International Journal of scientific research and management (IJSRM)
||Volume||4||Issue||08||Pages||4530-4535||2016||
Website: www.ijsrm.in ISSN (e): 2321-3418

Effect of Route Reflection on IBGP Convergence and an approach to reduce convergence time

Santhosh S, M Dakshayini,

M.Tech, CNE Student
Dept. of ISE, BMSCE
santhosh.devadiga@gmail.com
Professor, Dept. of ISE,
BMSCE, Bengaluru
dakshayini.ise@bmsce.ac.in

Abstract— Request for Comments (RFCs) for Border gateway Protocol (BGP) suggest that the network topology using BGP must have full mesh of IBGP sessions to avoid routing loops. Routing loop is a condition where packets are routed between two or more routers resulting in slow convergence and routing instability. Route reflection is one of the strategies to avoid full mesh of internal BGP sessions between BGP speaking routers in an autonomous system. This paper focuses on demonstrating an effect of using route reflection on convergence time in a small network. A mixed network scenario of external and internal BGP sessions is considered for demonstration. Results of the simulation have shown that, route reflection can significantly increase the convergence time avoiding full mesh. *Also an approach to reduce the convergence time has been proposed*. Proposed algorithm is based on the principle of creating logical group of router reflectors. The same concept can be scaled for larger networks for different scenarios.

Keywords: IBGP full mesh, convergence time, route reflector, Convergence time reduction algorithm

I. INTRODUCTION

An autonomous system (AS) can be thought of as an entity comprising of a number of network devices under a single technical administration. The network devices concerned with are typically routers and layer three switches. For example consider a university campus having ten routers belonging to various departments. All these routers are managed by the university's IT department. Here an autonomous system consists of ten routers. In most of the cases one of the routers is connected to an ISP to provide internet service to the users or Virtual private network with other universities. Alternatively two or more routers can be used to connect to the outside world through multiple ISPs to provide resiliency and improve fault tolerance.

BGP is conceded as routing protocol of the internet backbone. There are two well-known forms of BGP called IBGP (internal BGP – within an autonomous system) and EBGP (External BGP -between multiple autonomous systems). BGP session between the two routers within a university campus is an example of IBGP. BGP session between a router in the university campus and ISP's router is an example of EBGP. Each autonomous system is assigned and uniquely identified by a number called AS number.

It is important to prevent routing loops in any of the routing protocol and BGP is not an exception. It follows different loop prevention strategies in EBGP and IBGP. In case of EBGP, packets cross multiple autonomous systems. This makes an obvious choice of using autonomous system number as a parameter to check for loop. A loop is detected when a router finds multiple entries for the same AS number in the AS path. AS path is the list of all AS numbers traversed by the packet.

Whereas IBGP cannot use this technique because routing updates does not go out of the AS. Only approach given by the initial BGP RFCs is to have full mesh of IBGP session amidst all the BGP speaking routers in the AS. This implies that a network containing N routers need to establish and manage N*(N-1)/2 IBGP sessions, which can be very hectic and tumultuous for the network administrators as they need to do manual configuration on all the routers.

Rest of the paper is organized as follows. Section II discusses study model for route reflection. Section III focuses on the impact on convergence time. Section IV illustrates an approach to reduce convergence time when route reflection is used.

II. STUDY MODEL OF ROUTE REFLECTION

.

A model considered for this work is as shown in figure 1. In this model, a selected router is made as router reflector (RR) in an AS. All IBGP speaking routers in the AS form IBGP sessions with route reflector only. All the non RR IBGP speaking routers are called route reflector clients (RRC). An RR receives BGP updates from an RRC and forwards it to other RRC or to an EBGP neighbor depending on the destination address.

Route reflection gained a lot of popularity from the outset of the concept of preventing full mesh in IBGP. Vendors and

.

operators started adopting this technology in their designs without detailed testing or analysis because of the reduced number of BGP sessions it results in.

In figure 1 Router – R1 is configured as an RR. Remaining three routers are configured as RRCs.

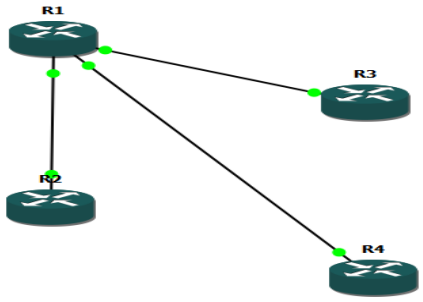

Figure1. R1 Configured as an RR

Multiple routers can be configured as RRs to improve fault tolerance at the RR level. If connectivity with primary route reflector RR1 and any of the RRCs goes down for some reason such as link or interface failure, power failure etc. R2 takes the responsibility of RR and BGP communication is unaffected. This is shown in figure 2.

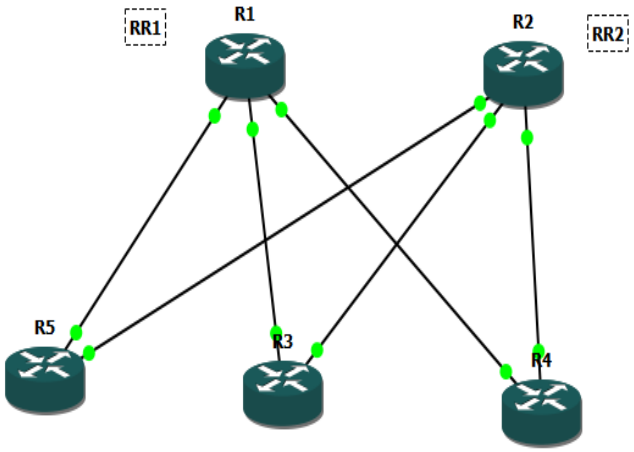

Figure2: Multiple RRs provide resiliency at the RR level.

Because of introduction of an RR in the AS, reachability information need to travel more than one hop to reach its destination within AS. This could be a potential reason for a loop if counter measures are not taken. Addition of two BGP attributes namely 'Cluster list' and 'Originator ID' takes care of the loop problem because of the extra hop. Originator ID is the router ID of the source router. Router ID is in general configured as highest loopback IP address on the router. Cluster list contains a list of routers IDs. A router updates the cluster list with its router ID if it is not found in the cluster list. The packet is discarded otherwise.

Router reflection has some advantages – it reduces the number of IBGP sessions drastically to N-1 where N is the number of BGP speaking routers in the AS. It may also reduce the operational expenditure to some extent i.e. addition and deletion of a BGP session becomes easier for the administrator. The size of the routing table and number of updates are also reduced compared to full mesh.

These advantages come at a cost. RR can affect the convergence time of BGP and it can even reduce the path diversity. *Path diversity* refers to the scenario where a router can select best path to a destination when multiple choices are available. A better path to a destination may not even be considered just because it is not learnt by an RR. This can affect QoS sensitive traffic. One of the simplest methods to avoid path diversity problem is to give meticulous attention to the placement of RR in the network design. A well placed RR can strategically solve the path diversity problem.

Effect on convergence time: using an example of a small network, effect of RR on convergence time can be studied. This study is specific to the scenario considered and the results cannot be generalized as there are various parameters which can affect the convergence time. For the same scenario convergence time is calculated using RR and full mesh for comparison.

III. IMPLEMENTATION TO CALCULATE CONVERGENCE TIME

Convergence time is one of the most important performance metrics of a routing protocol. When there is a change in the network topology, it takes some time for the update to reach all the routers in the network. Alternatively convergence can be thought of as the process of restoring a network back to its normal state when the problems that caused the outage are fixed.

In the considered example shown in figure 3 and figure 7 routers R1, R2, R3 and R4 are IBGP speakers with AS 299. R1 connects the AS to the external world.R1 has an EBGP session with R5. Router R5 is in AS 399. R1 and R4 are chosen as the farthest points in the network for implementation. Selecting farthest points gives better approximation of convergence time.

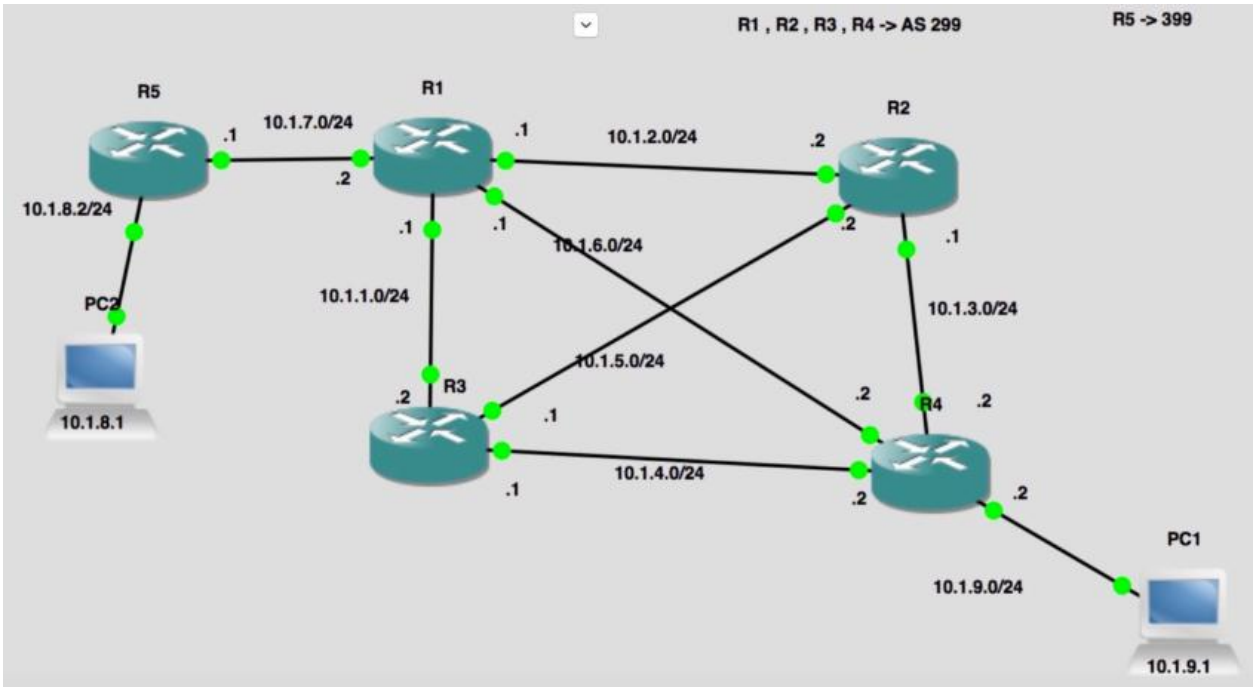

Figure 3: Full mesh IBGP

Figure 4 below shows the output taken at R4 proving PC2 is reachable from R4 using BGP.

```
R4#sh ip route 10.1.8.1
Routing entry for 10.1.8.0/24
  Known via "bgp 299", distance 200, metric 0
  Tag 399, type internal
  Last update from 10.1.7.1 00:17:25 ago
  Routing Descriptor Blocks:
  * 10.1.7.1, from 10.1.6.1, 00:17:25 ago
      Route metric is 0, traffic share count is 1
      AS Hops 1
      Route tag 399
```

Figure 4: routing table entry

A network outage is simulated by deliberately shutting down R1's interface that connects to R5. This makes R4 loose connectivity to R5 through all possible paths. Now convergence time can be measured as follows.

A continuous ping from PC1 to PC2 is started when the interface is turned up at R1 to bring back the connectivity to life. Figure 5 shows the instant at which it is turned up. Ping from PC1 to PC2 did not start to work immediately when problem is fixed. After a lot of packer drops ping starts to work. The round trip time for all ICMP packets are shown an each line of the ping output. Sum of all round trip times given an approximation of convergence time.

a. Continuous ping result at PC1 is shown in figure 6. Complete output could not be shown as the logs are more than a page. RTT for each packet is shown in the fourth field at each response. *Calculated convergence time is approximately equal to 20 seconds for full mesh topology*

```
R1(config-if)#
*Feb 26 00:04:05.839: %LINK-3-UPDOWN: Interface FastEthernet2/0, changed state to up
R1(config-if)#
*Feb 26 00:04:05.839: %ENTITY_ALARM-6-INFO: CLEAR INFO Fa2/0 Physical Port Administrative State Down
*Feb 26 00:04:06.839: %LINEPROTO-5-UPDOWN: Line protocol on Interface FastEthernet2/0, changed state to up
R1(config-if)#
*Feb 26 00:04:34.691: %BGP-5-ADJCHANGE: neighbor 10.1.7.1 Up
```

Figure 5: problem that caused the outage is fixed.

```
*10.1.9.2 icmp_seq=20 ttl=255 time=40.288 ms (ICMP type:3, code:1, Destination host unreachable)
*10.1.9.2 icmp_seq=21 ttl=255 time=5.128 ms (ICMP type:3, code:1, Destination host unreachable)
*10.1.9.2 icmp_seq=22 ttl=255 time=4.213 ms (ICMP type:3, code:1, Destination host unreachable)
*10.1.9.2 icmp_seq=23 ttl=255 time=2.018 ms (ICMP type:3, code:1, Destination host unreachable)
*10.1.9.2 icmp_seq=24 ttl=255 time=2.106 ms (ICMP type:3, code:1, Destination host unreachable)
*10.1.9.2 icmp_seq=25 ttl=255 time=3.274 ms (ICMP type:3, code:1, Destination host unreachable)
*10.1.9.2 icmp_seq=26 ttl=255 time=11.798 ms (ICMP type:3, code:1, Destination host unreachable)
*10.1.9.2 icmp_seq=27 ttl=255 time=45.254 ms (ICMP type:3, code:1, Destination host unreachable)
*10.1.9.2 icmp_seq=28 ttl=255 time=2.333 ms (ICMP type:3, code:1, Destination host unreachable)
*10.1.9.2 icmp_seq=29 ttl=255 time=4.620 ms (ICMP type:3, code:1, Destination host unreachable)
*10.1.9.2 icmp_seq=30 ttl=255 time=9.637 ms (ICMP type:3, code:1, Destination host unreachable)
10.1.8.1 icmp_seq=31 timeout
10.1.8.1 icmp_seq=32 timeout
84 bytes from 10.1.8.1 icmp_seq=33 ttl=61 time=44.440 ms
84 bytes from 10.1.8.1 icmp_seq=34 ttl=61 time=42.023 ms
84 bytes from 10.1.8.1 icmp_seq=35 ttl=61 time=37.120 ms
```

Figure 6: continuous ping from PC2 to PC1

b. Same experiment is repeated for this network setup by configuring R1 as route reflector. This is shown in figure 7.

For this scenario with R1 acting as route reflector, using the same technique described earlier convergence time is found to be approximately 30 seconds.

This clearly shows that introduction of an RR caused 50% increase in a very small network consisting of few routers.

IV. CONVERGENCE TIME REDUCTION PROTOCOL FOR IBGP NETWORKS WITH RRS

Considering a scenario where multiple RRs are used to provide route reflection and these RRs connect to multiple ISPs, a method can be applied to reduce convergence time. BGP routers use keep-alive packets to inform its presence to its neighbors. Default value of keep alive timer is one minute.

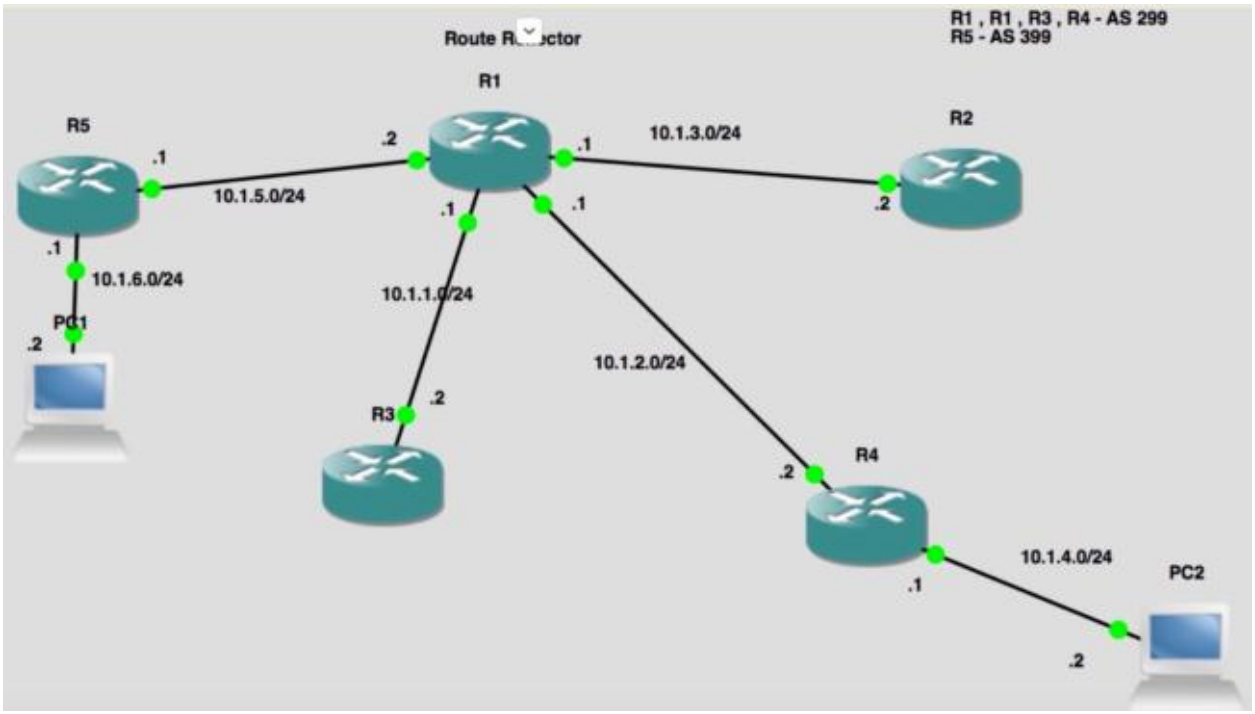

Figure 7: IBGP with Router reflector

A BGP speaker expects keep-alive packets from its active neighbors periodically every minute. When it does not receive keep-alive from a particular neighbor it waits for 3 keep-alive periods that is by default 3 minutes before terminating the BGP session. If it is possible to alter this behavior, convergence can be made up to 3 minutes faster than usual.
BGP policies must be defined in ordered to make the failover to second RR when the connection goes down to the primary RR. The amount of time for which it has to wait must be less than default hold timer value. This calls for a very high degree of robustness in the failover strategy among the RRs.

A logical group of all router reflectors can be created called RR_GROUP. An RR_GROUP is identified by using its unique logical GROUP_ID. Each RRC does not connect to individual RRs but it connects to RR_GROUP's logical ID. Each RR in a group is assigned a priority based on the computing abilities of the router, its position in the network etc. The priority assignment is done manually to avoid overhead of a master election mechanism.

This calls for necessity of a protocol that can effectively synchronize communication between route reflectors. The only goal of the protocol is to coordinate among all the route reflectors in RR_GROUP and communication from RR clients to RR Group. Each RR has an ID which is locally significant within the RR_GROUP called RR_ID. The status of current master RR is propagated using a small packet to all the backup RRs. This packet has very low overhead and includes only the status of the master RR and its RR_ID. Because of the small size of the packet it can be sent more frequently not causing any effect on the network performance. When the master RR status goes down, backup RR must pick up the role of master RR in next update period.

Reducing the hold timer to a suitable value and using the protocol explained above, a sufficient reduction in convergence time can be obtained.

Proposed convergence time reduction Algorithm

Step 1 : Initialization phase:

Each RR is given a RR_ID, it can be router's loopback IP for instance.

An RR_GROUP is given an ID called GROUP_ID. The IP addresses of the interfaces on all RRs that connect to the IBGP network are in the same subnet (IBGP LAN subnet).Then GROUP_ID can be an IP address from the same IBGP LAN subnet provided the ip address is used on any of the physical interfaces.

RR clients are aware of only the GROUP_ID. RR clients use GROUP_ID in a way similar to default gateway to communicate with RRs.

Step 2: Maintenance and failover phase:

The goal here is to ensure high availability.

Master RR is selected based on the computing abilities of the RR and its placement in the network.

Each RR is given a priority in the group. Active router at any point of time with the highest priority value becomes the master router in the group.

Low overhead keep alive packets are broadcasted by the master within the group periodically in short intervals.

When a member of the group does not receive the keep alive from master for 20 seconds it is considered dead and control is switched over to the router with next highest priority (back up RR). Backup RRs are made to wait only for one third of the standard BGP keep-alive period.

If back up RR also broadcasts its status to other routers, it is possible to take care scenarios where both current master and backup router become unavailable at the same time.

The keep alive packets contain only the necessary information such as Master route reflectors RR_ID, GROUP_ID, and priority value. This results in saving time when switching over from master to back up RR thus resulting in reduced convergence time.

V.CONCLUSION

Exact value of convergence time depends on several parameters such as size of the network, processing ability of the RR, strategic placement of the RR, number of RRCs, throughput of the links connecting routers etc. When using RRs in networks that require stringent QoS requirements it is very important to design the network to harness maximum performance. This paper provides a novel approach to reduce the convergence time by taking advantage of the BGP timers and creating a logical group of route reflectors.

ACKNOEDGEMENT

The authors would like to acknowledge and thank Technical Education Quality Improvement Program me [TEQIP] Phase 2, BMS College of Engineering and SPFU [State Project Facilitation Unit], Karnataka for supporting the research work

REFERENCES

[1] Yuri Breitbart ,Minos Garofalakis , Anupam Gupta , Amit Kumar ,"On Configuring BGP route reflectors" in conf rec Communication Systems Software and Middleware, 2007. COMSWARE 2007. 2nd International Conference , pp. 1-12

[2] Jong Han Park ,Ricardo Oliveira ,Shane Amante , Danny Mcpherson ,"BGP Route reflection revisited" IEEE Communications Magazine , July 2012 , pp 70-75

[3] Nick Feamster, Jennifer Rexford , "Network-Wide Prediction of BGP Routes" IEEE/ACM Transactions on Networking , 2007 , pp. 253-266

[4] R. Musunuri,J. A. Cobb " A complete solution for iBGP stability" Communications, 2004 IEEE International Conference June 2004, pp. 1177 – 1181

[5] U. Bornhauser , "Root causes for iBGP routing anomalies" Local Computer Networks (LCN), 2010 IEEE 35th Conference , pp 480 – 487

[6] H. Gobjuka , "Forwarding-loop-free configuration for IBGP networks" Networks, 2003. ICON2003. The 11th IEEE International Conference , pp 31 – 37

[7] Li Xiao , Communications, 2003. ICC '03. IEEE International Conference, pp 1765 - 1769 vol3

[8] V. Van den Schrieck , "BGP Add-Paths: The Scaling/Performance Tradeoffs" , IEEE Journal on

Selected Areas in Communications (Volume:28 , Issue: 8), pp 1299 – 1307

[9] B. Wang , "The research of BGP convergence time" , Information Technology and Artificial Intelligence Conference (ITAIC), 2011 6th IEEE Joint International , pp 354 - 357