BGP FAST FAILURE DETECTION USING BFD

INF645 Independent Project

By Nico

Contents

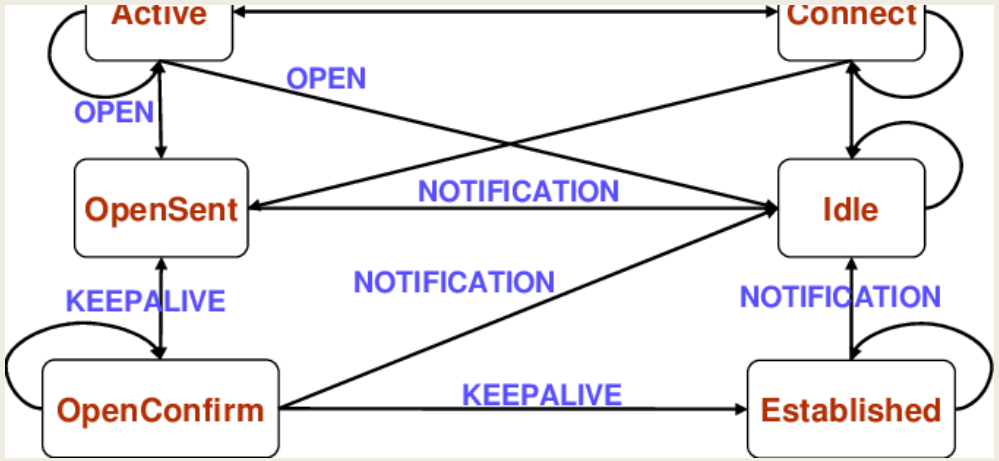
- Introduction
- Related research work
- Platform implementation
- Future outlook
- Q&A

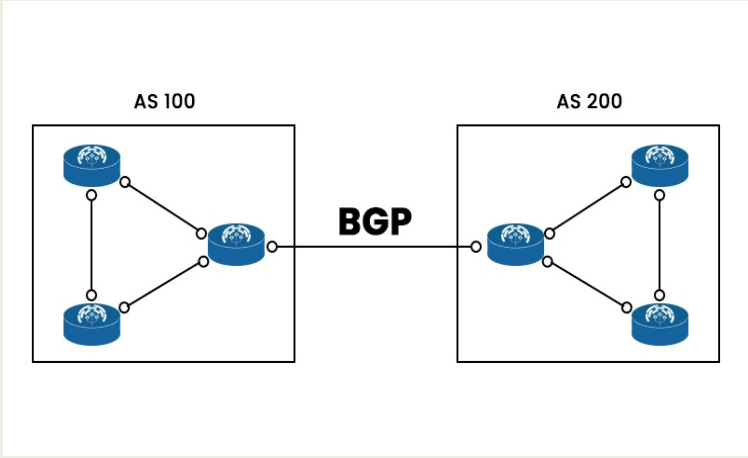
Introduction

- Motivation for choosing this topic:
  - Last year's internship at Cisco Meraki makes me feel more interested in network engineer programming
  - This year's internship is still at the same team, preparing for my upcoming new project about BGP network
  - For my future career plan, I am looking to continue to work as a network engineer

Introduction

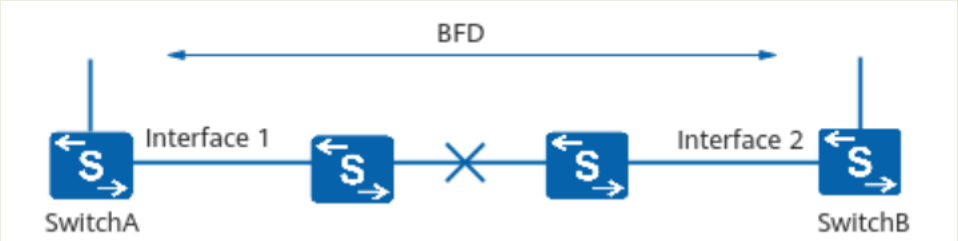
- BGP
  - Border Gateway Protocol: Achieves reachability between Autonomous Systems
  - KEEPALIVE packets for detecting network failures and rerouting(60s)

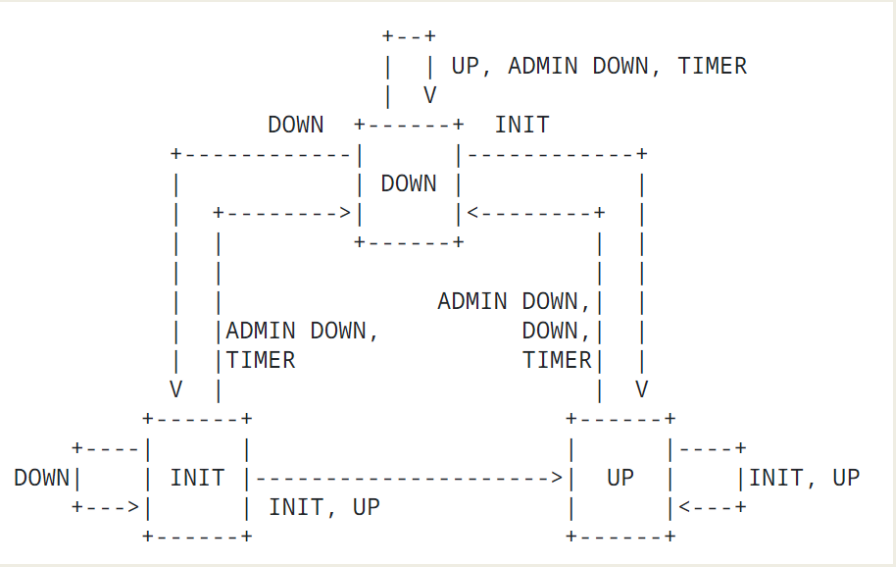




Introduction

- BFD
  - Bidirectional Forwarding Detection
  - Implement light-load failure detection, which takes only milliseconds, enhancing reliability.
  - Quickly detect a broad range of failures, including interface, data link, and forwarding engine failure.

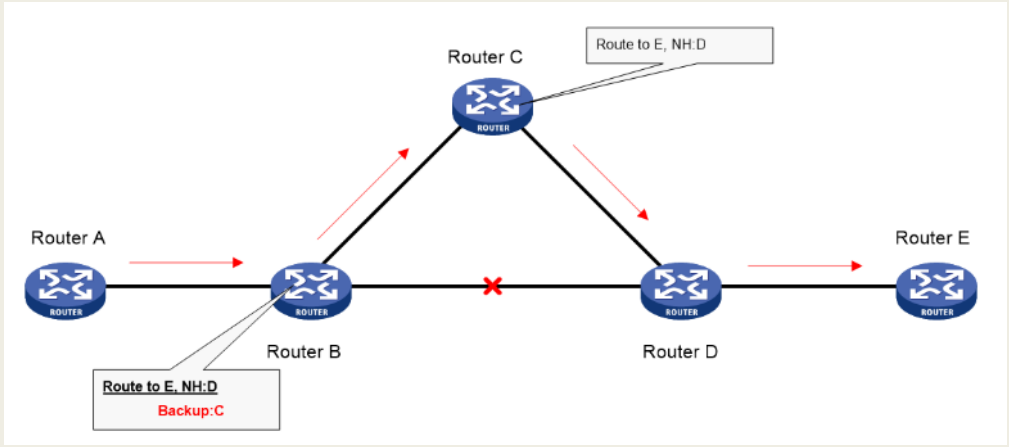




Related research work:
BGP retransmission combined with BFD

- Chanllenge using BGP:
  - Slow convergence time: BGP relies on periodic updates
  - Degradation of network performance and reliability.

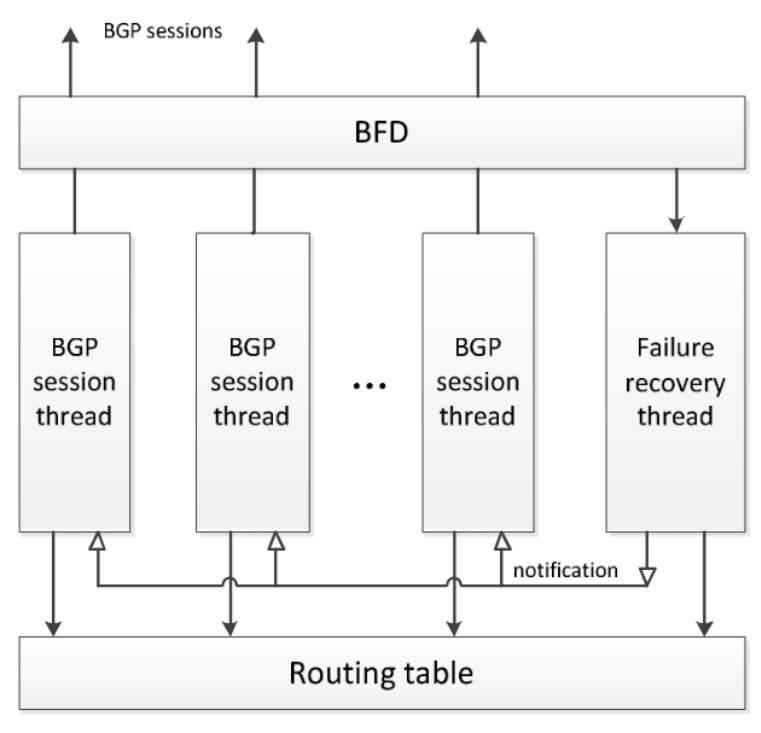
- Advantages of integrating BGP and BFD together
  - Reduce convergence time by instantly notifying BGP of link failures
  - BFD provides low overhead, short duration failure detection methods.



Related research work:
Multithread recovery algorithm in BGP

- Core idea:
  - Assigning the failure recovery process to multiple threads
  - Running in parallel with multiple BGP session threads

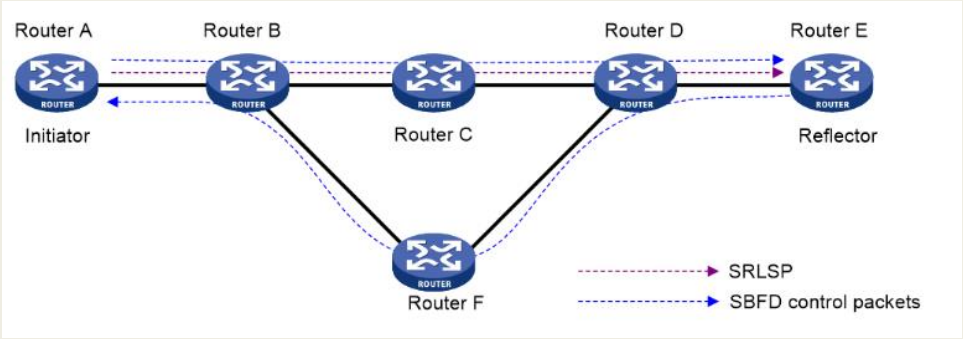
- Dynamic Route Generation
  - Detection
  - Finding alternative path and verify
  - Path combination
  - Handling exception



Related research work:
Seamless Bidirectional Forwarding Detection

- Core ideas:
  - Simplified BFD state machine (UP and DOWN states) to shorten session negotiation time
  - Shorter session negotiation time, faster detection than traditional BFD

- Working mechanism
  - Initiator and reflector
  - Link state detection is required at one end



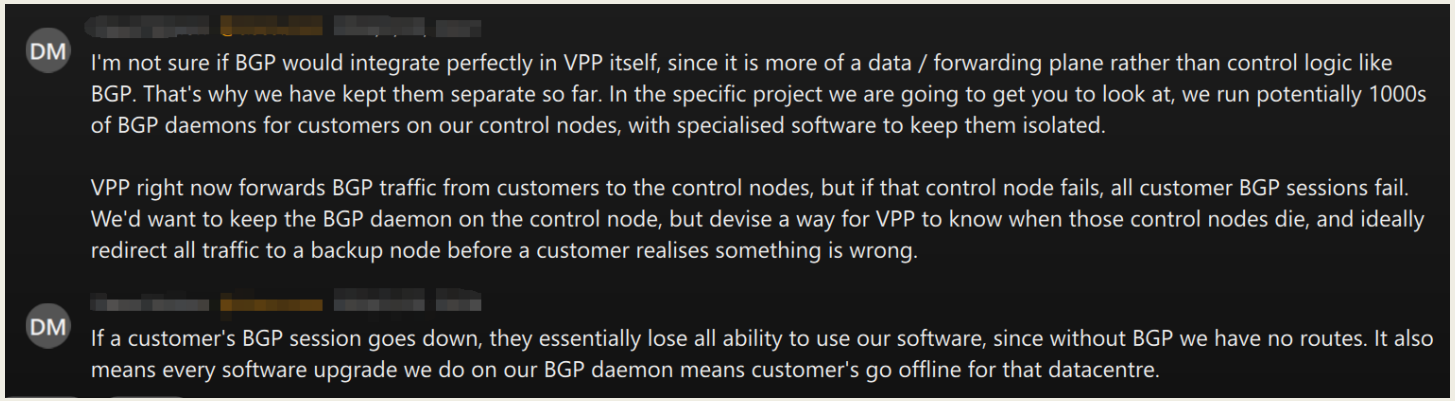
Related research work:
Comparison between BGP and BFD

- Frequency and response time:
  - BGP: second, layer 4
  - BFD: millisecond, layer 3
- Protocol design:
  - BGP: Exchange of routing information, and KEEPALIVE messages are only a mechanism used to maintain neighbor relationships.
  - BFD Designed for fast failure detection, can be used in OSPF, BGP and even in hardware
- Network burden
  - BGP: Increasing the frequency will clog up the network.
  - BFD: Centralize optimization of the failure detection process

Platform implementation

Attempts to build a platform

1. VPP



2. FRR

- Open source routing software project
- Relatively complex interface configuration, not well measured and visualized

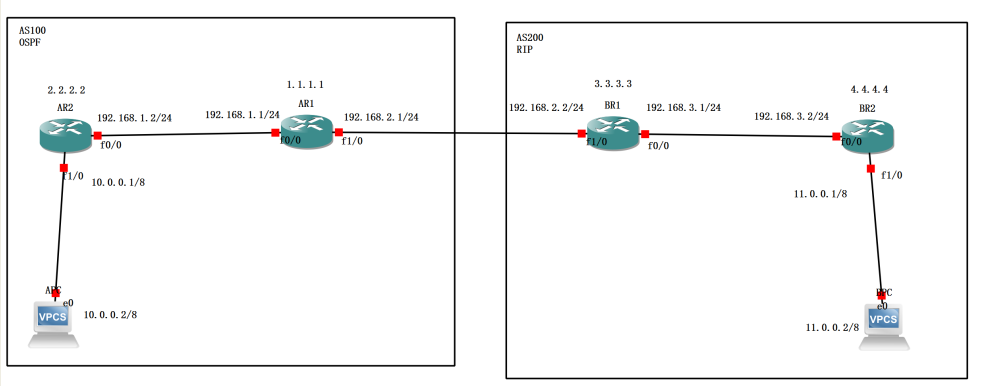
3. GNS3

- Allows users to build complex virtual networks
- Measurement is easy and visualization makes the network architecture clear.

Platform implementation

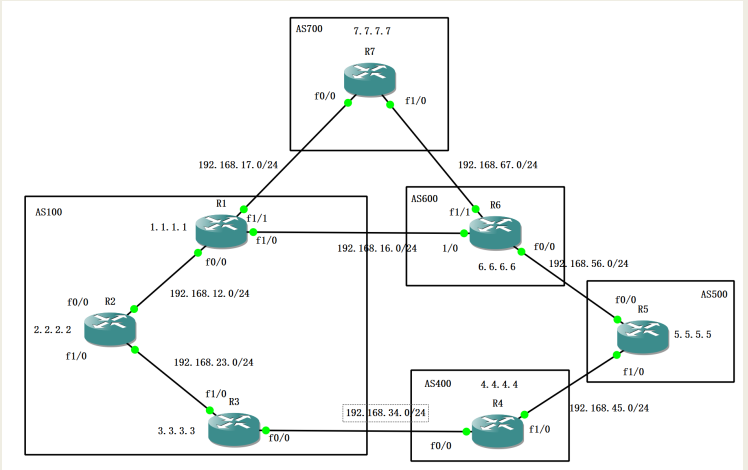
1. BGP implementation between ASes with different IGP protocols

Testing the performance gains from deploying BFD on a single BGP connection

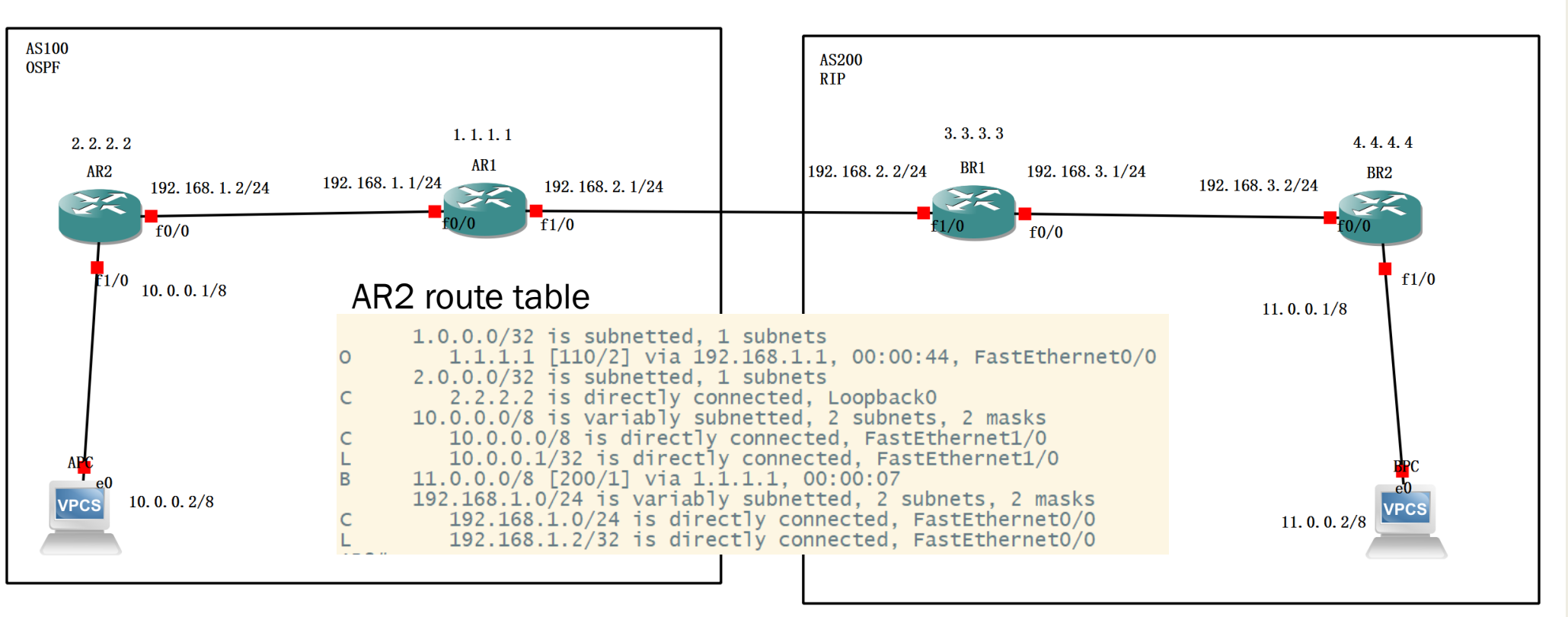


2. Platform building based on BFD link detection in BGP

Tested the deployment of BFD for BGP route convergence speed improvement between complex autonomous systems



BGP implementation between ASes with different IGP protocols

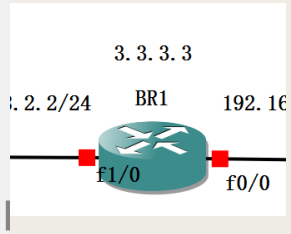


AR2 route table

```
      1.0.0.0/32 is subnetted, 1 subnets
O        1.1.1.1 [110/2] via 192.168.1.1, 00:00:44, FastEthernet0/0
      2.0.0.0/32 is subnetted, 1 subnets
C        2.2.2.2 is directly connected, Loopback0
      10.0.0.0/8 is variably subnetted, 2 subnets, 2 masks
C        10.0.0.0/8 is directly connected, FastEthernet1/0
L        10.0.0.1/32 is directly connected, FastEthernet1/0
B     11.0.0.0/8 [200/1] via 1.1.1.1, 00:00:07
      192.168.1.0/24 is variably subnetted, 2 subnets, 2 masks
C        192.168.1.0/24 is directly connected, FastEthernet0/0
L        192.168.1.2/32 is directly connected, FastEthernet0/0
```

BGP implementation between ASes with different IGP protocols

416	1220.136245	192.168.2.2	192.168.2.1	BGP	73 KEEPALIVE Message
419	1220.355463	192.168.2.1	192.168.2.2	TCP	60 179 → 43565 [ACK] Seq=570 Ack=570 Win=15815 Len=0
426	1269.653937	192.168.2.1	192.168.2.2	BGP	73 KEEPALIVE Message
428	1270.954216	192.168.2.1	192.168.2.2	TCP	73 [TCP Retransmission] 179 → 43565 [PSH, ACK] Seq=57
429	1273.511700	192.168.2.1	192.168.2.2	TCP	73 [TCP Retransmission] 179 → 43565 [PSH, ACK] Seq=57
430	1278.628201	192.168.2.1	192.168.2.2	TCP	73 [TCP Retransmission] 179 → 43565 [PSH, ACK] Seq=57
432	1288.839807	192.168.2.1	192.168.2.2	TCP	73 [TCP Retransmission] 179 → 43565 [PSH, ACK] Seq=57
436	1309.317714	192.168.2.1	192.168.2.2	TCP	73 [TCP Retransmission] 179 → 43565 [PSH, ACK] Seq=57
439	1319.850762	192.168.2.1	192.168.2.2	BGP	73 KEEPALIVE Message
447	1350.229132	192.168.2.1	192.168.2.2	TCP	92 [TCP Retransmission] 179 → 43565 [PSH, ACK] Seq=57
452	1378.188193	192.168.2.1	192.168.2.2	BGP	73 KEEPALIVE Message
455	1391.170394	192.168.2.1	192.168.2.2	TCP	111 [TCP Retransmission] 179 → 43565 [PSH, ACK] Seq=57
458	1400.751090	192.168.2.1	192.168.2.2	BGP	75 NOTIFICATION Message

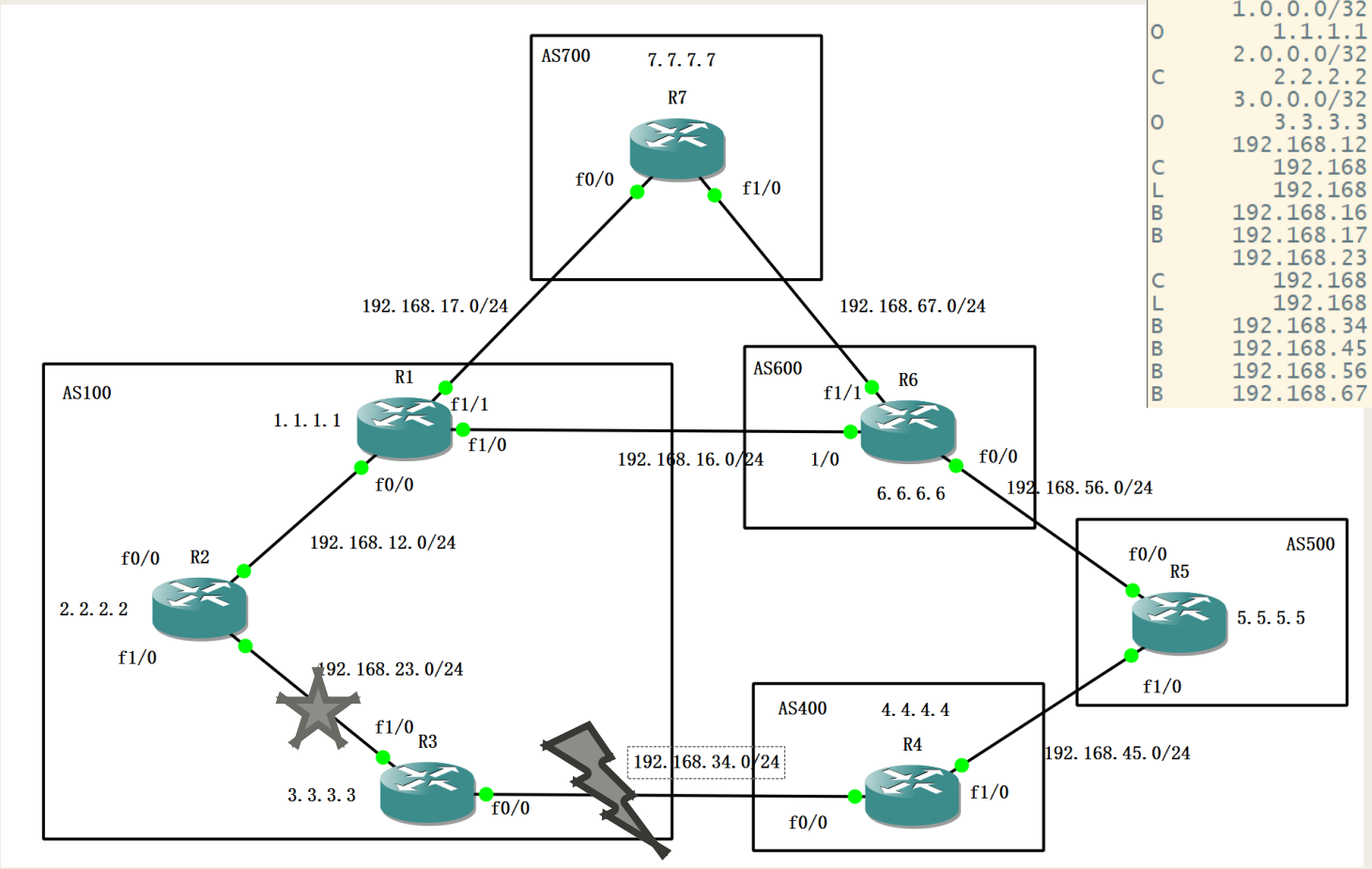


1400-1200 = 200s

533	45.700624	192.168.2.1	192.168.2.2	BFD Control	66 Diag: No Diagnostic, State: Up, Flags: 0x00
534	46.026939	192.168.2.1	192.168.2.1	BFD Echo	54 Originator specific content
535	46.435047	192.168.2.1	192.168.2.1	BFD Echo	54 Originator specific content
536	46.593173	192.168.2.1	192.168.2.2	BFD Control	66 Diag: No Diagnostic, State: Up, Flags: 0x00
537	46.886445	192.168.2.1	192.168.2.2	BFD Control	66 Diag: Echo Function Failed, State: Down, Flags: 0x
538	46.897003	192.168.2.1	192.168.2.2	TCP	60 179 → 13834 [FIN, PSH, ACK] Seq=20 Ack=20 Win=1577

46.9-45.7=1.2s

Platform building based on BFD link detection in BGP



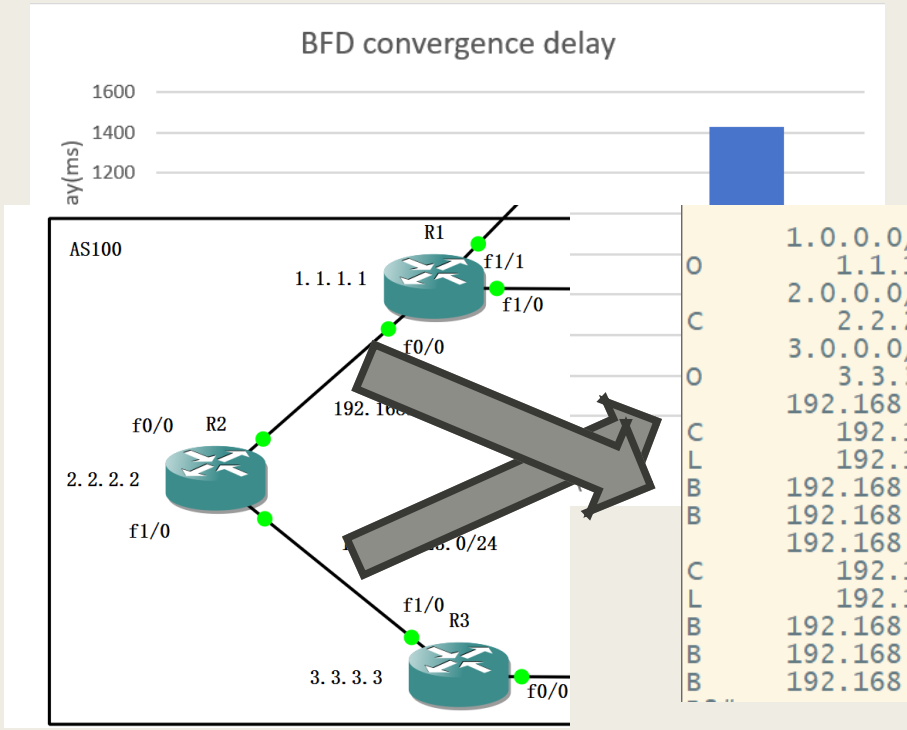
R2

```
      1.0.0.0/32 is subnetted, 1 subnets
O        1.1.1.1 [110/2] via 192.168.12.1, 00:04:42, FastEthernet0/0
      2.0.0.0/32 is subnetted, 1 subnets
C        2.2.2.2 is directly connected, Loopback0
      3.0.0.0/32 is subnetted, 1 subnets
O        3.3.3.3 [110/2] via 192.168.23.3, 00:04:42, FastEthernet1/0
      192.168.12.0/24 is variably subnetted, 2 subnets, 2 masks
C        192.168.12.0/24 is directly connected, FastEthernet0/0
L        192.168.12.2/32 is directly connected, FastEthernet0/0
B     192.168.16.0/24 [200/0] via 3.3.3.3, 00:03:54
B     192.168.17.0/24 [200/0] via 3.3.3.3, 00:03:54
      192.168.23.0/24 is variably subnetted, 2 subnets, 2 masks
C        192.168.23.0/24 is directly connected, FastEthernet1/0
L        192.168.23.2/32 is directly connected, FastEthernet1/0
B     192.168.34.0/24 [200/0] via 3.3.3.3, 00:04:13
B     192.168.45.0/24 [200/0] via 3.3.3.3, 00:04:13
B     192.168.56.0/24 [200/0] via 3.3.3.3, 00:04:13
B     192.168.67.0/24 [200/0] via 3.3.3.3, 00:03:55
```

```
Tracing the route to 192.168.17.7
VRF info: (vrf in name/id, vrf out name/id)
  1 192.168.23.3 20 msec 16 msec 16 msec
  2 192.168.34.4 [AS 400] 44 msec 16 msec 64 msec
  3 192.168.45.5 [AS 400] 48 msec 60 msec 64 msec
  4 192.168.56.6 [AS 500] 96 msec 104 msec 76 msec
  5 192.168.67.7 [AS 600] 112 msec 104 msec 104 msec
```

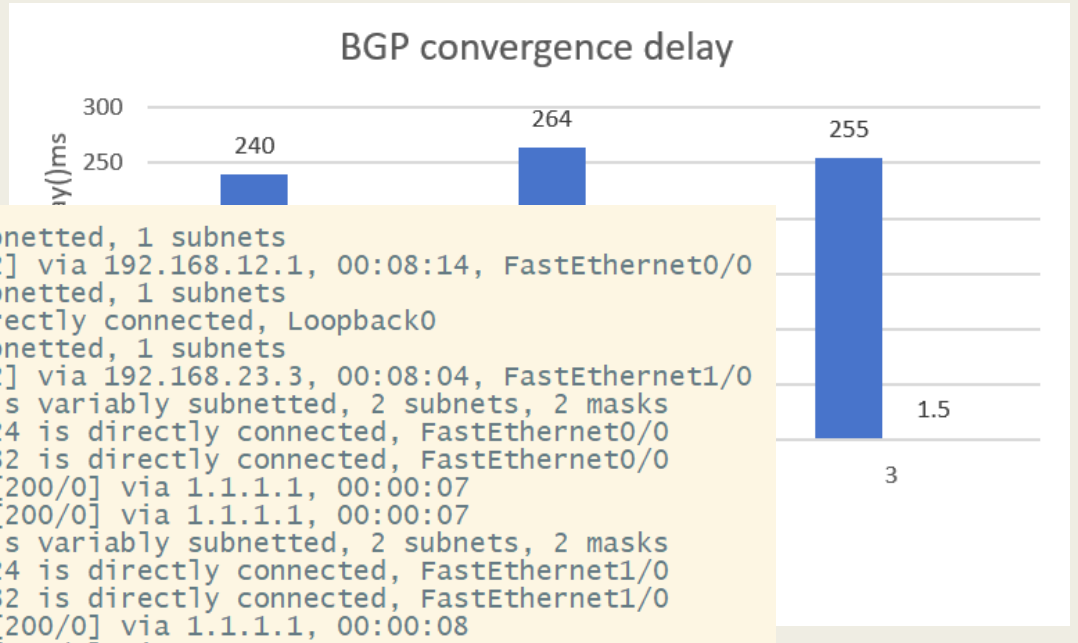
Platform building based on BFD link detection in BGP

BFD convergence delay:




```
     1.0.0.0/32 is subnetted, 1 subnets
O       1.1.1.1 [110/2] via 192.168.12.1, 00:08:14, FastEthernet0/0
     2.0.0.0/32 is subnetted, 1 subnets
C       2.2.2.2 is directly connected, Loopback0
     3.0.0.0/32 is subnetted, 1 subnets
O       3.3.3.3 [110/2] via 192.168.23.3, 00:08:04, FastEthernet1/0
     192.168.12.0/24 is variably subnetted, 2 subnets, 2 masks
C       192.168.12.0/24 is directly connected, FastEthernet0/0
L       192.168.12.2/32 is directly connected, FastEthernet0/0
B    192.168.16.0/24 [200/0] via 1.1.1.1, 00:00:07
B    192.168.17.0/24 [200/0] via 1.1.1.1, 00:00:07
     192.168.23.0/24 is variably subnetted, 2 subnets, 2 masks
C       192.168.23.0/24 is directly connected, FastEthernet1/0
L       192.168.23.2/32 is directly connected, FastEthernet1/0
B    192.168.45.0/24 [200/0] via 1.1.1.1, 00:00:08
B    192.168.56.0/24 [200/0] via 1.1.1.1, 00:00:08
B    192.168.67.0/24 [200/0] via 1.1.1.1, 00:00:08
```

BGP convergence delay:



Future outlook

- Combine the retransmission algorithm with the BFD
- Try to implement all the topology in VPP system, include BGP and BFD protocol
- Explore the other usage of BFD: in IBGP, EBGP and hardware

Q&A

Thank you