flags: X - disabled, A - active, D - dynamic,
C - connect, S - static, r - rip, b - bgp, o - ospf, m - mme,
B - blackhole, U - unreachable, P - prohibit
 0 ADb  dst-address=192.0.2.0/24 gateway=10.40.11.2
        gateway-status=10.40.11.2 reachable via  ether2 distance=20 scope=40 target-scope=10
        bgp-as-path="51894,51894,51894,12578,12578,174,7713,64510" bgp-origin=igp
        received-from=TO-12578

 1  Db  dst-address=192.0.2.0/24 gateway=10.40.12.2
        gateway-status=10.40.12.2 reachable via  ether1 distance=20 scope=40 target-scope=10
        bgp-as-path="5518,1299,4800,64510,64510,64510,64510,64510,64510" bgp-origin=igp
        received-from=TO-5518
[admin@AS51894] >
```

Study Case – Metric Manipulation

- Please note that if you don’t have anything to lookup the route from your destination, it would be hard because we can’t ensure the route back was correct and you only can use your feeling for that LOL :P because even we already manipulate the metric with BGP Prepend, we don’t know if our destination do the metric manipulation again to deny our metric manipulation
- The only ways to ensure the route back is only **Looking Glass** or we need more effort to ask your destination to do traceroute to your network and see the path
- Looking Glass servers are computers on the Internet running one of a variety of publicly available Looking Glass software implementations. A Looking Glass server (or LG server) is accessed remotely for the purpose of viewing routing information. Essentially, the server acts as a limited, read-only portal to routers of whatever organization is running the LG server.
- Please also note that Path A route selection are not recovered by itself, so if Path B has an issue on it, you need to re-route again

Conclusion

In **BGP** We Trust!

References

- https://wiki.mikrotik.com/wiki/Manual:IP/Route
- https://wiki.mikrotik.com/wiki/Manual:Routing/BGP
- https://wiki.mikrotik.com/wiki/Manual:Routing/Routing_filters
- https://wiki.mikrotik.com/wiki/Manual:Simple_BGP_Multihoming
- https://wiki.mikrotik.com/wiki/Manual:BGP_Best_Path_Selection_Algorithm
- https://wiki.mikrotik.com/wiki/Manual:Route_Selection_Algorithm_in_RouterOS
- https://en.wikipedia.org/wiki/Border_Gateway_Protocol
- https://www.bgp4.as/
- https://bgp.he.net/

Feel so hard to maintain your routing table?

Let me help you!

michael@takeuchi.id

https://www.facebook.com/mict404

https://www.linkedin.com/in/michael-takeuchi/

I am available for remote network engineer or freelance project vacancy :P

Question & Answer

Slide is available in my GitHub repository
https://github.com/mict404/slide/