http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 7 (1): 200-204, 2008
ISSN 1812-5638

Improvements in the Path Vector Approach for Inter-Domain Routing

Jawwad Haider, Muhmmad A.U. Khan and Sohail Razzaq
Department of Electrical Engineering, COMSATS Institute of Information Technology, Abbotabad, Pakistan

Abstract: The Border Gateway Protocol (BGP) is an inter-autonomous system routing protocol designed for TCP/IP Internets. Currently in the Internet Border Gateway Protocol (IBGP) deployments are configured such that all Border Gateway Protocol (BGP) speakers within a single Autonomous System (AS) must be fully associated so that any external routing information must be re-distributed to all other routers within that Autonomous System (AS). This represents a serious scaling problem. It over loads the memory and processor utilization of Internet backbone routers, increases the traffic load on internet backbone links there by increasing the over all convergence time of the Internet. This document describes the use and design of a method known as Path Relay (PR) technique to alleviate the need for full association (IBGP). This technique results in increasing the network performance in terms of backbone router processor and memory utilization backbone link bandwidth utilization and over all convergence time of the network. This performance up gradation is clearly evident from the results of the pilot deployment of PR.

Key words: Border gateway protocol, autonomous system, full association, path relay technique, convergence time, inter-domain routing

INTRODUCTION

Currently in the Internet, Border Gateway Protocol (BGP) deployments are configured such that all BGP speakers within a single Autonomous System (AS) must be fully associated and any external routing information must be re-distributed to all other routers within that AS. For n BGP speakers within an AS that requires to maintain n×(n-1)/2 unique internal BGP sessions. This full association requirement clearly does not scale when there are a large number of IBGP speakers each exchanging a large volume of routing information, as is common in many of today's internet networks. It over loads the memory and processor utilization of Internet backbone routers, increases the traffic load on internet backbone links there by increasing the over all convergence time of the internet (Fig. 1).

This scaling problem has been well documented and a number of proposals have been made to alleviate this (Haskin, 2004; Traina, 2005).

The main issue the with the IDRP route server (Haskin, 2004) technique is that it requires the internet backbone topology to be reconstructed which is a great management over head.

While the limited AS confederations for BGP (Traina, 2005) deals fairly well with the IBGP neighbors but it is not possible for the non complaint IBGP peers to be a part of original AS or domain without any loss of BGP network layer reach ability information.

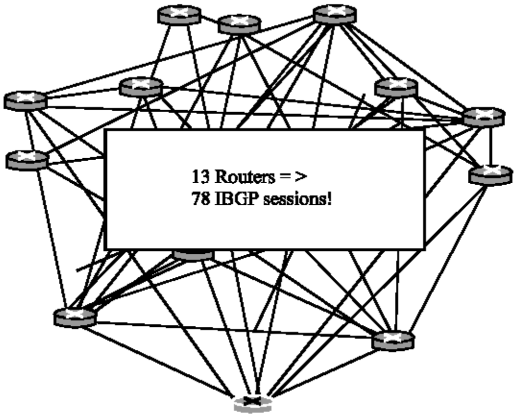

Fig. 1: Internal border gateway protocol scaling issues

This document represents another alternative in alleviating the need for a full association and is known as PR technique. This approach allows a BGP speaker (known as Path Relay) to advertise IBGP learned routes to certain IBGP peers. It represents a change in the commonly understood concept of IBGP and the addition of two new optional transitive BGP attributes to prevent loops in routing updates.

This study is made to satisfy the following objectives:

- Any alternative must be both simple to configure as well as understand.

Corresponding Author: Jawwad Haider, Department of Electrical Engineering, COMSATS Institute of Information Technology, Abbotabad, Pakistan

- It must be possible to reduce the Internet backbone link bandwidth utilization, load on processor and memory of back bone routers.
- The alternative must be able to reduce the convergence time of the internet for its better performance.
- It must be possible to transition from a full association configuration without the need to change either topology or AS. This is an unfortunate management overhead of the technique proposed in Haskin (2004).
- It must be possible for non compliant IBGP peers to continue be part of the original AS or domain without any loss of BGP routing information. Which is a deficiency in the technique proposed in Traina (2005).

These objectives are motivated by operational experiences of a very large and topology rich network with many external connections.

PATH RELAY (PR) TECHNIQUE

The basic idea of PR is very simple to comply with first objective of present study. Let us consider the simple example shown in Fig. 2.

In AS n there are three IBGP speakers (routers R-1, R-2 and R-3). With the existing BGP model, if R-1 receives an external route and it is selected as the best path it must advertise the external route to both R-2 and R-3. R-2 and R-3 (as IBGP speakers) will not re-advertise these IBGP learned routes to other IBGP speakers.

If this rule is relaxed and R-3 is allowed to advertise IBGP learned routes to IBGP peers, then it could re-advertise (or relay) the IBGP routes learned from R-1 to R-2 and vice versa. This would eliminate the need for the IBGP session between R-1 and R-2 as shown in Fig. 3.

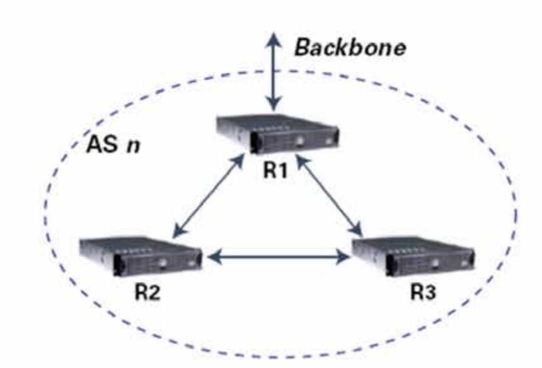

Fig. 2: Full association internal border gateway protocol

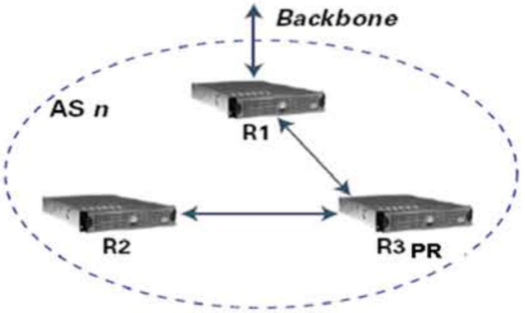

Fig. 3: Path relay scheme internal border gateway protocol

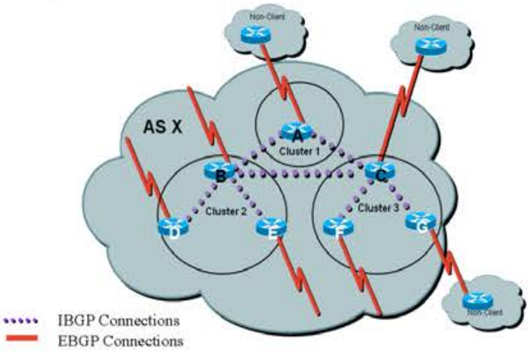

Fig. 4: PR Terminology

The PR scheme is based upon this basic principle.

TERMINOLOGY AND CONCEPTS

We use the term PR to describe the operation of a BGP speaker advertising an IBGP learned route to another IBGP peer. Such a BGP speaker is said to be a PR and such a route is said to be a relayed route.

The internal peers of a PR are divided into two groups:

- Client peers
- Non-Client peers

A PR relays routes between these groups and may relay routes among client peers. A PR along with its client peers forms a Cluster. The Non-Client peer must be fully associated but the Client peers need not be fully associated. Figure 4 shown a simple example outlining the basic PR components using the terminology noted above. Routers A, B and C are the Path Relays in the clusters 1, 2 and 3, respectively (Fig. 4).

Cluster 1 does not have no client or non-client peers, while routers D and E are the PR client peers in cluster 2 and routers F and G are the client peers in cluster 3.

OPERATION

When a PR receives a route from an IBGP peer, it selects the best path based on its path selection rule. After the best path is selected, it must do the following depending on the type of the peer it is receiving the best path from:

- Route from Non-Client IBGP peer
  Reflect to all the Clients.
- Route from a Client peer

Reflect to all the Non-Client peers and also to the Client peers (Hence the Client peers are not required to be fully associated).

An AS could have many PRs. A PR treats other PRs just like any other internal BGP speakers. A PR could be configured to have other PRs in a Client group or Non-client group.

In a simple configuration the backbone could be divided into many clusters. Each PR would be configured with other PRs as Non-Client peers (thus all the PRs will be fully associated). The Clients will be configured to maintain IBGP session only with the PR in their cluster. Due to path relaying, all the IBGP speakers will receive reflected routing information.

It is possible in an AS to have BGP speakers that do not understand the concept of PR (let us call them conventional BGP speakers). The Path Relay Scheme allows such conventional BGP speakers to co-exist. Conventional BGP speakers could be either members of a Non-Client group or a Client group. This allows for an easy and gradual migration from the current IBGP model to the Path Relay model. One could start creating clusters by configuring a single router as the designated PR and configuring other PRs and their clients as normal IBGP peers. Additional clusters can be created gradually.

REDUNDANT PRs

Usually a cluster of clients will have a single PR. In that case, the cluster will be identified by the ROUTER_ID of the PR. However, this represents a single point of failure so to make it possible to have multiple PRs in the same cluster; all PRs in the same cluster can be configured with a 4-byte CLUSTER_ID so that a PR can discard routes from other PRs in the same cluster.

LOOP AVOIDANCE

When a route is relayed, it is possible through mis-configuration to form route re-distribution loops. The PR method defines the following attributes to detect and avoid routing information loops:

ORIGINATOR_ID: ORIGINATOR_ID is a new optional, non-transitive BGP attribute. This attribute is 4 bytes long and it will be created by a PR in relaying a route. This attribute will carry the ROUTER_ID of the originator of the route in the local AS. A BGP speaker should not create an ORIGINATOR_ID attribute if one already exists. A router which recognizes the ORIGINATOR_ID attribute should ignore a route received with its ROUTER_ID as the ORIGINATOR_ID.

CLUSTER_LIST: Cluster-list is a new optional, non-transitive BGP attribute. It is a sequence of CLUSTER_ID values representing the relayed paths that the route has passed.

When a PR relays a route, it must prepend the local CLUSTER_ID to the CLUSTER_LIST. If the CLUSTER_LIST is empty, it must create a new one. Using this attribute a PR can identify if the routing information is looped back to the same cluster due to mis-configuration. If the local CLUSTER_ID is found in the cluster-list, the advertisement received should be ignored.

CONFIGURATION AND DEPLOYMENT CONSIDERATIONS

Care should be taken to make sure that none of the BGP path attributes defined above can be modified through configuration when exchanging internal routing information between PRs and Clients and Non-Clients. Their modification could potential result in routing loops.

In addition, when a PR relays a route, it should not modify the following path attributes: NEXT_HOP, AS_PATH, LOCAL_PREF and MED (Rekhter and Li, 2005) their modification could potential result in routing loops.

The BGP protocol provides no way for a Client to identify itself dynamically as a Client of a PR. The simplest way to achieve this is by manual configuration.

One of the key components of the path relay approach in addressing the scaling issue is that the PR summarizes routing information and only relays its best path, there by reducing the loads on the backbone internet links bandwidth and routers memory and processors and hence comply with objectives of present study.

Both MEDs and IGP metrics may impact the BGP oute selection. Because MEDs are not always comparable and the IGP metric may differ for each router (Johnson and Martey, 2002), with certain path relay topologies the path relay approach may not yield the same route selection result as that of the full IBGP associated approach. A way to make route selection the same as it would be with the full IBGP associated approach is to make sure that path relays are never forced to perform the BGP route selection based on IGP metrics which are significantly different from the IGP metrics of their clients, or based on incomparable MEDs (Halabi, 2000). The former can be achieved by configuring the intra-cluster IGP metrics to be better than the inter-cluster IGP metrics and maintaining full association within the cluster. The latter can be achieved by:

- Setting the local preference of a route at the border router to relay the MED values.
- Or by making sure the AS-path lengths from different ASs are different when the AS-path length is used as a route selection criteria.(Russ White and Mcpherson, 2004).
- Or by configuring community based policies using which the relay can decide on the best route (Siganos and Faloutsos, 2004).

One could argue though that the latter requirement is overly restrictive and perhaps impractical in some cases. One could further argue that as long as there are no routing loops, there are no compelling reasons to force route selection with path relays to be the same as it would be with the full IBGP association approach.

To prevent routing loops and maintain consistent routing view, it is essential that the network topology be carefully considered in designing a path relay topology. In general, the path relay topology should be congruent with the network topology when there exist multiple paths for a prefix (Johnson and Martey, 2002). One approach could be the POP-based relay, in which each POP maintains its own path relays serving clients in the POP and all path relays are fully associated. In addition, clients of the relays in each POP are often fully associated for the purpose of optimal intra-POP routing and the intra-POP IGP metrics are configured to be better than the inter-POP IGP metrics.

STATISTICS OF A PILOT DEPLOYMENT

Based upon the above concepts of path relay technique a test implementation on a pilot network is done with 13 routers and 78 IBGP full association links. The detail of the equipment and links used is given as follows:

Table 1: Over all network performance comparison

Sr. No.	Without PR (A)	With PR (B)	Difference C = [(A-B)/A]×100%
Memory Utilization (%)	11.3	3.7	67.25
CPU Utilization (%)	5.6	2.1	62.50
Link bandwidth Utilization (%)	20.0	3.1	74.50
Convergence time	29.0 sec	21.0 sec	27.60%

Router used	:	Cisco 2620 XM series
Standard Memory of Each	:	64 MB DRAM
CPU: Intel CPU	:	Dual P3 1.4 Ghz 128k Cache
WAN Link capacity	:	64Kbps

Average of 13 Routers and 78 Links

The column A in the following table gives us the percentage utilization of routers memory, CPU, link bandwidth utilization and the convergence time of the network, respectively without using PR in the network.

While the column B in the above comparison table represents the same items but with incorporating the PR in the network. It is clearly that the percentage improvement in the overall network performance in terms of reduces memory and CPU utilization at each router and the link bandwidth utilization between two IBGP routers and the convergence time of the whole network after a routing update. Hence the results clearly comply with present study objectives. This performance upgrade is shown in the column C of the performance comparison (Table 1).

CONCLUSIONS

Although this extension to BGP does not change the underlying security issues inherent in the existing IBGP (Heffernan, 1998). But still it should be considered as a major breakthrough in resolving Internet scalability issues. To enumerate the benefits of this technique; following are few: It solves the IBGP full-association problem by accommodating the no complaint IBGP peers (Traina, 2005) complying with objective of present study and without requiring any change in internet topology (Haskin, 2004) complying with objective of present study. It would be beneficial for ISPs when the number of internal neighbor statements becomes excessive. Packet forwarding will not be affected in the internet plane. As depicted vividly in our test deployment results discussed above, that incorporation of this feature in standard BGP protocol will make very low memory and CPU and link bandwidth usage and the over all performance of the Inter-domain routing will improve substantially.

ACKNOWLEDGMENTS

The author would like to thank Philip Smith and Khalid Raza for the many discussions resulting in this research. This idea was developed from an earlier discussion between Tony Li and Dimitri Haskin.

In addition, the author would like to acknowledge valuable review and suggestions from Dr. Asmat Ullah on this document.

REFERENCES

Halabi, S., 2000. Internet Routing Architectures. 2nd Edn., Cisco Press.

Haskin, D., 2004. A BGP/IDRP route server alternative to a full mesh routing. RFC, 1863.

Heffernan, A., 1998. Protection of BGP sessions via the TCP MD5 Signature Option. RFC., 2385.

Johnson and Abe Martey, 2002. Troubleshooting IP Routing Protocols. 1st Edn., Cisco Press.

Rekhter, Y. and T. Li, 2005. A Border Gateway Protocol 4 (BGP-4). RFC, 1771.

Siganos, G. and M. Faloutsos, 2004. Analyzing BGP policies: Methodology and Tools. http://www.ieee-infocom.org/2004/Papers/34_5.

Traina, P., 2005. Limited autonomous system onfederations for BGP. RFC, 1965.

White, R. and D. Mcpherson, 2004. Practical BGP. 1st Edn., Addison-Wesley Professional.