I. J. Computer Network and Information Security, **2018, 6, 1-15**
Published Online June 2018 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijcnis.2018.06.01

Ensure Symmetrical Traffic Flow, to prevent the Dropping of Response Packet by the Firewall, on the Active-Active Data Centers

Irwan Piesessa
Computer Science Department, BINUS Graduate Program - Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480
E-mail:irwan.piesessa@gmail.com

Benfano Soewito
Computer Science Department, BINUS Graduate Program - Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480
E-mail: bsoewito@binus.edu

Received: 23 February 2018; Accepted: 11 May 2018; Published: 08 June 2018

Abstract—This paper illustrates the problem in the Active-Active Data Centers of an organization, where response traffic from the destination server is dropped by the firewall because the initial traffic from the client departs from another firewall in different Data Center (asymmetric traffic). This problem can be solved by two proposed solutions, namely the implementation of the BGP Community attributes and OSPF over GRE tunnel. The case study also compares both proposed solutions in terms of recovery time, packet loss, ICMP response time and TCP three-way handshake time for HTTP connection.

Index Terms—Asymmetric routing, symmetric routing, Border Gateway Protocol (BGP), Open Shortest Path First (OSPF), Generic Routing Encapusulation (GRE).

I. Introduction

Routing is the process of finding the best path from the source system to the destination system on a network [1]. Asymmetric routing is a problem that often occurs in organizations that have two data centers to connect to the internet or to other networks through different firewalls. In asymmetric routing, the first TCP packet departs from the source with a syn flag (Synchronize) [2] to destination using one path and uses a different path when the response, TCP packet from the destination server (with the syn-ack flag) is sent back to the source [3]. The response packets from the server (for example, TCP (Transmission Control Protocol) syn-ack packet)) will be dropped by the firewall on the return path. This is because this firewall does not have a connection or session table for the related response packets. The connection or session table for clients that transmits initial packets (for example, TCP syn packet), resides on the firewall that is on the packet depart path.

Some solutions have been offered in terms of firewall configuration, for example, is to disable the tcp-check feature [4]. However, this will increase the security risk on the network. This paper offers a solution from the side of the routing protocol without reducing the level of network security (without disabling security features on the firewall).

An organization in Indonesia ("Organization A") have a Primary and Secondary Data Center that connected their Indonesian Head Office to their overseas Regional Data Center. Their overall IGP (Interior Gateway Protocol) Diagram can be seen in fig. 1.

"Organization A" uses OSPF as its IGP routing protocol where OSPF (Open Shortest Path First) is currently the most commonly used IGP protocols [5]. OSPF uses link-state or technology based on the SPF (Shortest Path First) to build and calculate the shortest route/path to all known destinations [6].

For EGP (Exterior Gateway Protocol), they use BGP (Border Gateway Protocol). BGP has many attributes or parameters that can be used to determine the best route to a network. An Administrator can select policies that match his/her organization's needs using BGP. With these advantages BGP was designed in the early 1990s [7] is de-facto used as an inter-domain routing protocol on the internet [8]. "Organization A" BGP network topology can be seen in fig. 2.

To load balance the links to the Primary Data Center (DC-A) and the Secondary Data Center (DC-B), traffic from the Head Office were divided as below:

- Intranet and all web application traffic to Regional DC, will enter DC-B first, then it will be forwarded to DC-A and will go to Regional DC through DC-A firewall.
- Email and other traffic to Regional DC will go to DC-A first, then it will be forwarded to DC-A via DC-A Firewall.

To achieve the above routing policy, the BGP local-preference attribute is used. Local-preference is a BGP attribute that belongs to a well known discretionary type that is included in each router update from one router to another router within the same Autonomous System [9]. If two or more routes have the same network prefix in the BGP table, then if the routes have same weight value, a route with higher local-preference value will be selected [10].

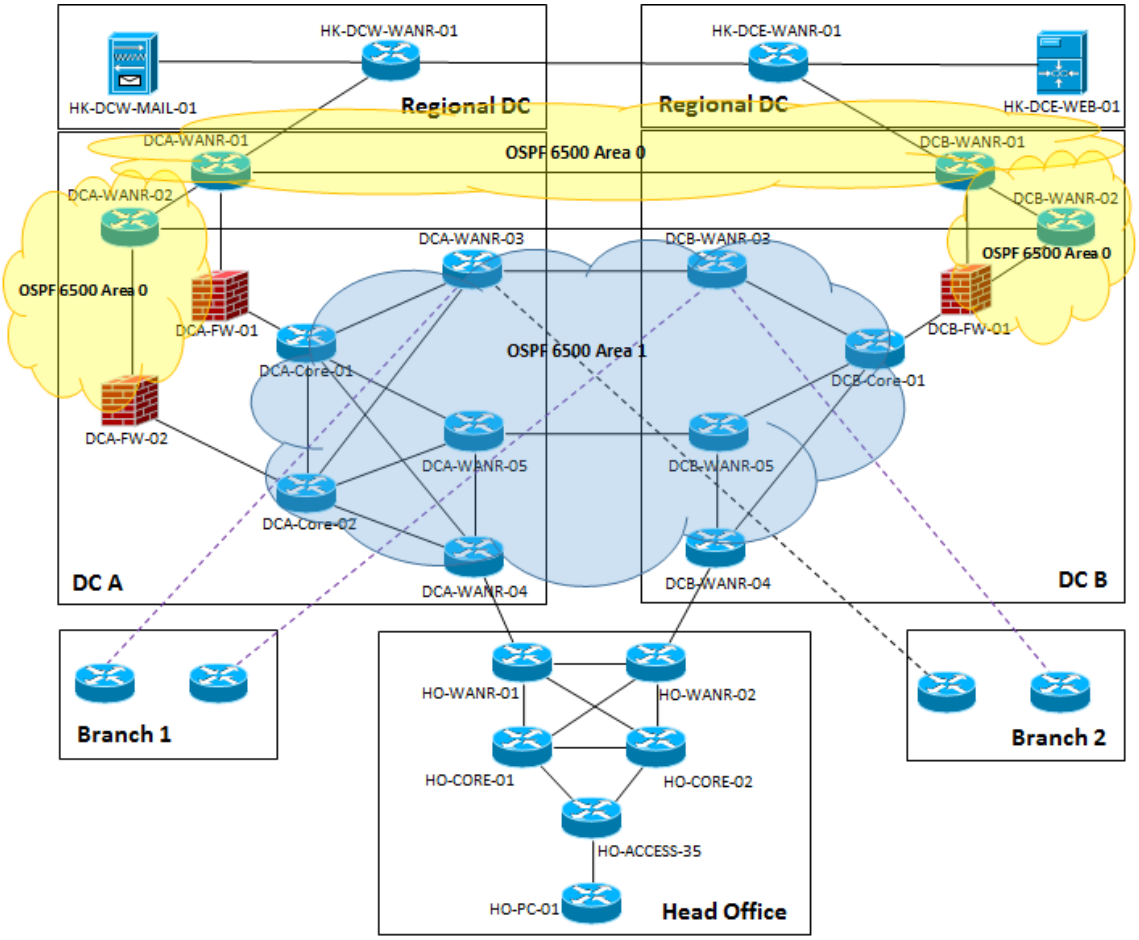

Fig.1. OSPF Topology

In “Organization A” head office network, on WAN (Wide Area Network) router that's connected to the DC-A, network prefix for Intranet and web application servers are set to have a local-preference value of 150. And other network prefixes that include the email server will have local-preference value 200. On Head Office WAN router that's connected to DC-B, the intranet and web application server network prefix are set to have a local-preference value of 200, and other network prefix set to 100. The local-preference configuration of the WAN routers in the network head office can be seen in table 1.

Table 1. BGP Local-preference Setting

BGP Local Preference Value		
Router	Intranet+Web Servers Network Prefixes	Mail & Other Network Prefixes
HO-WANR-01	150	200
HO-WANR-02	200	100

Based on BGP local-preference attribute on table 1, the traffic flow is shown in fig. 3.

The “Organization A” routing policy also states that traffic from the Head Office must flow through their dedicated Data Center routers (IE DCB-WANR-04, DCB-WANR-05, DCA-WANR-04 and DCA-WANR-05) to distribute their WAN routers workload in DC-A and DC-B. To achieve this policy, these routers are reconfigured the Administrative Distance (AD) values of their BGP routing protocols. AD is a feature that routers use to select the best path when there are two or more different routes to the same destination of two different routing protocols. AD defines the reliability of routing protocols. Each routing protocol has different priority values [11].

DCB-WANR-04, DCB-WANR-05, DCA-WANR-04 and DCA-WANR-05 are configured with AD 105 for internal BGP, so BGP routes will be preferred over OSPF that has default AD, which is 110 [12]. They allocate other domestic WAN routers (IE DCA-WAN-03 and DCB-WANR-03) for their branch traffic.

To speed up the routing convergence time, BGP keepalive and holdtime setting on dedicated Data Center WAN routers for Head Office are reconfigured. BGP

holdtime is the time value set by the router to determine whether the bgp peer is live or not If within a specified holdtime the router does not receive a keepalive message from its peer, then it will declare that the BGP peer is dead [13]. Usually holdtime time is three times the keepalive time [14].

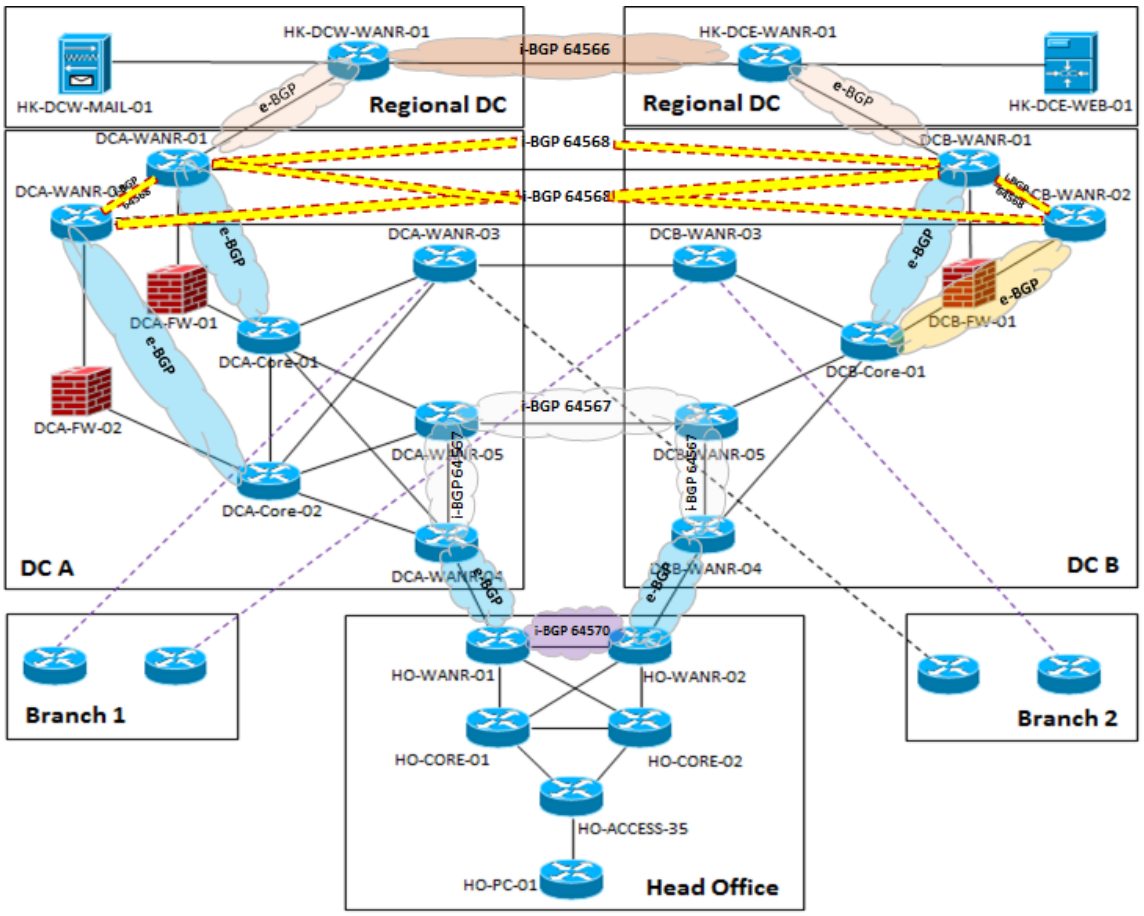

Fig.2. BGP Topology

DCB-WANR-04, DCB-WANR-05, DCA-WANR-04 and DCA-WANR-05 are reconfigured with a 5-second BGP keepalive time and a 10-second BGP holdtime. With this configuration, the BGP speaker router will send a keepalive message to the BGP peer in every 5 seconds. If within 10 seconds there is no BGP keepalive message received by the router, it will be declared that BGP peer is down, and all BGP routes received from the peer will be removed from the BGP table.

We simulate "Organization A" networks, including the Head Office network in Indonesia, their Primary and Secondary Data Centers (DC-A and DC-B) and Overseas Data Centers, with the same topology, the same router vendor and network operating systems, the same IGP and EGP routing protocols (OSPF and BGP), the same OSPF Areas and network type, and the same BGP routing policies (including the same BGP attributes, timers and Administrative Distance) as well as the same traffic flow for intranet, email and other applications.

To confirm the network traffic flow according to fig. 3, below is the traceroute output from the client PC at Head Office (HO-PC-01) to the intranet web server:

```
HO-PC-01#traceroute 10.15.3.100
Tracing the route to 10.15.3.100
  1 10.23.35.1 32 msec 12 msec 32 msec
  2 10.23.33.5 64 msec 36 msec 12 msec
  3 10.23.33.18 124 msec 56 msec 176 msec ← HO-WANR-02
  4 10.23.255.5 276 msec 184 msec 136 msec ← DCB-WANR-04
  5 10.19.0.54 140 msec 140 msec 132 msec ← DCB-WANR-05
  6 10.18.255.1 168 msec 144 msec 100 msec ← DCA-WANR-05
  7 10.18.0.13 188 msec 144 msec 108 msec ← DCA-Core-01
  8 10.18.0.38 112 msec 108 msec 152 msec
  9 10.18.0.26 196 msec 160 msec 172 msec
 10 10.19.2.2 164 msec 156 msec 160 msec
 11 10.15.0.13 236 msec 180 msec 172 msec
 12 10.15.0.18 220 msec 212 msec 216 msec ← Intranet Web Server
```

II. Asymmetric Routing Problem

The problem occurs when the inter-DC link (DCA-DC-B link between DCA-WANR-05 - DCB-WANR-05) down. Traffic to the intranet server from the Head Office network will remain through the Head Office - DC-B WAN link. Then, because of their routing policy, traffic

from Head Office will not flow through their dedicated branches WAN routers (IE DCB-WANR-03 and DCA-WANR-03), but will be forwarded by DCB-WANR-05 to the DC-B Core Router and go to the intranet server in Regional DC via DC-B firewall. And because of their routing policy as well, the response traffic will return to Head Office through DC-A via DC-A firewall. This is explained in the traffic flow diagram in fig. 4.

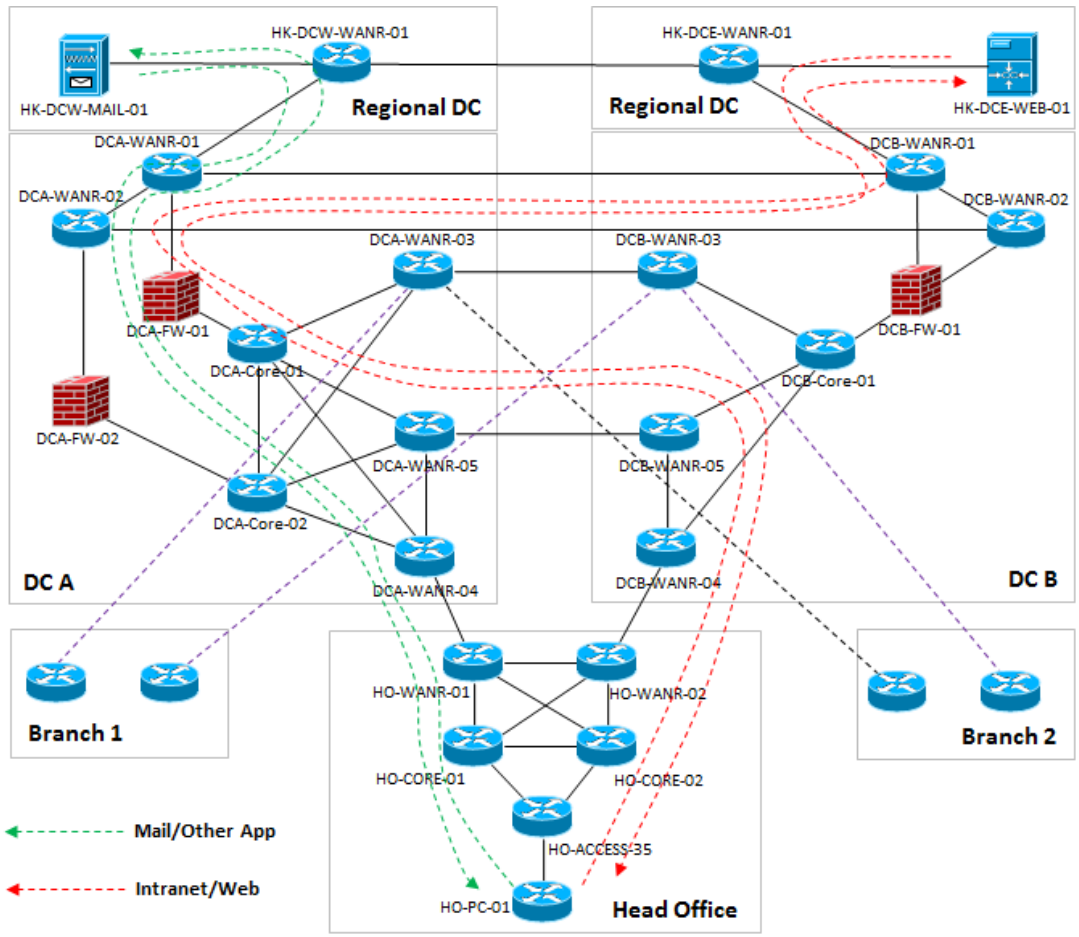

Fig.3. Routing Policy for Network Traffic Flow

Because of this, the connection to the intranet server can not be done. This can be seen from the ping results that run from the client PC to the intranet server before and after the inter-DC link is disconnected as follows:

```
HO-PC-01#ping 10.15.3.100 rep 200
Type escape sequence to abort.
Sending 200, 100-byte ICMP Echos to
10.15.3.100, timeout is 2
seconds:  !!!!!!!!!!!!!!!!!!!!!...UUU..UUUUU
UUUUUUUUUUUUUUU
```

III. Proposed Solutions

After analyzing their network architecture and routing policy, then to solve this case, we propose two solutions as follows:

- Using the BGP community
- Using OSPF peering over GRE (Generic Routing Encapsulation) tunnel

A. Using the BGP Community

Community is an optional transitive attribute of the BGP. BGP routers can use this attribute to determine whether to accept or not and determine the priority level of the received BGP route [15].

We propose to use BGP communities in the BGP networks to detect whether the inter-DC link between DCA-WANR-05 and DCB-WANR-05 is in normal condition (up) or not (down). We configured the DCA-WANR-05 router to add the BGP community attribute to all routes sent to the DC-B (DCB-WANR-05) router and will be advertised to DCB-WANR-04 and HO-WANR-02. Both WAN routers at Head Office (IE HO-WANR-01 and HO-WANR-02) will know about the conditions of inter-DC link based on the routes they receive from DC-B. If there is a community string on the route, then the inter-DC link is in normal condition. And if it does not exist, then they will state that the inter-DC link is down.

We use 64567: 1000 as a BGP community value on the DCA-WANR-05 router to route table that will be advertised to the DCB-WANR-05 router.

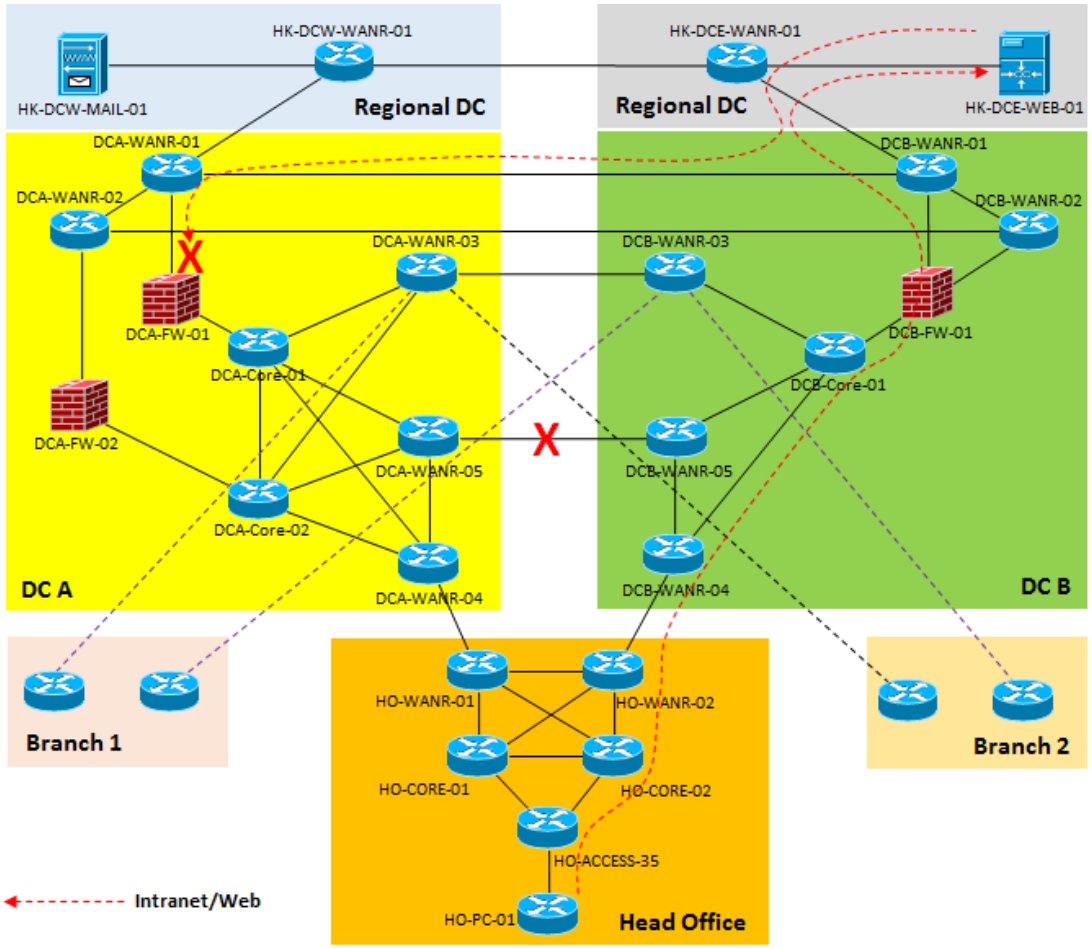

Fig.4. Asymmetric Routing Problem

In HO-WANR-02 router, we set the BGP rule that if the route with the intranet server network prefix is received from DCB-WANR-04 and has a BGP community with value 64567: 1000, we will set the route to have BGP local-preference with value 200. For other routes we will set the route to have a BGP local-preference BGP with value 100. This BGP route table will be installed in the router's routing table and advertised to the HO-WANR-01 router. The BGP rule on the HO-WANR-02 router at Head Office network can be seen in table 2.

Table 2. BGP rule on HO-WANR-02

Router HO-WANR-02		
If Match (and)		Then Set
Network = Intranet Servers Prefix	Community = 64567:1000	local-preference = 200
Network = Other Prefixes		local-preference = 100

Verification of the BGP best route to the Intranet server on both of the WAN routers in the Head Office, can be seen in fig. 5 and 6. For route received from the DCB-WANR-04 router, which informs that for network intranet server prefix (10.15.3.100/32) that have community BGP 64567: 1000, it will go to Regional DC via router HO-WANR-02 (10.23.1.4) then forwarded to DCB-WANR-04 (10.23.255.5).

```
Paths: (3 available, best #1, table Default-IP-Routing-Table, RIB-failure(17))
Flag: 0x9E0
 Advertised to update-groups:
   1
 64567, (received & used)
  10.23.255.5 (metric 1) from 10.23.1.4 (10.23.1.4)
    Origin incomplete, metric 1000, localpref 200, valid, internal, best
    Community: 64567:1000
```

Fig.5. BGP route table on HO-WANR-01 router

```
BGP routing table entry for 10.15.3.100/32, version 153
Paths: (2 available, best #1, table Default-IP-Routing-Table)
 Advertised to update-groups:
   2
 64567
  10.23.255.5 from 10.23.255.5 (10.19.1.5)
    Origin incomplete, metric 1000, localpref 200, valid, external, best
    Community: 64567:1000
```

Fig.6. BGP route table on HO-WANR-01 router

With the combination of local-preference and the BGP community attributes, if the inter-DC link between DCA-WANR-05 and DCB-WANR-05 is down, HO-WANR-02 will not receive the intranet server prefix from the DC-B router that has the BGP community with a value of 64567: 1000. Please note that we have created a BGP rule in the HO-WANR-02 router, that the route to be given the highest local-preference (200), is the route for the intranet

server network prefix (10.15.3.100/32) and has the BGP community attribute 64567: 1000 (using AND operation). Because of this, the router will delete the previous route, and will install the intranet server prefix originating from the DC-A via the HO-WANR-01 router. Then, all HO WAN routers will direct traffic to the intranet server to the DC-A via HO-WANR-01 (10.23.1.3) and then through DCA-WANR-04 (10.23.255.1).

Once the BGP community solution configuration has been implemented in the simulator, we simulate by disconnecting the inter-DC link between DCA-WANR-05 and DCB-WANR-05. The BGP route table for the intranet server prefix (10.15.3.100) on the HO-WANR-01 and HO-WANR-02 routers after the inter-DC link between DCA-WANR-05 and DCB-WANR-05 is disconnected, can be seen in fig. 7 and 8.

```
BGP routing table entry for 10.15.3.100/32, version 107
Paths: (2 available, best #1, table Default-IP-Routing-Table)
Flag: 0x9C0
 Advertised to update-groups:
   2
 64567
  10.23.255.1 from 10.23.255.1 (10.18.1.5)
    Origin incomplete, metric 1000, localpref 150, valid, external, best
```

Fig.7. BGP route table on HO-WANR-01 after inter-DC link down

```
HO-WANR-02#sh ip bgp 10.15.3.100
BGP routing table entry for 10.15.3.100/32, version 157
Paths: (2 available, best #2, table Default-IP-Routing-Table, RIB-failure(17))
Flag: 0x820
 Advertised to update-groups:
   1
 64567, (received & used)
  10.23.255.1 (metric 2) from 10.23.1.3 (10.23.1.3)
    Origin incomplete, metric 1000, localpref 150, valid, internal, best
```

Fig.8. BGP route table on HO-WANR-02 after inter-DC link down

The network flow to the intranet server after inter-DC link between DCA-WANR-05 and DCB-WANR-05 is disconnected and the BGP community configuration is implemented can be seen in fig. 9.

Output of the ping command on HO-PC-01 to the intranet web server which confirms that traffic has recovered after inter-DC connection between DCA-WANR-05 and DCB-WANR-05 disconnected is shown below:

```
HO-PC-01#ping 10.15.3.100 rep 20
Sending 20, 100-byte ICMP Echos to
10.15.3.100, timeout is 2 seconds:
!!!!!!!!!...!!!!!!!!
Success rate is 85 percent (17/20), round-
trip min/avg/max = 180/243/328 ms
HO-PC-01#
```

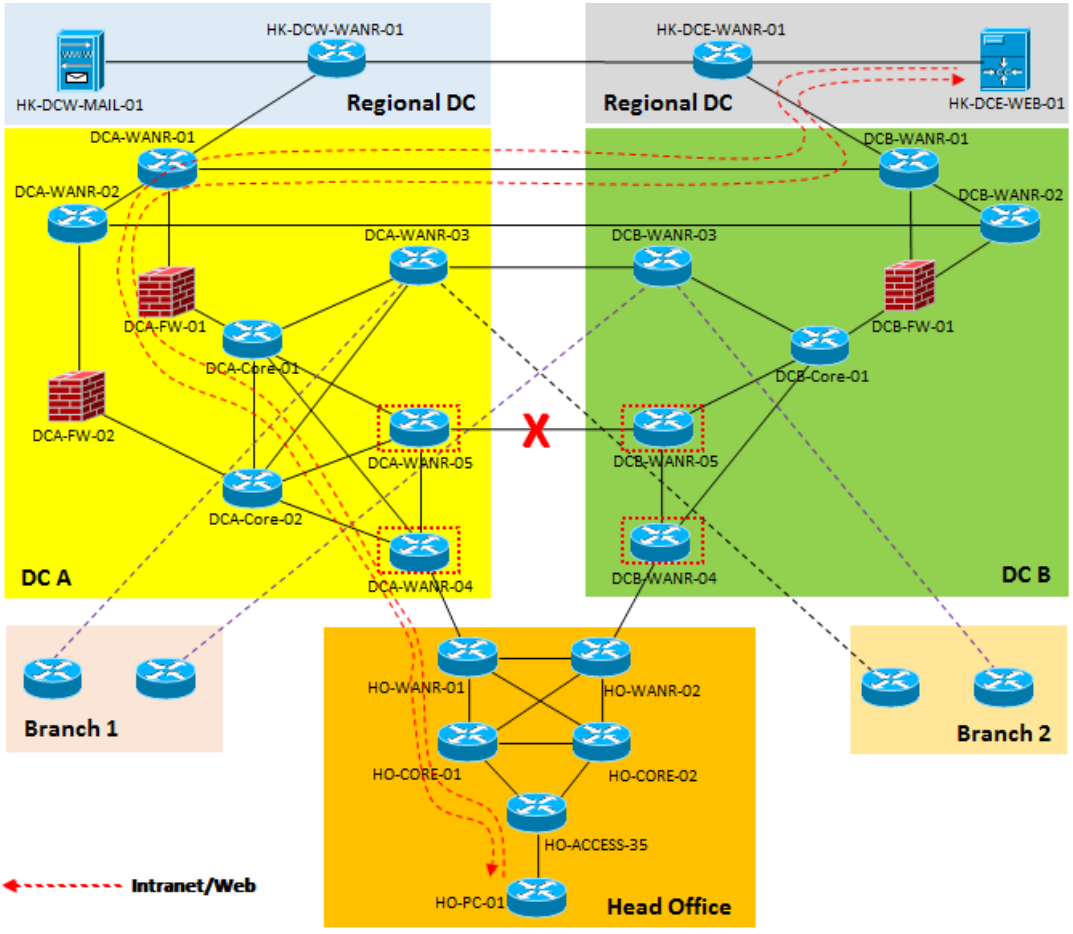

Fig.9. Traffic flow from Head Office to Intranet Server after the Inter-DC link down

There are 3 icmp (Internet Control Message Protocol) echo-request packets that get no reply during the route transition when the inter-DC link is disconnected. The traceroute output below from the Head Office PC to the intranet server confirms that network traffic to the intranet server is now via DC-A:

```
HO-PC-01#traceroute 10.15.3.100
Tracing the route to 10.15.3.100
  1 10.23.35.1 32 msec 44 msec 16 msec
  2 10.23.33.5 56 msec 60 msec 60 msec
  3 10.23.33.14 124 msec 36 msec 84 msec ←
HO-WANR-01
  4 10.23.255.1 192 msec 84 msec 88 msec ←
DCA-WANR-04
  5 10.18.0.5 160 msec 88 msec 120 msec ←
DCA-Core-01
  6 10.18.0.38 140 msec 140 msec 160 msec
  7 10.18.0.26 180 msec 204 msec 152 msec
  8 10.19.2.2 180 msec 164 msec 192 msec
  9 10.15.0.13 208 msec 204 msec 184 msec
 10 10.15.0.18 264 msec 264 msec 272 msec ←
Intranet Web Server
```

Based on the BGP route table, ping, http (Hypertext Transfer Protocol) access and traceroute commands from our network simulations, we can state that BGP community solutions can solve asymmetric routing problems on "Organization A" networks.

B. Using OSPF over GRE Tunnel Solution

We propose an alternative solution, based on the OSPF routing protocol that used as IGP in DC-A and DC-B. The idea is to utilize the inter-DC link between DCA-WANR-03 and DCB-WANR-03 in accordance with company policy allocated for use only by the branch network. Because in DCA-WANR-04, DCA-WANR-05, DCB-WANR-04 and DCB-WANR-05, BGP internal routing protocol (with AD 105) is preferred over OSPF (AD 110), we introduce the route to regional DC in OSPF that will be promoted and installed in the DCB-WANR-05 routing table when the BGP peering session between DCA-WANR-05 and DCB-WANR-05 is disconnected. OSPF peering will be established on top of the new GRE Tunnel. The GRE is a protocol used to encapsulate a packet over other protocols that run over the network layer [16]. GRE runs over a virtual point-to-point connection. The GRE packet structure can be seen in fig. 10 [17]:

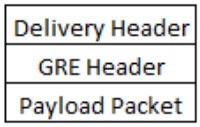

Fig.10. GRE Packet Structure

We propose to use the new loopback IP address as the source and destination address. This IP address will not be advertised to IGP to prevent routing loops. The new loopback IP address will be routed through static routes or other routing protocols that have a lower AD than OSPF. This OSPF peering will always be available both when the inter-DC link between DCA-WANR-05 and DCB-WANR-05 is normal or disconnected (down).

OSPF peering diagram running on the GRE Tunnel we propose can be seen in fig. 11 and 12.

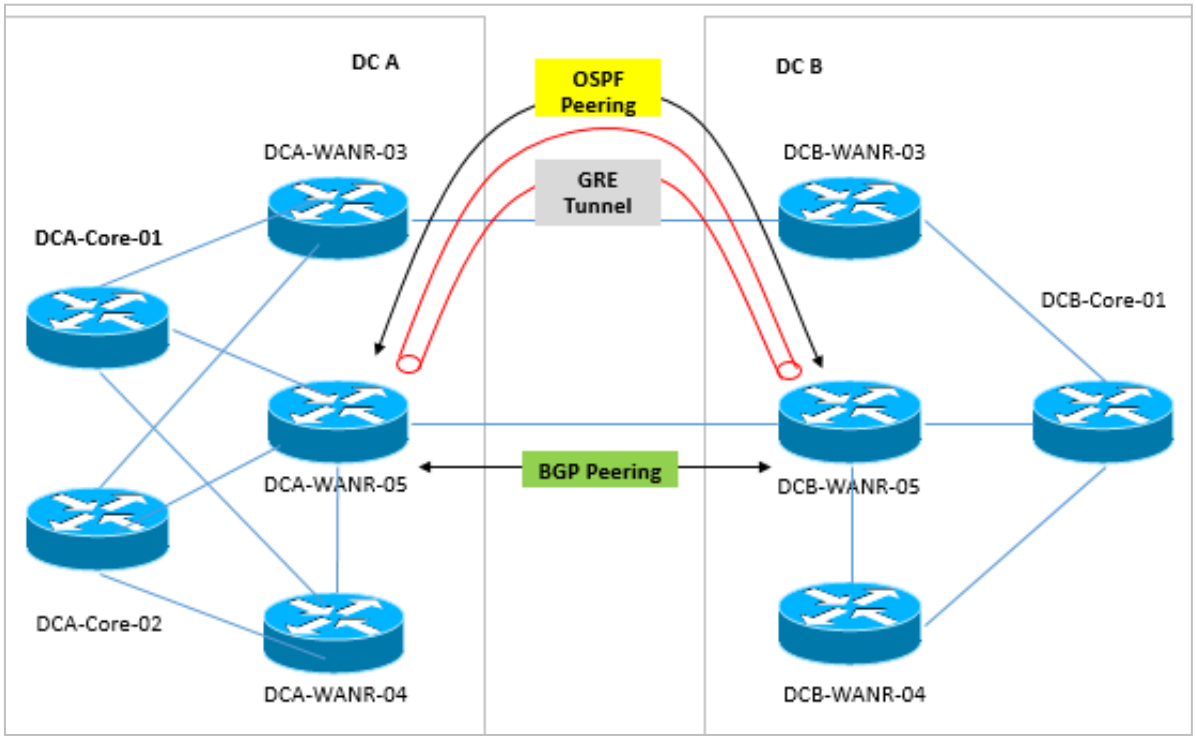

Fig.11. OSPF over GRE tunnel with inter-DC link between DCA-WANR-05 and DCB-WANR-05 in Normal Condition

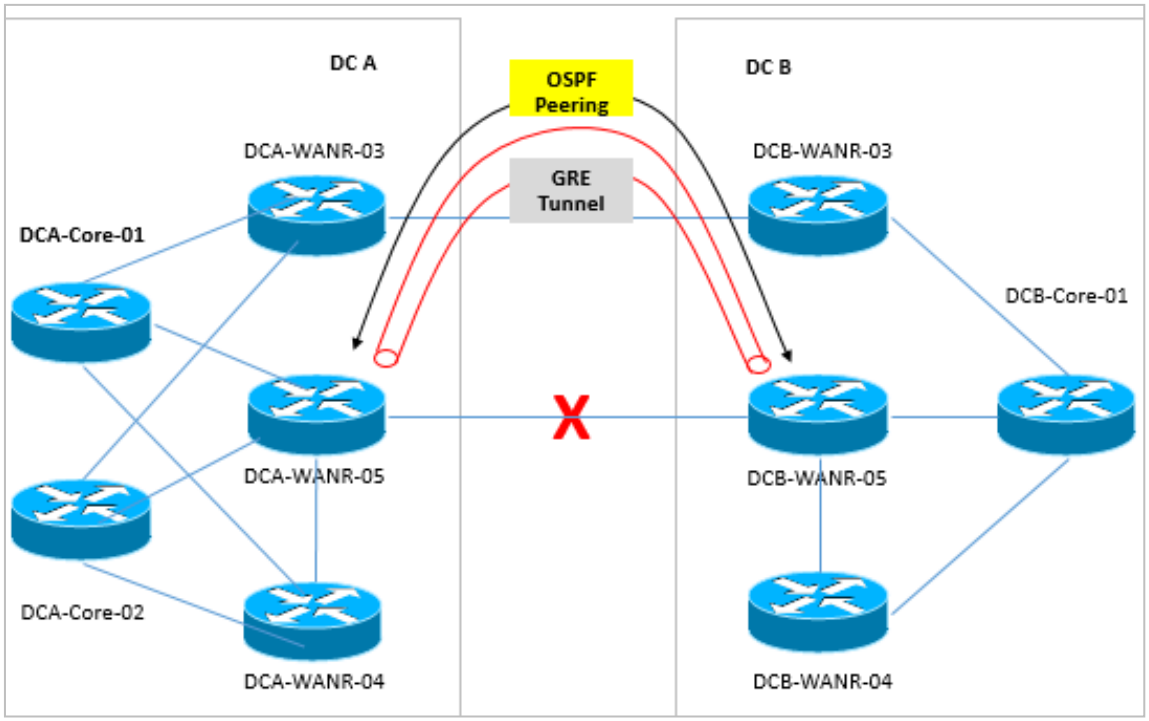

Fig.12. OSPF Peering over GRE tunnel is still available when the inter-DC link between DCA-WANR-05 and DCB-WANR-05 is disconnected.

After we define the configuration of OSPF peering over GRE Tunnel that needs to be applied to the network, then we implement it in the simulator. OSPF peering session information between DCA-WANR-05 and DCB-WANR-05 over GRE Tunnel is shown in fig. 13 and 14.

```
Neighbor 10.18.1.6, interface address 10.18.254.1
  In the area 1 via interface Tunnel100
  Neighbor priority is 0, State is FULL, 6 state changes
  DR is 0.0.0.0 BDR is 0.0.0.0
  Options is 0x52
  LLS Options is 0x1 (LR)
```

Fig.13. OSPF peers over GRE Tunnel (Interface Tunnel 100) on DCA-WANR-05 router (peering to DCB-WANR-05)

Once the OSPF peering over GRE Tunnel configuration has been implemented in the simulator, we simulate by disconnecting the inter-DC link between DCA-WANR-05 and DCB-WANR-05. The network traffic flow that going to intranet server after the inter-DC link between DCA-WANR-05 and DCB-WANR-05 disconnected can be seen in fig. 15.

Output of ping and telnet to TCP port 80 on HO-PC-01 to the intranet web server which ensures that the network traffic is successfully recovered after the inter-DC link between DCA-WANR-05 and DCB-WANR-05 is disconnected, can be seen below:

```
HO-PC-01#ping 10.15.3.100 repeat 20
Sending 20, 100-byte ICMP Echos to
10.15.3.100, timeout is 2 seconds:
!!!!!!!!!!!...!!!!!!
Success rate is 85 percent (17/20), round-
trip min/avg/max = 204/285/364 ms
HO-PC-01#telnet 10.15.3.100 80
Trying 10.15.3.100, 80 ... Open
test
HTTP/1.1 400 Bad Request
Date: Fri, 01 Mar 2002 00:19:41 GMT
Server: cisco-IOS
Accept-Ranges: none
400 Bad Request
[Connection to 10.15.3.100 closed by foreign
host]
HO-PC-01#
```

The output of the traceroute of the PC in the Head Office below ensures that now the network traffic flowing to the intranet server will enter DC-B first and then redirected to the DC-A via a new GRE tunnel between DCA-WANR-05 and DCB-WANR-05:

```
HO-PC-01#traceroute 10.15.3.100
Type escape sequence to abort.
Tracing the route to 10.15.3.100
   1 10.23.35.1 20 msec 68 msec 4 msec
   2 10.23.33.5 36 msec 4 msec 16 msec
   3 10.23.33.18 120 msec 48 msec 60 msec
   4 10.23.255.5 96 msec 180 msec 292 msec
←DCB-WANR-04
   5 10.19.0.54 288 msec 288 msec 148 msec
←DCB-WANR-05
   6 10.18.254.1 192 msec 80 msec 236 msec
←GRE Tunnel
   7 10.18.0.13 248 msec 160 msec 192 msec
   8 10.18.0.38 352 msec 256 msec 212 msec
   9 10.18.0.26 232 msec 184 msec 196 msec
  10 10.19.2.2 232 msec 268 msec 324 msec
  11 10.15.0.13 304 msec 276 msec 288 msec
  12 10.15.0.18 376 msec 280 msec 264 msec
```

Based on the simulation, we can argue that the OSPF

over GRE Tunnel solution solved the asymmetric routing problem on the “Organization A” network.

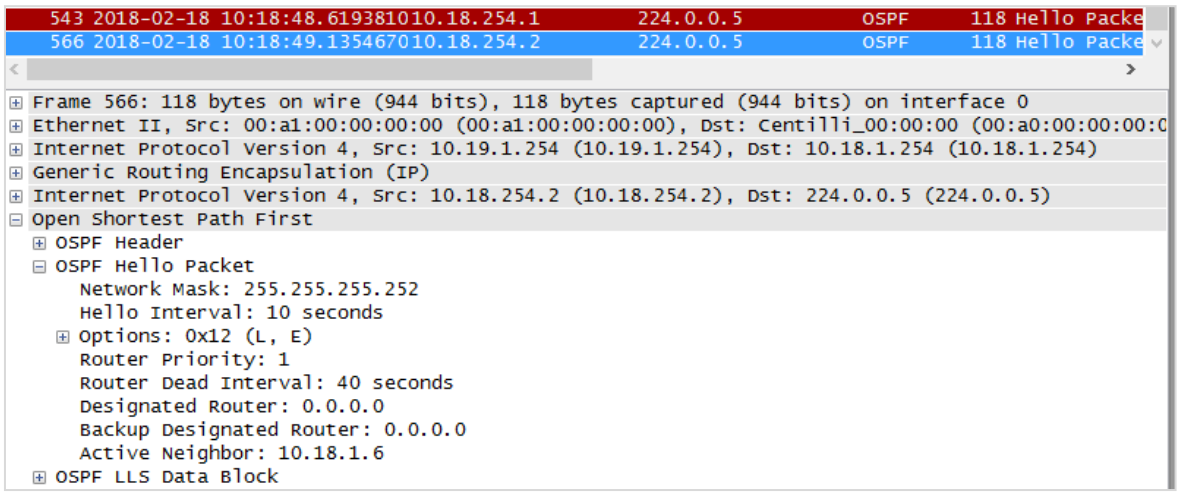

Fig.14. Network Packet Analyzer reveals OSPF Hello packet over GRE Tunnel

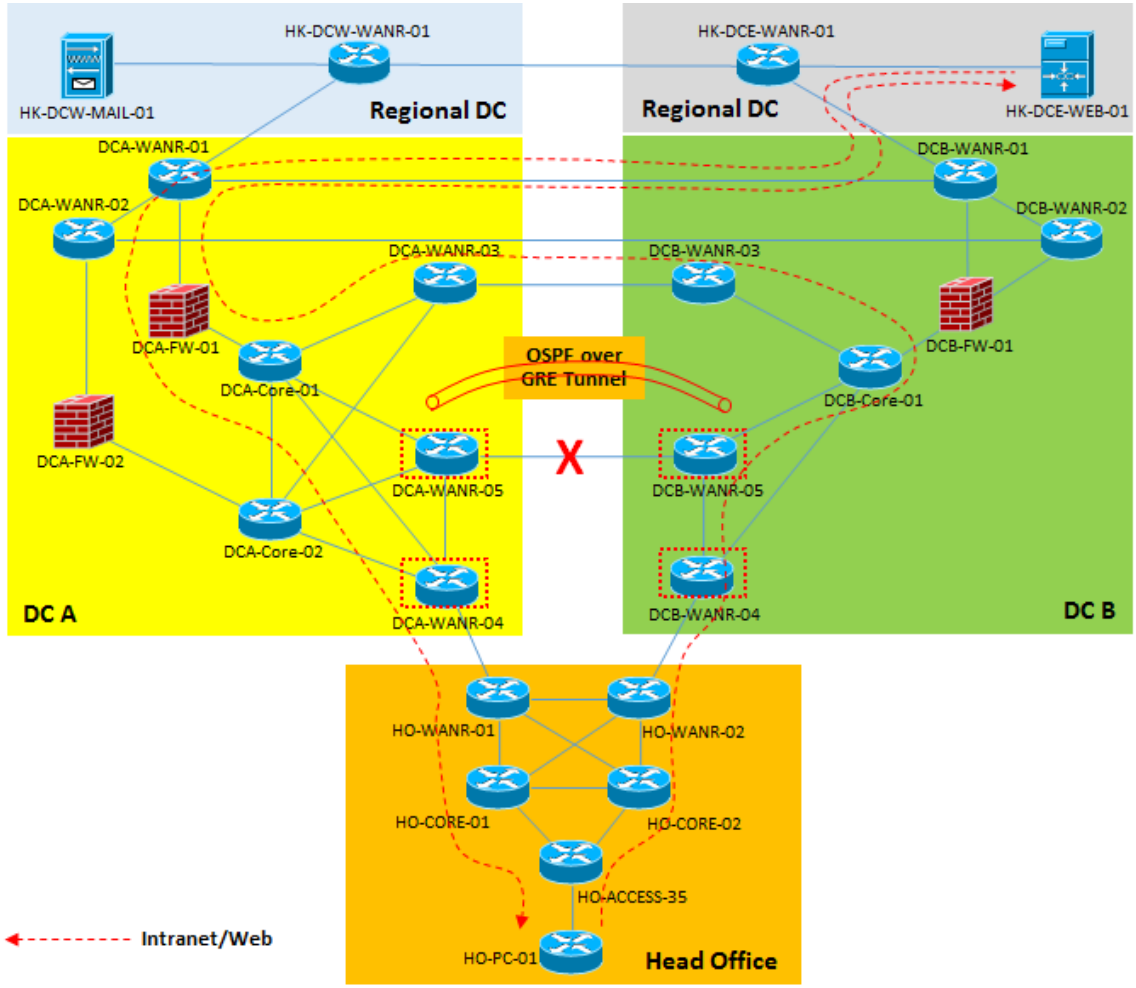

Fig.15. Traffic flow from Head Office to Intranet Server after the Inter-DC link disconnected and OSPF over GRE Tunnel solution applied

IV. Evaluation of the Proposed Solutions

We use the following parameters to evaluate both the solutions we offer:

1. Recovery time.
2. Packet loss.
3. ICMP echo-reply time.
4. TCP three-way handshake time for HTTP.

A. Recovery Time

Recovery time is how fast the flowing traffic can be recovered after the inter-DC link between DCA-WANR-05 to DCB-WANR-05 is disconnected. In every test, we record icmp echo packet time from client PC that does not get the reply from the intranet server (*IE*x). (*IE*x) is sent by the PC client to the intranet server before the inter-DC link between DCA-WANR-05 and DCB-WANR-05 is disconnected. We also record icmp echo packet time sent by the client PC to the intranet server after the traffic can be restored either by the BGP Community or OSPF over the GRE Tunnel solution (*IE*y). The recovery time (RT) can be calculated by using the following formula.

$$RT = IEy - IEx \quad (1)$$

ICMP echo is an ICMP packet with type 8 [18].

We tested both solutions 30 times. The result of all tests in terms of recovery time is shown in table 3.

Table 3. Recovery Time

Test	BGP Community (second) (a)	OSPF over GRE Tunnel (second) (b)
1	6.02035	6.14471
2	5.98595	5.97645
3	5.99584	6.25964
4	6.01240	5.99296
5	6.00371	6.18292
6	6.01631	6.02487
7	6.00475	6.14596
8	6.01561	6.01575
9	6.00210	6.00673
10	6.00322	6.55538
11	6.02087	5.89943
12	6.01559	5.98652
13	6.02839	6.18638
14	5.63500	6.36581
15	5.98416	5.95480
16	6.00923	6.22222
17	6.05173	5.99795
18	6.00117	6.66694
19	5.98637	5.99063
20	6.12940	6.17405
21	5.55354	6.34709
22	6.15758	6.14246
23	6.01870	6.30851
24	5.98481	6.96698
25	5.99764	5.99581
26	5.93496	6.02806
27	5.99675	6.36990
28	5.99669	6.02636
29	5.90589	5.99147
30	6.09340	6.16391
Average	5.98540	6.16969
Difference (b-a)		0.18428
Percentage of difference (b-a)		3.0%

Comparison of recovery time of each test for both solutions can be seen in fig. 16.

We can see that based on table 3 and fig. 16, when we use the BGP community solution, the average recovery time is 184.3 ms faster than OSPF over GRE tunnel.

According to the author's analysis, recovery time in BGP Community solutions is faster, because the router that decides to turn traffic to PDC after getting information about inter-DC interruption is the HO-WANR-01 router that has two BGP routes to go to the intranet server (10.15.3.100/32), which is through SDC via HO-WANR-02 router and through PDC via PDC-WANR-04 router. The best route is through SDC. When the inter-DC link is disconnected, the HO-WANR-01 router will receive this information and will delete the route to the intranet server via SDC. Then route to the intranet server via PDC will be the best BGP route and promoted to the routing table of the router.

So the processes performed by the HO-WANR-01 router are:

1. Delete BGP route to the intranet server 10.15.3.100 via SDC
2. Making the BGP route to the intranet server 10.15.3.100 via PDC (pre-existing in the BGP table as an alternate route) becomes the best BGP route to 10.15.3.100
3. Promote BGP route to the intranet server 10.15.3.100 to the routing table of the router.

As for the OSPF over GRE Tunnel solution, when the SDC-WANR-05 router receives disconnected information on PDC-SDC inter-DC links, the following are the following processes:

1. Perform a recalculation process to determine the best route to the intranet server 10.15.3.100 (partial process spfQ)
2. Added a new OSPF route to the intranet server 10.15.3.100 to the OSPF database.
3. Delete the OSPF database for the route to the intranet server 10.15.3.100 via PDC-SDC inter-DC link.
4. Promote the new OSPF route to the intranet server to the routing table of the router.

More processes are being performed on OSPF over the GRE Tunnel solution to find the best route after PDC-SDC inter-DC link interrupted. This makes recovery time on OSPF over GRE Tunnel solution longer than BGP Community solution.

The author's analysis is based on the process log of old route deletion and new route installation, either for OSPF over GRE Tunnel solution or BGP Community solution. The following is a log of the process of deleting and adding routes to the intranet server of the SDC-WANR-05 router when the PDC-SDC inter-DC link is disconnected (using OSPF over GRE Tunnel solution):

```
*Mar  1 01:17:02.579: RT: del 10.15.3.100/32
via 10.18.255.1, bgp metric [105/1000]
*Mar  1 01:17:02.579: RT: delete subnet route
to 10.15.3.100/32
*Mar       1    01:17:02.579:    RT:    NET-RED
10.15.3.100/32
*Mar   1 01:17:02.863: OSPF process partial
spfQ LSA id 10.15.3.100: mask 255.255.255.255,
type 5 adv_rtr 10.18.1.1, age 629, seq
```

```
0x80000006 (Area dummy area)
*Mar  1 01:17:02.863: OSPF: Start partial
processing Type 5 External LSA 10.15.3.100,
mask 255.255.255.255, adv 10.18.1.1, age 629,
seq 0x80000006, metric 1000, metric-type 2
*Mar  1 01:17:02.867: Add better path to LSA
ID 10.15.3.100, gateway 10.18.254.1, dist
1000
*Mar   1 01:17:02.867: Add path: next-hop
10.18.254.1, interface Tunnel100
*Mar   1 01:17:02.867: OSPF: delete lsa id
10.15.3.100, type 5, adv rtr 10.18.1.1 from
delete list
*Mar     1   01:17:02.867:   network   update
dest_addr 10.15.3.100 mask 255.255.255.255
gateway 10.18.254.1
*Mar   1 01:17:02.871: RT: SET_LAST_RDB for
10.15.3.100/32
NEW rdb: via 10.18.254.1
*Mar  1 01:17:02.871: RT: add 10.15.3.100/32
via 10.18.254.1, ospf metric [110/1000]
*Mar       1    01:17:02.871:    RT:    NET-RED
10.15.3.100/32
*Mar  1 01:17:02.875:     Add External Route
to  10.15.3.100.  Metric:  1000,  Next  Hop:
10.18.254.1
```

Based on the logs above, seen in the routing table of the old route deletion process and the new route promotion / installation takes 296 milliseconds (time in the last line of the logs minus the time on the first line).

The following are the process logs of old route deletion and new route installation when using BGP Community solution on the HO-WANR-01 router when PDC-SDC inter-DC link is disconnected:

```
HO-WANR-01#
Mar   1 03:07:01.659: BGP(0): 10.23.1.4 rcv
UPDATE about 10.15.3.100/32 -- withdrawn
Mar   1 03:07:01.667: BGP(0): Revise route
installing 1 of 1 routes for 10.15.3.100/32 ->
10.23.255.1(main) to main IP table
Mar     1   03:07:01.667:   RT:   closer   admin
distance for 10.15.3.100, flushing 1 routes
Mar       1    03:07:01.667:    RT:    NET-RED
10.15.3.100/32
Mar   1 03:07:01.667: RT: SET_LAST_RDB for
10.15.3.100/32
NEW rdb: via 10.23.255.1
Mar   1 03:07:01.671: RT: add 10.15.3.100/32
via 10.23.255.1, bgp metric [20/0]
Mar       1    03:07:01.671:    RT:    NET-RED
10.15.3.100/32
Mar     1   03:07:01.675:   BGP(0):   10.23.1.4
NEXT_HOP  part  1  net  10.15.3.100/32,  next
10.23.255.1
HO-WANR-01#
```

Based on the above logs, we can see that the old route deletion process and the promotion / installation of new routes in the BGP table take 12 milliseconds (faster than when using OSPF over GRE solution).

B. Packet Loss

In this test, packet loss is an ICMP echo packet (from client PC) that does not get icmp echo-reply packet (from intranet server) during the transition period. Number of packet loss after we tested two solutions we offered 30 times can be seen in table 4.

The comparison of number of ICMP-echo packet that does not get ICMP echo-reply packet for each test of both solutions can be seen in fig. 17.

Based on table 4 and fig. 17, we can see that when we use the BGP community solution, the average number of icmp echo-request packets that get no reply is 23.1% lower than the OSPF over GRE Tunnel solution.

Table 4. Number of ICMP Echo Packet That Does Not Get a Reply

Test	BGP Community Solution (a)	OSPF over GRE Tunnel (b)
1	3	4
2	3	3
3	3	4
4	3	3
5	3	3
6	3	3
7	3	4
8	3	3
9	3	3
10	3	6
11	3	4
12	3	3
13	3	4
14	3	5
15	3	3
16	3	4
17	3	3
18	3	7
19	3	3
20	3	3
21	3	5
22	3	4
23	3	5
24	3	4
25	3	3
26	3	4
27	3	6
28	3	3
29	3	4
30	3	4
Average	3	3.9
Difference (b-a)		0.9
Percentage of difference (b-a)		23.1%

The number of icmp echo-request packets that get no response is closely related to the recovery time value of each solution. So in accordance with the author's analysis on the recovery time test results, the number of ICMP-echo-request packets that get no response in the BGP Community solution were fewer, as the number of processes performed to remove and promote new routes to the intranet server in the routing table on the SDC-WANR-05 (on the OSPF over GRE Tunnel solution) is much more than the process done by the HO-WANR-01 router in the BGP Community solution.

C. ICMP Echo-request Reply Time (The Ping Delay)

This parameter is measured from the average ICMP

echo-reply packet time as the response packet for the ICMP echo packet sent by the client PC in the Head Office to the intranet server at Regional DC after the traffic is restored. In every test, we sent the icmp echo packet 100 times from the PC client to the intranet server after the traffic returned to normal while the inter-DC link remained disconnected. ICMP echo reply is an ICMP packet with type 0.

The average ping delay from the client PC to the intranet server when we use both solutions after the inter-DC link is disconnected can be seen in table 5.

Table 5. Average of ICMP Echo-reply Time

Test	BGP Community Solutions (miliseconds)	OSPF over GRE Tunnel (miliseconds)
1	247	394
2	230	312
3	285	373
4	240	381
5	258	351
6	229	320
7	234	311
8	232	286
9	244	293
10	225	333
11	224	303
12	231	309
13	228	323
14	228	297
15	226	301
16	224	324
17	226	299
18	229	298
19	240	323
20	230	330
21	244	292
22	244	296
23	246	294
24	226	309
25	221	306
26	231	317
27	219	351
28	224	295
29	233	326
30	233	303
Average	234	318
Difference (b-a)		84
Percentage of difference (b-a)		26.4%

The comparison of the average pings delay of each test for both proposed solutions is shown in fig. 18. Based on table 5 and fig. 18, we can see that when we use the BGP community solution, the average ping delay is 84ms (26.4%) faster than the OSPF over GRE tunnel solution

In BGP Community solution, the ICMP-echo-request reply time is faster because the number of hops traversed by PC HO-PC-01 to the Intranet server in the Regional Data Center is less than when using OSPF over GRE Tunnel solution, IE 10 hops versus 14 hops.

D. TCP Three-Way Handshake Time for HTTP

In every test, we record the time when the client PC initiates a connection by sending a TCP syn packet [19] to the intranet server with destination port 80 (TCP*s*) and the time when the client PC transmit TCP ack packet to the intranet server (TCP*a*), in response to the TCP syn-ack packet sent by the intranet server. The HTTP TCP three-way handshake time (TCP*t*) can be calculated by using the following formula.

$$TCPt = TCPa - TCPs \quad (2)$$

The time required to complete a TCP three-way handshake for HTTP connection between the client PC and intranet server when we use both solutions after inter-DC link is disconnected, can be seen in table 6.

Table 6. TCP Three-way Handshake Time for HTTP Connection

Test	BGP Community (a)	OSPF over GRE Tunnel (b)
1	238.4	374.1
2	242.4	538.3
3	258.8	349.7
4	238.2	355.8
5	238.3	397.8
6	228.4	263.9
7	313.9	368.7
8	199.2	377.7
9	281.3	280.8
10	252	296.3
11	219.6	220.7
12	252.4	319.1
13	320.2	274.2
14	187.7	382.2
15	218.5	484.2
16	201.4	293
17	193.5	320.7
18	232	369.1
19	338.4	295.8
20	176.1	388.6
21	232	396.9
22	231.8	328.2
23	189.2	308.4
24	183.2	231
25	178.6	463
26	230.8	330.4
27	290.5	379.8
28	221.8	328.9
29	199.7	361.5
30	196.4	254.6
Average	232.8	344.4
Difference (b-a)		111.6
Percentage of Difference (b-a)		32.4%

The comparison graph of the length of time it takes to complete a TCP three-way handshake for HTTP connection on both of our proposed solutions is shown in fig. 19.

Based on table 6 and fig. 19, we can see that when we use the BGP community solution, the average TCP three-way handshake time for HTTP connection is 111.6ms (32.4%) faster than the OSPF over GRE tunnel solution. This is because of the number of hops traversed by PC HO-PC-01 to the Intranet server when using the BGP Community solution less than when using the OSPF over GRE Tunnel solution, which is 10 hops to 14 hops.

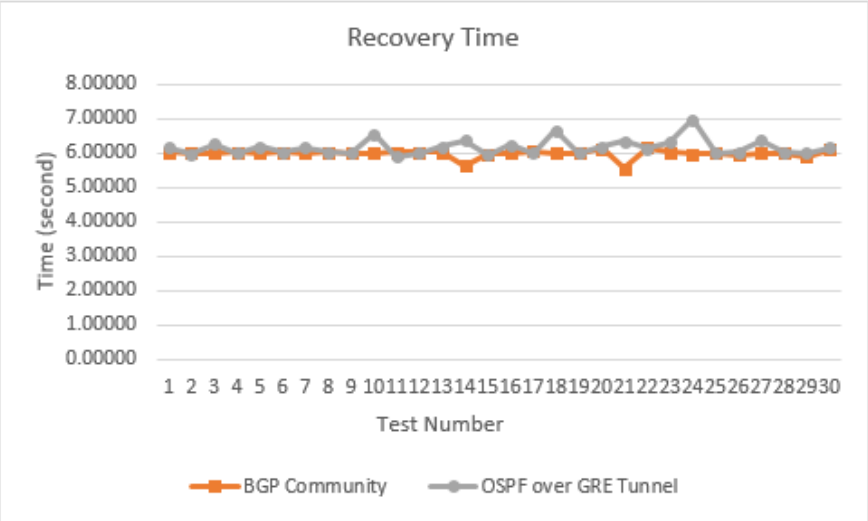

Fig.16. Recovery Time Graph

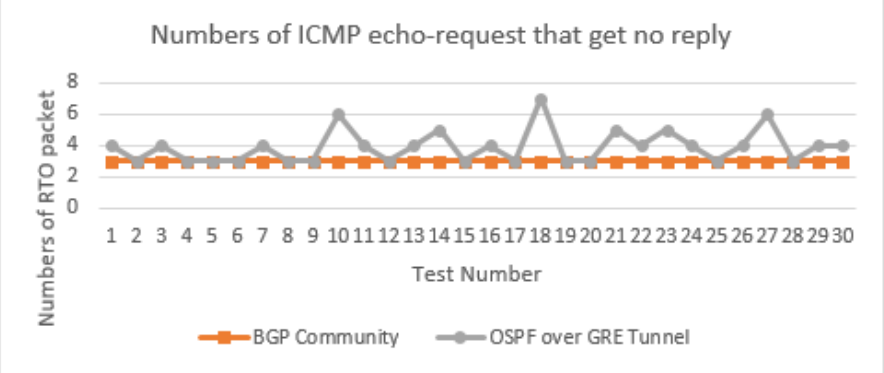

Fig.17. Number of ICMP Echo Packet That Does Not Get a Reply

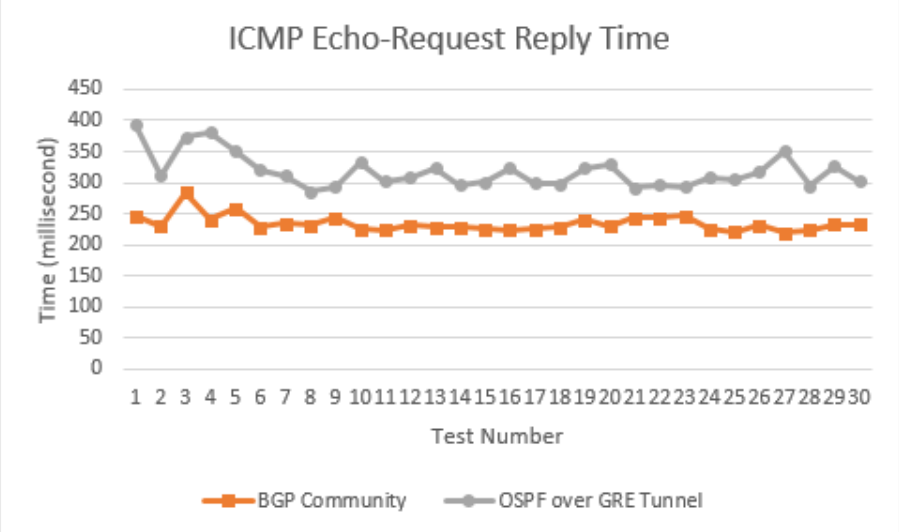

Fig.18. ICMP Echo-request Reply Time Graph

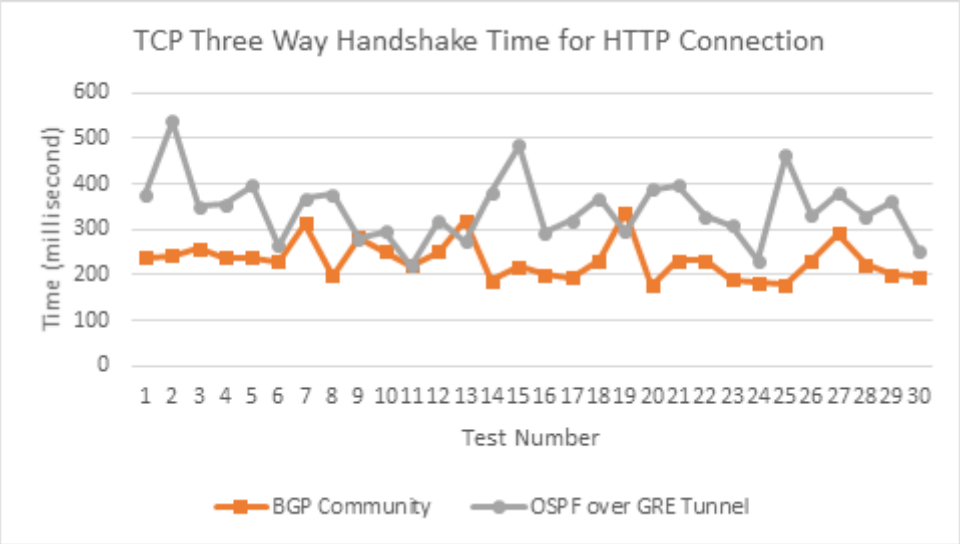

Fig.19. TCP three-way Handshake Time for HTTP Connection Graph

V. Conclusion

In the active-active data center that uses firewalls on each Data Center, asymmetric routing is very dangerous, because it will make the dissolution of communication between clients in the Data Center with the server on the other network. This happens because the initial traffic goes through a firewall located in one Data Center and its response traffic back through the firewall located in the other Data Center, so the response traffic will be dropped by the firewall. "Organization A" encountered this asymmetric routing problem.

We have simulated "Organization A"'s active-active Data Center network using the same IGP and EGP routing protocols, the same router operating system, the same routing policy and the same traffic flow. We also reproduce the same problem, ie dropped intranet web traffic when the inter-DC link is disconnected. Based on Analysis of "Organization A"'s network architecture and routing policy, we design, implement, verify and test the two solutions we offer using a simulator. Based on the 30 tests we conducted (that based on the 4 parameters we tested, IE recovery time, packet loss (the number of icmp echo-request packets that did not get replies), ping delay, and the time that required to complete the TCP three-way handshake for HTTP connection), we concluded that the BGP community solution is better than the OSPF over GRE Tunnel solution to be applied to "Organization A" networks, as well as other organizations that have similar network topology and routing policies.

References

[1] Osunade, O (2012). A Packet Routing Model for Computer Networks. International Journal of Computer Network and Information Security. 4 (2012) 13-20.

[2] Bogdanoski, M., Shuminoski, T., & Risteski, A. (2013). Analysis of the SYN Flood DoS Attack. International Journal of Computer Network and Information Security.8 (2013) 1-11.

[3] Durai, A. (2008). Asymmetric Routing and Firewalls. Retrieved from Cisco Systems: https://www.cisco.com/web/services/news/ts_newsletter/tech/chalktalk/archives/200903.html

[4] Piens, T. (2015). DotW: Issues with Asymmetric Routing. Retrieved from Palo Alto Networks: https://live.paloaltonetworks.com/t5/Featured-Articles/DotW-Issues-with-Asymmetric-Routing/ta-p/65456

[5] Magnani, D., Carvalho, I., Noronha, T. (2016). Robust Optimization for OSPF Routing. International Federation of Automatic Control. 49-12 (2016) 461-466.

[6] Periyasamy, P., Karthikeyan, E. (2013). Survey of Current Multipath Routing Protocols for Mobile AD Hoc Networks. International Journal of Computer Network and Information Security. 12 (2013) 68-79.

[7] Hiran, R., Carlsson, N., & Shahmehri, N. (2017). Collaborative Framework for Protection Against Attack Targeting BGP and Edge Networks. Computer Networks, 122 (2017) 120–137.

[8] Cheng, P., Zhang, B., Massey, D., & Zhang, L. (2010). Identifying BGP Routing Table Transfers. Computer Networks, 55 (2011) 636–649.

[9] Rekhter, Y., Li, T., (1995). RFC 1771: A Border Gateway Protocol 4 (BGP-4). Retrieved from Internet Engineering Task For19 (IETF): https://tools.ietf.org/html/rfc1771

[10] Cisco Systems (2016). BGP Best Path Selection Algorithm. Retrieved from Cisco Systems: https://www.cisco.com/c/en/us/support/docs/ip/border-gateway-protocol-bgp/13753-25.html

[11] Cisco Systems (2013). What is Administrative Distance?. Retrieved form Cisco Systems: https://www.cisco.com/c/en/us/support/docs/ip/border-gateway-protocol-bgp/15986-admin-distance.html

[12] Parkhurst, W. (2002). Cisco OSPF Command and Configuration Handbook.Cisco Press. Indianapolis, USA.

[13] Cisco Systems. (2013). Cisco IOS IP Routing: BGP Command Reference. Retrieved from Cisco Systems: https://www.cisco.com/c/en/us/td/docs/ios/iproute_bgp/command/reference/irg_book/irg_bgp4.html

[14] Juniper Networks. hold-time (Protocols BGP). Retrieved on 20 February 2018 from Juniper Networks: https://www.juniper.net/documentation/en_US/junos/topics/reference/configuration-statement/hold-time-edit-protocols-bgp.html

[15] Chandra, R., Traina P. (1996). RFC 1997: BGP Communities Attribute. Received from Internet Engineering Task Force (IETF): https://tools.ietf.org/html/rfc1997
[16] Dommety, G. (2000). RFC 2890: Key and Sequence Number Extensions to GRE. Retrieved from Internet Engineering Task Force (IETF): https://tools.ietf.org/html/rfc2890
[17] Farinacci, D., Li, T., Hanks, S., Meyer, D., Traina, P. (2000). RFC 2784: Generic Routing Encapsulation (GRE). Retrieved from Internet Engineering Task Force (IETF): https://tools.ietf.org/html/rfc2784
[18] Postel, J. (1981). Internet Control Message Protocol. RFC 1972. Retrieved from Internet Engineering Task Force: https://tools.ietf.org/html/rfc792.
[19] Pandya, P. (2013). TCP/IP Packet Analysis. Computer and Information Security Handbook. (2013) 499-512.

Authors' Profiles

Irwan Piesessa received the Bachelor of Computer degrees from the High School of Informatics Management Muhamadiyah Jakarta, Indonesia, in 2005. Currently Irwan is studying in the postgraduate program in Informatics Engineering Department at Bina Nusantara University, Jakarta.

Previously, he worked in one of the largest telecommunication companies in Indonesia as a network security engineer. Now he works in a global telecommunications company as a senior network consultant. He has more than 15 years experience in Enterprise, Service Provider, and Data Center network environment. He has a high interest in IP network management, network security and network design, planning and engineering.

Benfano Soewito completed his bachelor level education in the F MIPA (since April 10, 2008 changed into the Faculty of Science and Technology) Department of Physics, Airlangga University, Surabaya, Indonesia. Then Benfano completed his master level education in the field of Computer Engineering at the Department of Electrical and Computer Engineering, Southern Illinois University, United States of America (USA) in 2004. Benfano Completed his doctoral level education in 2009 at the same faculty and university.

After completing his undergraduate level education, he worked at a company in Surabaya, which produces crystal oscillator as the head of engineering with the main task in the development of design specification, manufacturing technique and quality improvement of the crystal oscillator unit.

While completing his master and doctoral degree, he also worked as an assistant professor, research assistant, support computer specialist, and developed several websites inside the environment of Southern Illinois University, USA. After finishing his doctoral studies, he chose to pursue a career in the education field by joining Bakrie University. Since 2013, he has became a full-time lecturer at the Faculty of Graduate Programs at Bina Nusantara University, Jakarta. He has a high interest toward researches in computer science with a special interest in information technology as include Internet packet processing and scanning, router development as well as security and computer network.

How to cite this paper: Irwan Piesessa, Benfano Soewito,"Ensure Symmetrical Traffic Flow, to prevent the Dropping of Response Packet by the Firewall, on the Active-Active Data Centers", International Journal of Computer Network and Information Security(IJCNIS), Vol.10, No.6, pp.1-15, 2018.DOI: 10.5815/ijcnis.2018.06.01