International Refereed Journal of Engineering and Science (IRJES)
ISSN (Online) 2319-183X, (Print) 2319-1821
Volume 5, Issue 8 (August 2016), PP.16-23

Effective IBGP Operation without a Full Mesh topology

Muhammad H. Raza[1], Ankit K. Kansara[2], William Robertson[3]

The Date of Receipt and Acceptance Should Be Inserted Later

Abstract: This paper presents a set of novel and efficient set of alternative algorithms to the existing approaches to avoid a full mesh topology in internal border gateway protocol (iBGP). BGP exchanges routing information within an Autonomous System (AS) and among various ASes. All the routers inside an AS have to be connected in a full mesh topology to run iBGP. Managing a full mesh topology becomes difficult as the network size increases to reflect poor scalability. Route reflectors (RR) and BGP confederation are the two existing alternatives approaches to avoid full mesh topology at the cost of increased overheads and the possibility of network inconsistencies such as oscillations and persistent forwarding loops. The concept of central routing (CR) is an alternative solution, where the root node in an AS is responsible for all the control and management operations such as maintaining routing tables and calculating paths. The CR algorithms have been proven working successfully through simulations as a better option to avoid a full mesh oriented techniques for iBGP.

Keywords: BGP · iBGP · Route Reflector · BGP Confederation · Full Mesh Topology · Autonomous System

I. INTRODUCTION

Network management is a very vital factor for the proper operation of modern communication networks as a number of standard protocols are available for each layer of the TCP/IP model. Border Gateway Protocol (BGP), one of the Internet protocols, is implemented on gateway routers. An AS is a network

Muhammad H. Raza Internetworking Program, Faculty of Engineering, Dalhousie University, Halifax, B3H

or a group of networks, which is usually owned or controlled by a single ad- ministrative entity such as a business firm, a university, or a government or a non-government organisation.

BGP is configured on all the edge routers of an AS to exchange the routing information between them. BGP is divided into two groups: Exterior Border Gateway Protocol (eBGP) and Interior Border Gateway Protocol (iBGP). The routing information among various ASes and within one AS is specifically identified as Network Layer Reachability Information (NLRI) and routers have to maintain TCP connections among them by using port 179 for NLRI. BGP enabled routers that maintain TCP connections are known as neighbours or peers. Initially, the entire routing table is exchanged between peers and sub- sequently only updates are sent [1], [2], [3], and [4].

BGP operation faces a few operational challenges as described in [4]. For example, all the BGP enabled routers should be connected to each other in the form of a full mesh topology to implement iBGP. Forming a full mesh topology becomes cumbersome as the number of routers grows within an AS. Each router in an AS will exchange its full routing table with all other routers, which results in dramatic growth in the size of a routing table. Studies, such as [4] and [6], show an exponential growth in the size of routing tables in recent years. The configuration of a full mesh network topology cannot be scaled because each router in the topology has a BGP session with every other router [14], BGP update messages travel to all other routers in the topology, a copy of the advertisement is stored locally, and the addition of every new router in the topology required configuration of an iBGP session. The rapid growth in computing and storage technologies can provide an argument that a faster computation and more storage place can resolve the challenges faced by limited scaling capabilities of meshed BGP, but older and less capable routers cannot deal with the memory requirements and processing load, and cannot act as an effective gateway due to this increase in the size of routing tables. Replacing the older equipment with the newer equipment can translate into a huge investment.

The administrators of large networks use alternative concepts such as Route Reflector (RR) [11] and BGP confederation [12] to avoid the mesh topology requirement at the cost of increased overheads and the possibility of network inconsistencies. The concept of Central Routing (CR) is an alternative solution, where the root node in an AS is responsible for all the control and management operations such as maintaining routing tables and calculating paths. This paper presents a novel and efficient set of algorithms to imple- ment CR as an alternative to the existing approaches to avoid a full mesh topology in iBGP.

The rest of the paper is organized as follows. Section 2 analyzes the existing alternatives to avoid mesh topology. Section 3 reviews related work. Section 4 describes CR algorithms. Section 5 presents the simulation based evaluation of the algorithm. Section 6 concludes this paper and talks about the future work.

II. ANALYSIS OF EXISTING ALTERNATIVES OF MESH TOPOLOGY IN IBGP

When iBGP routers are connected via a full mesh topology, each router needs to maintain a Transport Control Protocol (TCP) session with all other iBGP routers in an AS. So whenever a router has any update to send, it has to send these updates to all other routers in an AS. If a router is connected to all other routers through one link, that link will be overloaded with all TCP traf- fic whenever there is any update. This raises a question of network scalability. RR [11] is a technique in which information is distributed without a full mesh topology of routers, which results in lowering network overheads. In RR, one router in an AS is selected as a Concentration Router or a RR which is con- nected to all other iBGP routers and sends updates to all other routers on behalf of the sender router [5]. The use of RR [11] in a networking topology can bring in certain advantages such as reduced memory and connection over- head, but these advantages come at the cost of increased complexity in the operation.

The drawback of RR is that it increases the overheads on the Concentra- tion Router, and if not configured properly, it may result in routing loops and an unstable network. Moreover, RRs have a single path for each destination, but this can create inconsistency when there is more than one RR in an AS as they may have different routes for a single destination, which may also result in routing loops. Also, the path selected by RR would not necessarily be the same as that of selected in the case of a full mesh topology [3] and [5]. Different RR may assign different paths to routers along a path and this will become a source of inconsistencies, which can cause improper functionality of protocols.

In BGP confederation, described in Internet Engineering Task Force (IETF) refrences; RFC5065 and RFC3065 [12], ASes with a large number of routers are divided into sub ASes. Each AS is identified globally by its number known as Autonomous System Number (ASN). Grouping one big AS into sub-ASes with different ASNs will increase the cost of purchasing ASNs and is not pragmatic to do so as ASNs are a limited resource. The sub-ASes in BGP confederation overcome these problems by being invisible to the outside world while any router of any sub-ASes is identified as a member of the main AS.

Though BGP confederation is a nice way to overcome the drawbacks of iBGP and route reflectors, but it increases the processing overhead and infor- mation complexity within and outside ASes. This overwhelmed complexity is the affect of a virtual BGP world that is created within an AS and that works like actual BGP. In many networks, BGP confederation works in conjunction with RR [3]. Overall, both the solutions, RR and BGP confederations, may not a) select the same feasible path as in case of a full mesh topology; b) cannot overcome the problem of exponential growth of routing tables, and c) updates go to all routers, which results in communication overheads.

III. ANALYTIC REVIEW OF RELATED WORK

Contemporary research on BGP has been under consideration by many re- search groups. For example, in [13], the authors devise an inference technique to pinpoint iBGP policies from public BGP data. They show that the majority of large transit providers and many small transit providers do apply policies iniBGP. They also propose configuration guidelines to achieve traffic engineer- ing goals with iBGP policies, without sacrificing BGP convergence guarantees
[13].

Similarly in [14], the element of this inter-domain routing system that has attracted the single-most attention within the research community has been the ”inter-domain topology”. [14] criticizes that almost from the get go, the vast majority of studies of this topology, from definition, to measurement, to modeling and analysis, have ignored the central role of BGP in this problem. [14] tries to establish that the legacy is a set of specious findings, unsub- stantiated claims, and ill-conceived ideas about the Internet as a whole. By presenting a BGP-focused state-of-the-art treatment of the aspects that are critical for a rigorous study of this inter-domain topology, the authors in [14] claim to demystify many ”controversial” observations reported in the existing literature.

In [15], by distinguishing between dissemination correctness and existing correctness properties, the authors show counter examples that invalidate some results in the literature. They claim to prove that deciding whether an iBGP configuration is dissemination correct is computationally intractable. They consider it even worse, determining whether the addition of a single iBGP ses- sion can adversely affect dissemination correctness of an iBGP configuration is also computationally intractable.

Our work falls in the category of earlier work (to eliminate the need for a full mesh topology in iBGP) on applying routing policy at route servers a the exchange points as presented in [7]. There are some other proposals that require very significant changes in the logic of BGP such as [8]. Whenever there is a discussion on the scale of BGP, many of the previous solutions take the system approach that requires modeling on the scale of an entire network and the decision making is also not on real time bases such as [9]. Another notable contribution is [10] that proposes an alternative to the full mesh topology in BGP but as per the critical review of this work presented in the following paragraphs; this solution becomes a lot complicated to use and is very resource intensive.

The solution in [10] consists of multiple building blocks and each module requires a different type of information and many constraints for assigning routes. For example, as per one constraint in [10], the IGP Viewer establishes IGP adjacency to one or more routers to receive IGP topology information. To ensure that the IGP Viewer never routes data packets, the links between the IGP Viewer and the routers should be configured with large IGP weights to ensure that the IGP Viewer is not an intermediate hop on any shortest path. An operator could dictate grouping for clustering routers needed for the Effective iBGP Operation Without a Full Mesh Topology 5 proper functioning of IGP Viewer. A complicated component of [10] is the BGP Engine that maintains an iBGP session with each router in the AS to learn about candidate routes and communicate its routing decisions to the routers.

Another complication in [10] needs the logic to assign routes in such a way that the routers along the shortest IGP path (from any router to its assigned egress router) must be assigned a route with the same egress router. Also, the RCS in [10] must assign a BGP route such that the IGP path to the next-hop of the route only traverses routers in the same partition.

Multiple complicated algorithms in [10] are responsible to implement a mul- tidimensional, complex, and complicated solution that requires a large number of assumptions and has to consider many operational constraints. In contrast to [10] and other previous research work, the work presented in this research paper narrowly focuses on the elimination of the requirement of a full mesh opology in iBGP with the help of only two simple algorithms and a few very pragmatic assumptions as described in the next session.

In [10], the evaluation is based on the performance as a function of the number of prefixes and routers by developing a router emulator tool that reads and plays back time stamped against the instrumented implementations of the components of [10]. The emphasis of testing in [10] is on performance parameters such as memory used, decision time, and overall processing time rather than proving the objective determination of the paths in iBGP.

Our paper objectively evaluates the performance of CR algorithms for their ability to determine routing paths in iBGP without the requirement of a full mesh topology with obvious benefit of a smaller routing table. The reduction in the size of the routing table is one of the major contributions of this work. CR based proposed scheme facilitates the accuracy of a full mesh topologywith better scalability capability.

IV. DESCRIPTION OF CENTRAL ROUTING (CR) ALGORITHMS

The proposal in this paper as an alternative approach to iBGP, provides theaccuracy of a full mesh topology and scalability of RR. In central routing, the path selection from a source to a destination is decided by the root router. When any iBGP router shows up in an AS, it sends its directly connected iBGP routers to the central node, and the central node updates its routing table accordingly. Two algorithms: (Central Routing Algorithm (CRA) and Algorithm for Shortest Path Search (ASPS)), are presented in the following two subsections.

4.1 Central Routing Algorithm (CRA)

In CRA, only the central node will have information regarding all the iBGP routers in an AS. Apart from the central node, all other iBGP routers will only 6 Muhammad H. Raza et al. have information about their directly connected neighbours. The discussion in this paper is based on clear assumptions and the algorithms are designed in such a way that the stated approach can be achieved with the implementation of a simple interface.

Pragmatic Assumptions:

1. Selection of Central Node will be predetermined, and there is only one central node per AS, however this limitation will be dealt with future changes in the CR algorithms according to the principles of backup and security features implemented on any router in a network.

2. The programming logic is installed on central node and an interface be- tween the programming logic and a message packet is capable of extracting and delivering information correctly.

3. The central node will not take part in any activity other than accepting neighboring information updates and calculating paths between sources and destinations.

4. The cost on each link is the same, but can be made adoptable to various cost functions in future.

The proposed algorithm has two parts; an algorithm for updating routing table and an algorithm for the shortest path search. The algorithm for updating a routing table consists of the following four steps and is explained in Figure 3.

Algorithm 1 Algorithm for Updating Routing Table

1. Search for source and peer id combination in the routing table.

2. If no match is found, it makes an entry in the table.

3. Go to Step 1 until all source-peer combinations are checked.
4. End.

The algorithm for the shortest path search is described in the following subsection.

4.2 Algorithm for Shortest Path Search (ASPS)

ASPS algorithm searches for the best possible path between a source and a destination and returns the entire string of every routers id between a source and a destination.

In case of more than one least hop count path between a source and a destination, the first matched path in database is selected. To search for the least hop count between a source and a destination, the algorithm accepts two parameters, source id and destination id.

The algorithm for the shortest path search consists of the following steps and is explained in Figure 4.

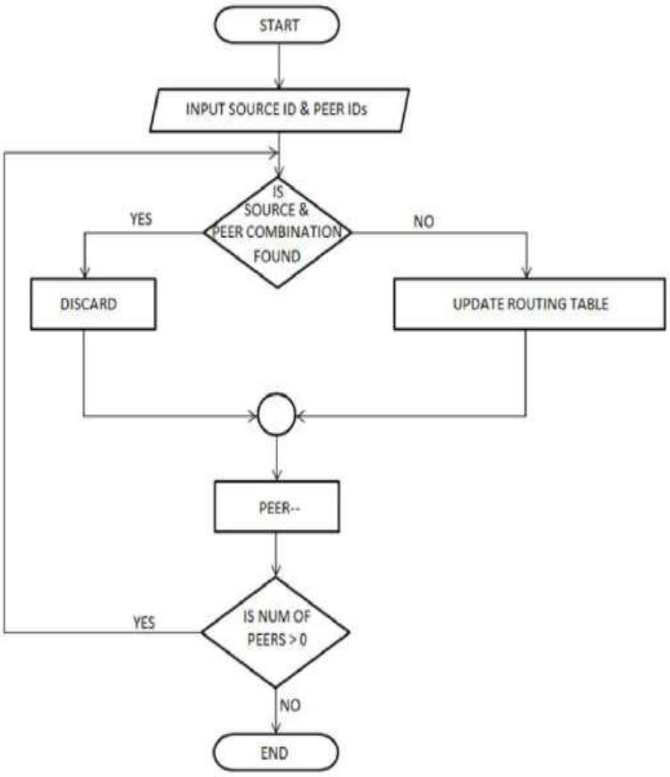
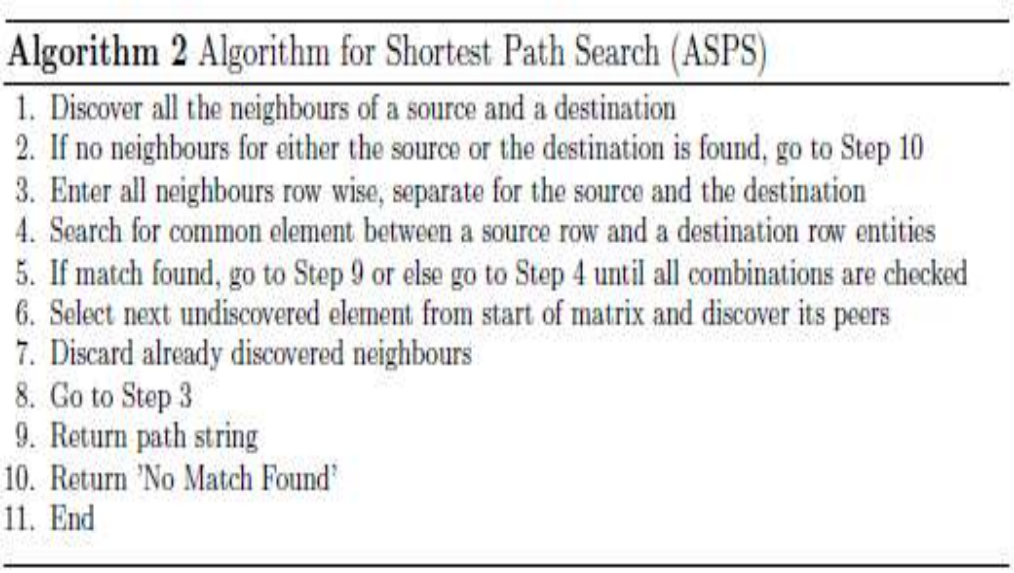


Fig. 1 Flowchart of the Algorithm for Updating Routing Tables



Algorithm 2 Algorithm for Shortest Path Search (ASPS)

```
1. Discover all the neighbours of a source and a destination
2. If no neighbours for either the source or the destination is found, go to Step 10
3. Enter all neighbours row wise, separate for the source and the destination
4. Search for common element between a source row and a destination row entities
5. If match found, go to Step 9 or else go to Step 4 until all combinations are checked
6. Select next undiscovered element from start of matrix and discover its peers
7. Discard already discovered neighbours
8. Go to Step 3
9. Return path string
10. Return 'No Match Found'
11. End
```

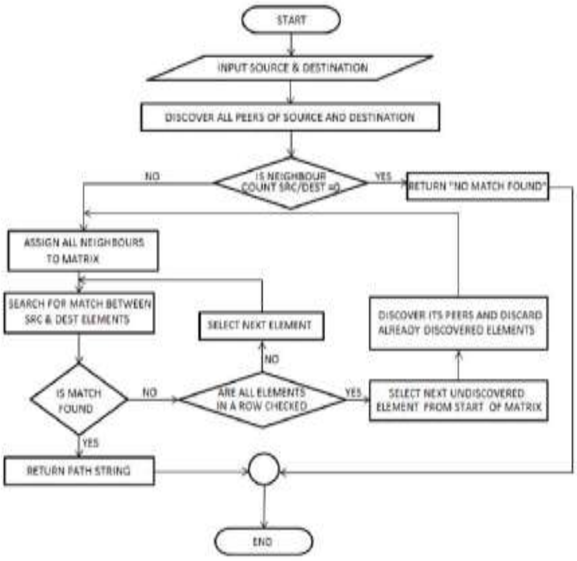

START
INPUT SOURCE & DESTINATION
DISCOVER ALL PEERS OF SOURCE AND DESTINATION
IS NEIGHBOUR COUNT SRC/DEST =0
NO
YES
RETURN "NO MATCH FOUND"
ASSIGN ALL NEIGHBOURS TO MATRIX
SEARCH FOR MATCH BETWEEN SRC & DEST ELEMENTS
SELECT NEXT ELEMENT
DISCOVER ITS PEERS AND DISCARD ALREADY DISCOVERED ELEMENTS
NO
IS MATCH FOUND
NO
ARE ALL ELEMENTS IN A ROW CHECKED
YES
SELECT NEXT UNDISCOVERED ELEMENT FROM START OF MATRIX
YES
RETURN PATH STRING
END

Fig. 2 Flowchart of the Algorithm for Shortest Path Search

5. Proof of the Concept of Central Routing Algorithms Through Scenario Based Evaluation

A network topology shown in Figure 5 was used to evaluate CR algorithms. The evaluation network consists of a total of 22 iBGP nodes in one AS.

If case of the use of a conventional iBGP with either a full mesh topology or RR or other such methods, each node keeps a track of all other routers in this AS. So at every node, a routing table will have at least 21 entries with one for each neighbour. In total, there will be 22 routing tables each with 21 entries, which will make 462 records. Moreover, the growth of a routing table will be exponential when newer nodes are added. This predicts large sized routing tables for medium to very large networks.

But unlike these traditional approaches, CR algorithms introduce a sharp plunge in the size of the routing tables at each node. The only node which will have a heavy routing table is the central node. All other nodes will have Effective iBGP Operation Without a Full Mesh Topology

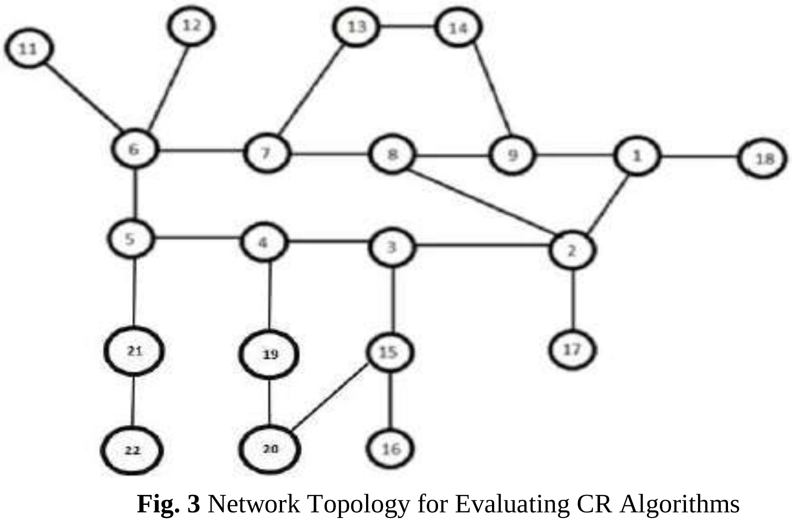


Fig. 3 Network Topology for Evaluating CR Algorithms

Comparatively small routing tables, with the entries of immediate neighbours only. This nature of the expected result will prove that CR algorithms reduces the table size and route exploration becomes efficient without the requirement of a full mesh topology.

For the evaluation of the CR algorithms, the simulation scenario and the CR algorithms were implemented with coding in Java and verified with simu- lations in NS-3. For better understanding, the processing is explained with the help of Java classes. To search for the least hop count between a source and
a destination, xget path method of xsearch path class will accept two param- eters, that is, a source id and a destination id. Method xget path will check whether the source id and destination id are same or not. If they are

not same, then it will first discover all the neighbours of both source and destination, and will arrange them in a two dimensional matrix. The size of matrix will be calculated by the number of nodes in an AS and the a connected neighbours to a router that has the highest id. Now all the entries will be matched with each other and if a match is found then a path will be returned. If a match is not found then starting from the beginning of the matrix, all the nodes will be explored and their neighbours will be discovered and arranged in a new row in the matrix. In this way, each of them (one by one) will be matched with their opposite nodes (source and destination neighbouring nodes are considered to be opposite nodes).

A user needs to define a source id and a destination id. Two sample runs of simulations are presented for the network shown in Figure 5.

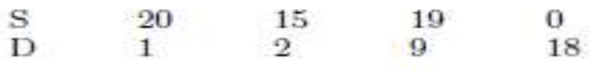
Initially, the algorithm checks whether the source and destination identifi- cations are same or not in an AS. If they are not the same, all the neighbours 10 Muhammad H. Raza et al. of both the source and destination are arranged in a two dimensional matrix. The size of the matrix is determined by the number of nodes in an AS and the number of highest connected neighbours to a router. The entries of this matrix are matched with each other and a match returns a path. If a match is not found, all the nodes are explored in the matrix, their neighbours are discovered and arranged as a new row in that matrix, and one by one each of them is paired as a source and destination pair.

In the 1st run, the source and destination ids are entered as 11 and 14 respectively. As a result, a path is obtained from node 11 to node 14 as a sequence of hops, that is 11.6.7.13.14. The resulting routing table of the central node of the sample network is shown in Table 4.

With Router 11 as a source and router 14 as a destination in an AS shown in Figure 5, Router 6 is the neighbour of the source router and Router 13 and Router 9 are the neighbours of the destination. The first iteration matrix will look like Table 1.

Along with the matrix of router ids, a separate array

Table 1 The First Iteration in the 1st Run

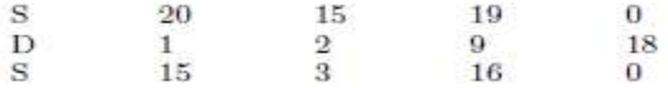


S	20	15	19	0
D	1	2	9	18

is also maintained to keep a track of router ids, and whether they belong to a source or a destination denoted by S and D respectively. The first column in the matrix is not included in the search and is needed only when a match is found between S and D to form a path. Also during the exploration, the nodes that have already been searched or discovered are not added again. In this example scenario, as there is no match between S and D, the first element of S, Router 6, is selected, and its neighbours are discovered and are matched with all the elements of D to get a matrix that is shown as Table 2.

The search and addition of new rows continues between S and D rows

Table 2 The Second Iteration in the 1st Run



S	20	15	19	0
D	1	2	9	18
S	15	3	16	0

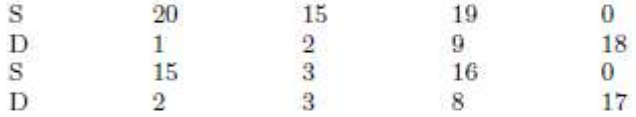
Until A match is found. The logic of CR algorithms avoids searching of the SS and DD pairs, but instead searches for an S and D pair.

Continuous discovery of the SD pairs shapes a matrix as shown in Table 3. The cells; c12, c13, and c31 mark the path towards a source and the cells: c22, c24, and c52 identify the paths towards a destination in Table 3.

In the 2nd run, the source and destination ids are entered as Router 19 and Router 1 respectively. As a result, a path is obtained from Router 19 to

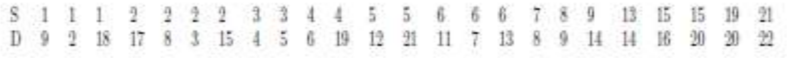
Effective IBGP Operation without a Full Mesh Topology 11

Table 3 The Third Iteration in the 1st Run



S	20	15	19	0
D	1	2	9	18
S	15	3	16	0
D	2	3	8	17

Table 4 Database of the Central Node in Figure 5



S	1	1	1	2	2	2	2	3	3	4	4	5	5	6	6	6	7	8	9	13	15	15	19	21
D	9	2	18	17	8	3	15	4	5	6	19	12	21	11	7	13	8	9	14	14	16	20	20	22

Router 1 as a sequence of hops, that is 19.4.5.6.7.8.9.1. The results of various iterations in the 2nd run are shown in Table 5, Table 6, Table 7, and Table 8. The explanation for various iterations in the 1st run holds true for the itera- tions in the 2nd run with different router ids.

From the evaluation process, it becomes clear that without a full mesh

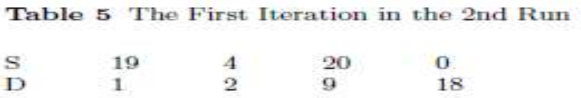


Table 5 The First Iteration in the 2nd Run

S	19	4	20	0
D	1	2	9	18

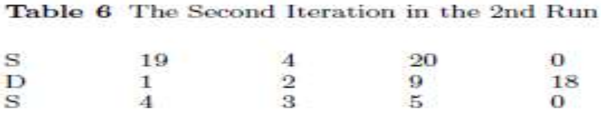


Table 6 The Second Iteration in the 2nd Run

S	19	4	20	0
D	1	2	9	18
S	4	3	5	0

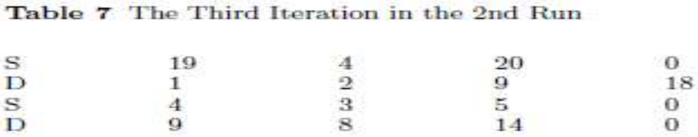
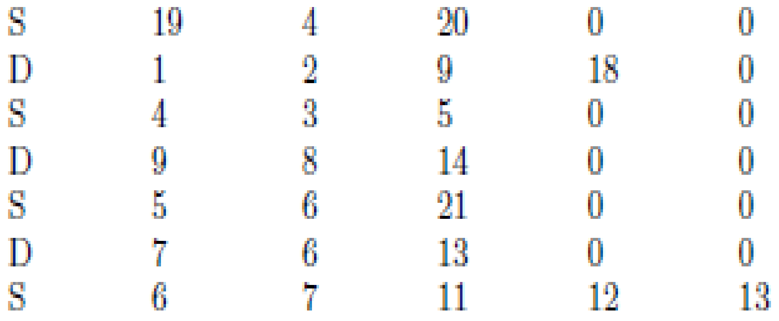


Table 7 The Third Iteration in the 2nd Run

S	19	4	20	0
D	1	2	9	18
S	4	3	5	0
D	9	8	14	0

Topology in iBGP and without the use of RR or BGP confederation, the CR iscapable of finding a path in iBGP. The sizes of the routing tables at all routers(except central node) is minimal and consisted of only immediate neighboursand is a maximum of four neighbors in the simulated topology as comparedto the requirement of a larger routing table at each node equal to 462 if afull mesh topology is used. The only node that has the largest routing table

Table 8 After a Few more iterations in the 2nd Run



S	19	4	20	0	0
D	1	2	9	18	0
S	4	3	5	0	0
D	9	8	14	0	0
S	5	6	21	0	0
D	7	6	13	0	0
S	6	7	11	12	13

Is the central node, Router 7, that has a maximum of 24 records. A clear cut reduction in the sizes of the routing tables in the entire topology makes the saving in the computation time and computational overhead with the use of CR scheme.

V. CONCLUSION AND FUTURE WORK

The algorithms to implement CR as an alternative to the existing approachesto avoid a full mesh topology in iBGP were proposed in this work. The fullmesh requirement was identified as the limitation of iBGP and weaknessesof the alternative solutions such as RR and BGP confederation were anal-ysed, and this analysis was the key motivation for this work. The proposedCR based scheme was successfully implemented through simulations and theresults proved CR to be a successful alternative to route reflectors and BGPconfederation. In future, the methods for the selection of a central node and its back upwill be investigated to add features to the existing CR methodology.

REFERENCES

[1]. Huitema, Christian. Routing in the Internet. Prentice Hall PTR, 1999.

[2]. Stewart III, J. W. (1998). BGP4: inter-domain routing in the Internet. Addison-Wesley Longman Publishing Co., Inc.

[3]. Zhang, Randy, and Micah Bartell. BGP design and implementation. Cisco Press, 2003.

[4]. Narayanan, Amit. ”A Survey on BGP Issues and Solutions.” arXiv preprint arXiv:0907.4815 (2009).

[5]. Park, J. H., Oliveira, R., Amante, S., McPherson, D., and Zhang, L. (2012). BGP route reflection revisited. Communications Magazine, IEEE, 50(7), 70-75.

[6]. http://www.cisco.com/web/about/ac123/ac147/, lastly accessed on May 29, 2014.

[7]. R. Govindan, C. Alaettinoglu, K. Varadhan, and D. Estrin, .Route servers for inter- domain routing,. Computer Networks and ISDN Systems, vol. 30, pp. 1157.1174, 1998

[8].] A. Basu, C. H. L. Ong, A. Rasala, F. B. Shepherd, and G. Wilfong, .Route oscillations in IBGP with route reflection,. in Proc. ACM SIGCOMM, August 2002.

[9]. O. Bonaventure, S. Uhlig, and B. Quoitin, .The case for more versatile BGP route re- flectors.. Internet Draft draftbonaventurebgproutere ectors00.txt, July 2004.

[10]. Caesar, Matthew, et al. ”Design and implementation of a routing control platform.” Proceedings of the 2nd conference on Symposium on Networked Systems Design and Implementation. Volume 2. USENIX Association, 2005.

[11]. Effective iBGP Operation Without a Full Mesh Topology 13 Bates, Tony and Chen, Enke and Chandra, Ravi ”BGP route reflection: An alternative to full mesh Internal BGP (IBGP)”, 2006.

[12]. Traina, Paul, Danny McPherson, and John Scudder. ”Autonomous system confedera- tions for BGP. No. RFC 5065”, 2007.

[13]. Vissicchio, Stefano and Cittadini, Luca and Di Battista, Giuseppe. ”On iBGP routing policies”, Networking, IEEE/ACM Transactions, IEEE, 227–240, 2015.

[14]. Roughan, Matthew, et al. ”10 lessons from 10 years of measuring and modeling the internet’s autonomous systems.” Selected Areas in Communications, IEEE Journal on 29.9 (2011): 1810-1821.

[15]. Vissicchio, Stefano, et al. ”iBGP deceptions: More sessions, fewer routes.” INFOCOM, 2012 Proceedings IEEE. IEEE, 2012.