ISSN 2320-9984
Volume 1, Issue 1, April 2013 ISSN: 2320-9984 (Online)
International Journal of Modern Engineering & Management Research
Website: www.ijmemr.org

Refining BGP Efficiency : A Case Study

Reena Shakya
Department of Computer Science & Engineering
Madhav Prodyogiki Mahavidyalaya
Bhopal (M.P.) [INDIA]

Dr. Gireesh Kumar Dixit
Department of Computer Science & Engineering
Madhav Prodyogiki Mahavidyalaya
Bhopal (M.P.) [INDIA]
Email : gireeshdixit15@reiffmail.com

Abstract*—*A topology running the Border Gateway Protocol (BGP), the destination reachability can be interrupted due to connection failures and the time required for the network to converge could lead to quality of service degradation or even interruption, which is serious especially for real-time interactive applications. The path-exploration process and the time of the minimal route advertisement interval have a significant influence on the convergence time and in recent years several proposals have been made in order to reduce the convergence time. In this paper, we describe proposals to refine BGP efficiency. nodes.

1. INTRODUCTION

Currently the only protocol used on the Internet for the exchange of information between ISPs (Internet Services Provider) is known as BGP (Border Gateway Protocol). BGP is the external protocol that takes into account not only economic, social and technical relationships that are establish between ISPs, but also based on the attributes that characterize different existing connections between the terminal nodes, must determine what is the best way to be followed by packets of information exchanged between ASs (Autonomous System). To meet with this goal BGP defines characteristic elements such as routers types, databases, attributes, messages and timers that are essential for its proper operation.

BGP is considered a scalable protocol (handles a hierarchical structure that can be achieved through the use of attributes), robust and stable, which is characterized by avoiding loops. BGP is based on the path-vector routing scheme. An AS constructs a graph (tree) of the different systems that are connected taking into account the routing information learned from neighboring routers which can establish communicate according to their policies or criteria.

2. BGP CONVERGENCE PROBLEM

Convergence time is one of the most critical issues in a network due to the fact that as the network converges in the routing tables of routers could reside misinformation that causes errors when making routing decisions. The problem of convergence time on the Internet has been widely discussed in different studies [7], [8], [9], [10], [11] and concluded that there are mainly three factors that directly impact the convergence time in a BGP network: i) network topology, ii) the mechanism for selecting the best route (known as Path Exploration) and iii) the Minimum Route Advertisement Interval timer (MinRouteAdver). When the network topology changes, there is a delay equal to the time indicated by the timer MinRouteAdver between two updates sent by a BGP speaker related to the same destination prefix (i.e. destination IP address). This time is cumulative and, therefore, has a great impact over network convergence time. To illustrate the impact that

the MinRouteAdver timer has on convergence time, we present the following example, which shows the effect that an event such a node disconnection has on the routing tables for ease of explanation only, we consider a synchronous network, where at each round the router receives the messages sent in the previous round, calculates its new state, and sends new messages to its peers if required. The duration of each round is, for simplicity, 1s. Figure 1 shows a network of 5 nodes (Autonomous Systems or AS's) fully interconnected and, connected to the AS0, a network identified as dst (for purposes of example, the destination network). To understand the notation, the numbers inside the nodes are the AS's identification. The numbers beside each node indicate the last path announced by each network to dst.

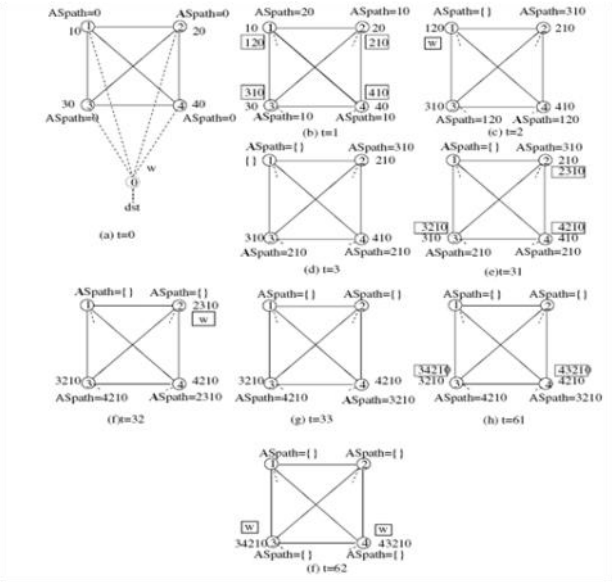

Figure 1. Illustration of convergence time problem in a Clique when an node disconnection event occurs [12]

At t = 0, all nodes can directly reach node 0, but then this node is dropped. In the next round all nodes choose a new route and should announce it. For this example, we analyze only the Node 1 which is illustrative for the other nodes.

At t = 1, Node 1 knows that the other nodes can have access to 0, so choose the best route. Since all paths are equal in length is chosen the one that has the lowest ID. Then, Node 1 will announce the new route 1-2-0. This is the first example of misinformation. Node 1 does not know that Node 2 currently has no access to Node 0 and makes assumptions with information that is no longer valid. Simultaneously, the other nodes do a similar process, but choose to Node 1 as the next hop to reach the node 0.

At t = 2, Node 1 learns that it cannot reach the Node 0 through Node 2 and immediately sends an update message to all nodes, indicating that they should withdraw (Withdrawn) this route.

AT t = 3, all nodes discovers that Node 1 has no access to Node 0 and choose a new route but cannot communicate it immediately since they must wait the time configured in the minRouteAdver to be able to send the new update. After 30 seconds (suggested time for minRouteAdver) the process is repeated again, Node 2 is chosen as the next hop and must wait another 30 seconds.

This example clearly shows that the combination of a highly connected topology (clique topology), the minRouteAdver timer and misinformation in the routing tables results in a relatively high convergence time, which can make unreliable some real time services currently offered on the Internet.

3. PROPOSED WORK

A. Ghost Flushing (GF) Improvement Proposal

The improvement proposal known as "Ghost Flushing" to correct the fact that a false information (or Ghost Information in the terminology used by the authors) generates another one and, in a network environment, such misinformation continues recursively. This means that if a node makes a decision based on information that is no longer valid, updates sent by this node may contain information that is also not valid. Subsequently, other nodes make their own decisions based on this misinformation, and so on. Sometime later, this false information will

be removed from the network, but require more messages, more updates and more time. If we also take into account the delay between updates due to MinRouteAdver timer it is easy to see that this may be the origin of high convergence time we are currently experiencing.

The procedure includes the following general activities:

To identify ghost information: information is classified as ghost information if the path to a destination network is upgraded to a longer path. If the path previously announced is no longer valid, there is false information in the network. first option is to immediately send an update message indicating the new path (this involves sending an implicit withdrawn). However, this action is not always feasible due to the limitation imposed by the MinRouteAdver timer; however it is always possible to send a withdrawn message that allows cleaning ghost information that has been published previously. It is important to note that a withdrawn message in this context has a different meaning than it has in the classic BGP4. In the original BGP protocol, a withdrawn means that a node has no routes to a destination and in the new proposal it means that the path that was previously announced and is no longer a valid path. This allows a router to send a withdrawn message despite of having alternative routes to a specific destination. minRouteAdver time did not elapse since the last announcement then send withdrawal (dst) to all neighboring BGP peers.

B. EPIC Improvement Proposal:

According to the analysis, the problem of BGP convergence time has its root on how this protocol performs the search for the optimal route to a given destination, what is known as "Path Exploration".

In vector based routing protocols, such as distance-vector and path-vector, the route selected by a router depends on the routes learned by its neighbors, which, in turn, depend on their neighbors and so on. This feature generates what is known as the process of scanning paths or path exploration which increases the convergence time of such protocols. It is important to notice that, in the case of vector-based protocols, the trajectory or path vector can prevent loops but does not avoid the exploration of paths.

Besides, considering that BGP works both externally (eBGP) and internally (iBGP) of the autonomous systems (AS's), the process of exploring trajectories becomes very complex and therefore take a long time. In BGP, the path exploration process is implemented using an attribute known as AS_PATH, which indicates the full path to be followed by a message originating at a given node to a specific node destination. This attribute is sent in different update messages and it indicates the preferred path to reach any destination node. Also, this attribute is used when a route is no longer valid and it should be eliminated, however, this attribute is not sufficient to distinguish whether other different paths are still valid or not. In BGP, there are only two ways to determine when a path s no longer valid:

When a router detects an event that changes the topology of the network (for example, a link failure, the termination of a BGP session, etc.). In this case, the router generates a message indicating that a path must be removed (if that is the case) and this router is called the "event originator."

When a router receives a message from a neighbor indicating that a particular path is no longer valid. In this case, the router generates a new message to its neighbors indicating that this route is no longer valid and including the AS_PATH that it has received without any change. This router is called "event propagator"

In practice, AS_PATH attributes sent by an "originator" or by a "propagator" are indistinguishable and contain no additional information to the path itself and, according to the proposal authors, this is the root of the path exploration problem, because when a router

receives a message with this attribute it cannot infer any additional information that allows it to identify additional invalid routes in its own routing table.

C. TIDR Improvement Proposal

"Traffic-aware Routing for Inter-Domain Routing Intenet Improved Stability for the design of this solution, the authors based on two important factors in Internet:

Internet access is not uniform: For quite some time has been observed that Internet traffic is not distributed evenly among the different autonomous systems [16], [17], [18], [19] and this observation has been consistent over time. For example, in studies of Rekhte and Chinoy in the late 80's, as in Fang and Peterson [18] in the late 90's and Rexford et al. [19] in 2002 showed that about 10% of Internet prefixes are responsible for more than 85% of network traffic. Additionally, it was observed that a large proportion of BGP update messages circulating in the network are caused by small percentage of highly active prefixes [20] and the most popular prefixes are very stable [19].

Prevalence of transient failures: It has been noted that a vast majority of link failures in Internet are transient and have a short duration. For example, a study of failures in the Sprint backbone links showed that about 50% of link failures were recovered in less than a minute, 80% in less than 10 minutes and 90% in less than 20 minutes [21]. Additionally, in [22] it is shown that about 50% of the problems caused by BGP misconfigurations last less than 10 minutes. In this solution, the authors consider two independent systems of any ASX and ASY. ASX autonomous system classifies all network prefixes into two categories: a class "significant" and a class prefix "no significant" prefixes, these classifications with respect to the autonomous system ASY.

In this solution, processing and dissemination of messages associated with each of these classes is handled differently, so that messages associated with significant prefixes are propagated with high priority and messages associated with the prefixes are not significant do it with a lower priority. Moreover, when the preferred route to a destination not significant fault and is replaced by another route on the ASX, the AS does not need to propagate the new route to its neighbor, if the fault is transient.

Thus, Tidra restricts the effect of transient faults and can reduce appreciably the sending of BGP update messages. On the other hand, any change in the routes associated with a significant prefix is always communicated to the neighbor (subject to MRAI timer) in order to ensure that the neighbor stays up to date information on the routes to their destinations significant. To achieve this, ASX has two types of timers associated with each of its neighbors, one of which is the standard MRAI BGP timer and the other is called the timer Tidra.

Tirdad timer is used only for the prefixes are not significant, while the timer MRAI prefixes used both significant as not significant. Ideally, Tidra timer duration should be long enough to allow most transient faults recover before the timer expires.

4. SIMULATIONS AND RESULTS

Simulation Tools

To evaluate the performance of each improvement proposal, we developed four simulators:

- Basic BGP4 Simulator for NS2
- BGP Ghost Flushing Simulator for NS2
- BGP EPIC Simulator for NS2
- BGP TIDR Simulator for NS2

To analyze the convergence times and the number of messages generated to achieve convergence in each proposal, we deployed 10 simulation campaigns on each generated

topology; faults were simulated in 30 randomly selected links and four different algorithms were compared: Original BGP4, BGP Ghost Flushing, BGP EPIC and BGP TIDR. In each of case the following were the measured variables:

Convergence Time:

This time is measured by the difference in time from the moment that the link failure is generated until no more update messages circulates through the network. Total messages: To measure this variable, we counted the number of update messages that are generated from the time the link failure occurs until the system converges.

Table 1. Average Convergence Time (secs)

Nodes	Original BGP4	BGP GF	BGP EPIC	BGP TIDR
50	387.00	387.00	260.00	4342.67
100	455.67	455.67	265.67	5636.00
150	499.33	469.33	277.33	7913.33
200	586.67	541.00	284.33	9100.33
250	598.00	547.00	285.00	9622.67
300	617.00	607.33	292.33	10194.00
350	623.33	633.67	314.33	10532.67
400	752.00	682.00	342.33	13152.33

5. CONCLUSIONS

Based on development work and the results presented in the previous chapter, we may get the following conclusions:

There is a high variability in the results of the variables analyzed in this study, with respect to the topology of the network over which measurements were made, even using the same routing algorithm. This shows that the variables analyzed in this study (convergence time and number of messages) have a high dependence on both the network topology and the location of the link failure.

Original BGP4 protocol presents the convergence times and the highest number of messages compared with the other algorithms tested. This allows us to conclude that this algorithm can be improved and, considering its impact on Internet traffic and stability, the search for solutions to its convergence problem is an important research field

REFERENCES:

[1] A. Tanenbaum, Computer networks, Boston, MA, USA: Prentice Hall, 2003, 869 p.

[2] A.M. Ospina Bolaños, Estado de Arte sobre el Border Gateway Protocol (BGP), Medellín, 2009, 107 p.

[3] J. MacFarlane, Network Routing Basics, Indianapolis, IN, USA: Wiley Publishing Inc., 2006, 395 p.

[4] B.A. Forouzan, Transmisión de datos y redes de comunicaciones, New York, NY, USA: McGraw Hill, 2007, 870 p.

[5] M.A. Sportack, IP routing fundamentals, San Jose, CA, USA: Cisco Press, 1999, 528 p.

[6] U. Black, IP routing protocols, Boston, MA, USA: Prentice Hall, 2000.

[7] Z. Jinjing; Z. Peidong, L. Xicheng, "On the Power-Law of the Internet and the Prediction of BGP Convergence," International Conference on Internet Surveillance and Protection, 2006. ICISP '06., vol., no., pp.17, 26-28 Aug. 2006

[8] C. Labovitz, A. Ahuja, R. Wattenhofer y S. Venkatachary, "The Impact Of Internet Policy And Topology On Delayed Routing Convergence," INFOCOM 2001. Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE, vol.1, no., pp.537 -546 vol.1, 2001.

[9] C.Labovitz, G.R. Malan, F. Jahanian,

"Origins Of Internet Routing Instability," INFOCOM '99. Eighteenth Annual Joint Conference -of the IEEE Computer and Communications Societies. Proceedings. IEEE, vol.1, no., pp.218-226 vol.1, 21-25 Mar 1999

[10] G. Huston, M. Rossi, G. Armitage, "A Technique for Reducing BGP Update Announcements through Path Exploration Damping", Selected Areas in Communications, IEEE Journal on, vol.28, no.8, pp.1271- 1286, October 2010

[11] R. Oliveira, Z. Beichuan; D. Pei; L. Zhang, "Quantifying Path Exploration in the Internet," Networking, IEEE/ ACM Transactions on, vol.17, no.2, pp.445-458, April 2009

[12] R. Zhang y M. Bartell, BGP Design and Implementation, San Jose, CA, USA: Cisco Press, 2003, 672 p.

[13] G. P. Rodrigo. Convergence Time Reduction in the BGP4 Routing Protocol using the "Ghost-Flushing" Technique and Other Proposals. 2004.

[14] Chandrashekar, J.; Duan, Z.; Zhang, Z.-L.; Krasky, J.;, "Limiting path exploration in BGP," INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE, vol.4, no., pp. 2337- 2348 vol. 4, 13-17 March 2005

[15] P. Chen; W. Hyung Cho; Z. Duan; X. Yuan;, "Traffic-Aware Inter-Domain Routing for Improved Internet Routing Stability," Proc. of Global Telecommunications Conference, 2008. IEEE GLOBECOM 2008. IEEE, vol., no., pp.1-6

[16] L. Kleinrock and W. Naylor, "On measured behavior of the ARPANetwork," Proc. of AFIPS Conference, 1974 National Computer Conference, May 1974.

[17] Y. Rekhter and B. Chinoy, "Injecting Inter-Autonomous System Routes Into Intra-Autonomous System Routing: A Performance Analysis", Proc. ACM SIGCOMM 1992 Computer Communication Review, vol. 22, no. 1, Ene. 1992.

[18] Fang, W.; Peterson, L., "Inter-AS traffic patterns and their implications", Proc. of Global Telecommunications Conference, 1999. GLOBECOM '99, vol.3, no., pp.1859-1868

[19] J.Rexford, J. Wang, Z. Xiao, y Y. Zhang, "BGP routing stability of popular destinations", Proc. of the 2nd ACM SIGCOMM Workshop on Internet measurement (IMW '02). ACM, New York, NY, USA, 197-202.

[20] Oliveira, R.V.; Izhak-Ratzin, R.; Beichuan Zhang; Lixia Zhang, "Measurement of highly active prefixes in BGP," Global Telecommunications Conference, 2005. GLOBECOM '05. IEEE, vol.2, no., pp. 5 pp.

[21] G. Iannaccone, C. Chuah, R. Mortier, S. Bhattacharyya, and C. Diot, "Analysis of link failures in an IP backbone," in Proc. of ACM SIGCOMM Internet Measurement Workshop, Nov. 2002.

[22] R. Mahajan, D. Wetherall, and T. Anderson, "Understanding BGP misconfiguration," in Proc. ACM SIGCOMM, Pittsburgh, PA, Aug. 2002.

[23] M. Piechowiak y P. Zwierzykowski, "Efficiency Analysis of Multicast Routing Algorithms in Large Networks," Networking and Services, 2007. ICNS. Third International Conference on, vol., no., pp.101, 19-25 June 2007 (2010) Waxman Network Topology Generator [en linea]. Disponible:-

[24] www.mathworks.com/matlabcentral/ fileexchange/2517-waxmannetwork.