Routing Instability in the BGP Protocol

Kwang-Il Lee, Mehdi Kalantari Khandani, Mark Shayman*

* Department of Electrical and Computer Engineering and Institute for Systems Research
University of Maryland, College Park, MD 20742
{kilee, mehkalan, shayman}@glue.umd.edu

Abstract **:** In this paper we will show some instability problems of the BGP routing protocol that can happen in the current Internet. We will show that the instabilities inside a domain in the IBGP protocol can be propagated to the outside of a domain and cause route instabilities in the external BGP domain. Our simulation experiments on simple network topologies conforms our results.

1. Introduction

Routing instability is one of the most important and pathological problems of the Internet. This kind of instability can cause loss of service, waste of network resources and service degradation for QoS demanding applications.

Border Gateway Protocol (BGP) is the most important and widely used inter-domain routing protocol, and so far it has been very successful in accommodating the fast growing demands of the Internet. The protocol has been successful in keeping the size of routing tables reasonable as the number of networks and hosts in the Internet increases very.

Like many other routing protocols, BGP suffers from route instability. The instability of routes in BGP routers can be very expensive because these instabilities incur very high costs by making major changes in traffic paths at high speed inter-domain links.

Many researchers have studied the stability and convergence properties of BGP routing protocol; authors of [8] have given the conditions under which route oscillations happen for the BGP routers in the Internet. They have given some constraints in the selection of those routes between the BGP routers to prevent oscillation. In order to prevent routing instabilities of BGP, the routing policies should be consistent, and they should not have any conflict, and the BGP convergence test can be performed by doing consistency test for the routing policies. The authors of [7] have shown that such tests are either NP-complete or NP-hard.

In [9], the authors have used the real data collected in the Internet to study the effect of BGP configuration errors on the end to end connectivity, and their experiments show that most of the times, the end to end connectivity is robust to the BGP configuration errors, but configuration can cause the waste of resources. In [10] the authors have used simulation and analysis to study the effect of loss of BGP messages due to congestion on the instability of BGP routes, and based on their results, they propose a prioritized treatment of BGP messaged by the routers. A survey on the problems of BGP can be found in [11].

In this paper we will study a special case of BGP instability that is triggered by congestion inside a routing domain. In this problem congestion inside a routing domain causes loss of IBGP connection between BGP routers of that routing domain. Loss of connection causes route flapping and instabilities in the routing tables.

There is a lot of research done on the effects of congestion on the behavior of BGP at the backbone routers, but less has been done about the intra-domain edge routers. However, the instabilities incurred by congestion among intra-domain edge routers can have similar effect. This kind of instability can be propagated to the other routers, and finally it may result a global instability. In this paper we will demonstrate this kind of instability.

The rest of this paper is organized as follows: In the Section 2 we will give a brief overview of BGP4. In Section 3 we give a few instances of the common instabilities of BGP in the Internet. In Section 4 we introduce a scenario in which congestion inside a domain and IGBP can cause instabilities of external BGP. In Section 5 the results of the simulation will be shown.

2. Overview of BGP4

The Internet is composed of many administratively independent domains called autonomous systems (AS). An AS is a set of routers and end users that are connected to each other and share a routing protocol such as RIP or OSPF to provide routing and connectivity of the hosts within that AS; thus an AS is sometimes called a routing

domain. The networks of a large organization or ISP are examples of autonomous systems.

To provide global connectivity to the hosts within an AS, we need to provide some kind of connectivity of that AS to the outside world. This connectivity is provided by the BGP routers at the boundaries of each AS. Each BGP router is responsible for different tasks like neighbor acquisition, neighbor reachability monitoring, and exchange of information about reachability of networks within or through that AS. Additionally BGP is equipped with a key property called Classless Inter-Domain Routing (CIDR). CIDR is a solution for more efficient use of IP addresses in the Internet, and keeping the size of routing tables small enough at BGP routers by using the concept of a supernet. In CIDR, several class C IP addresses are assigned to an organization instead of a huge unnecessary class B IP address. On the other hand, many of the networks with class C addresses with the same prefixes in their IP address may share the same entry in the routing table of a BGP router. This technique is called route aggregation.

The other important feature of BGP is use of path vector instead of distance vector. In path vector, the explicit paths to reach a destination are advertised to a neighbor BGP router instead of advertising a cost and next hop that is advertised in distance vector routing schemes. Path vector technique solves many problems of distance vector; with explicit routes, the loops are naturally avoided, and a network can define preferred paths instead of shortest paths that might have congestion or security problems.

BGP uses TCP as the reliable end to end transport protocol for making connection among BGP neighbors. This makes the protocol much simpler since many tasks like timer management, timeout, retransmission and transmission rate control are left to the TCP protocol.

There are four kinds of BGP messages that are exchanged through the TCP connection among the BGP peers:

1-OPEN
2-UPDATE
3-NOTIFICATION
4-KEEPALIVE

The OPEN message is used at the start for neighbor acquisition purpose. Once the connection is established without any error, the UPDATE message is used for exchanging paths among BGP peers. The NOTIFICATION message is for the purpose of handling the errors, and KEEPALIVE message is used for monitoring the connection and the link between the BGP peers.

The peers that exchange adjacency messages can belong to the same AS. This is for the case in which an AS has several BGP routers. The protocol under which these routers exchange the paths is similar to the case in which the BGP routers belong to different ASs. The messages are the same as the case that the routers belong to different ASs, and the underlying transport protocol is TCP. The only difference is that the BGP messages are exchanged among the routers by using the local routing protocol of the AS. BGP protocol requires all possible BGP pairs belonging to an AS establish an internal BGP connection and exchange the external paths they have learned through their connection to external peers. This situation is simply shown in Figure 1.

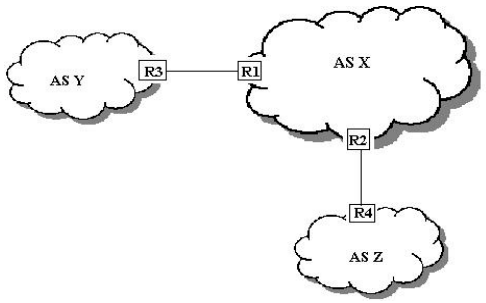

Figure 1: AS X has two BGP routers. The external paths learned by the two routers are exchanged through IBGP protocol.

In Figure 1, AS X has two BGP routers. Router R1 collects reachability information about the networks belonging to AS Y and the other networks that can be reached through AS Y. So does R2 about AS Z. Without exchange of the reachability information between R1 and R2 it will be impossible for AS Z to learn paths to reach AS Y. The protocol that is used for this internal BGP message exchange is called Internal BGP or IBGP. More information about BGP can be found in [1].

3. BGP instabilities

While BGP has satisfied many of the requirements of the very fast growing Internet, it has many drawbacks and vulnerabilities. Several problems like software bugs, TCP attacks or congestion can cause BGP instabilities in the routing tables of the BGP routers.

One of the classic problems of BGP is called the black-hole phenomenon. In this problem, a bug, hacker or manual bad configuration causes a BGP router to wrongly announce that through the AS to which it belongs, say AS X, there are good and low cost paths to some networks. This causes many BGP routers to update their routing table entries for these networks. So a huge amount of traffic will be forwarded to the AS X. This unexpected traffic causes

major problems like instability of routing tables, huge amount of packet loss and network resources, and finally a congestion collapse at the AS X and possibly its peers. In the literature of BGP instabilities, there are several well-documented instances of the black-hole phenomenon. More about black-hole and its occurrence can be found in [2], [3], and [4].

Another important vulnerability of BGP comes from the underlying transport protocol, which is TCP. TCP is vulnerable to several kinds of denial of service attacks, software bugs, and congestion in the underlying links. For instance a SYN flooding attack sends many TCP SYN messages toward a server and forces it to establish many connections and exhausts its resources such as memory and bandwidth. This kind of attack might be harmful for the other established and non-established TCP connections, and the connection intended for adjacent BGP peers may not be established. Or it might be lost at the time of attack if the bandwidth taken by the attack traffic causes heavy congestion.

Congestion can be pathological for BGP and causes instabilities in the routing tables. For instance if a huge amount of traffic is intended to be forwarded through a specific AS, it will be likely that the links that connect this AS to the external world are congested, and as a result it is likely that the periodic KEEPALIVE messages that are exchanged through an established TCP connection are lost, or if the congestion is heavy enough, the TCP connection used between the BGP peers times out. In either case the adjacency between two BGP routers will be lost.

In [5], the authors have shown how stressful conditions like congestion caused by the worm attacks Code-Red/Nimbda can affect the behavior of BGP. These attacks were aimed against ISPs in September 2001 and resulted in global instability of BGP in the Internet. The measurements of the authors show that the amount of BGP messages or BGP prefixes updated during the stressful conditions increased by more than a factor of ten over that during normal conditions. They have also shown it might take up to several days until the effects of such instabilities are damped and the routing tables of the BGP routers converge to a fairly stable state.

The above scenario is illustrated in the Figure 2. In this figure primarily the link between AS X and AS Y is congested due to a huge amount of traffic that is intended to be routed through this link. As a result of this major congestion in this link the TCP connection between R1 and R2 is lost, so for AS Y all networks that were formerly reachable through AS X will be announced unreachable. Soon afterwards, AS Z provides an alternative to reach these networks and the traffic that was formerly was routed through AS X now will be forwarded to AS Z to reach their destination networks. But again this traffic can cause congestion in the link between R1 and R3 and the same scenario happens for these two routers. On the other hand when the traffic is switched to AS Z, the link between R1 and R2 will be back to the normal status and BGP adjacency will be acquired between these two routers. Now the AS X announces good and low cost paths to AS Y for the networks that are now reachable through AS Z. Again the traffic will be switched back to AS X, and this problem might recur several times.

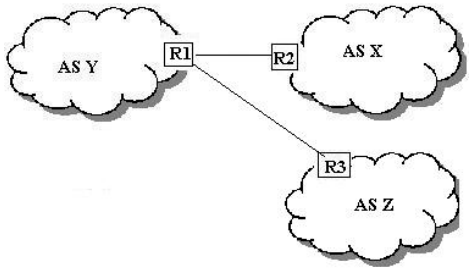

Figure 2: Congestion can cause instability of routing tables and several flapping of paths between AS X and AS Z

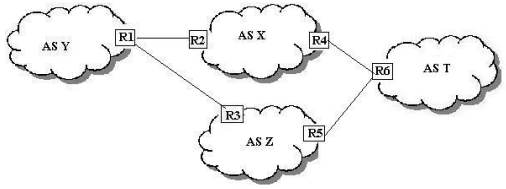

Figure 3: A scenario in which internal congestion in an AS causes BGP instability

Another possible instability happens when there is internal congestion in some links within an AS and this causes the TCP connection between two BGP routers of that AS to time out. In this case the two routers of the AS use IBGP protocol for the purpose of exchanging messages. Similar instability to the above case can happen in this case.

To illustrate the above scenario consider a topology like that in Figure 3. This scenario is similar to the previous scenario except for the fact that the bottleneck link is an internal link within AS X. Both IBGP messages and high bandwidth traffic are intended to pass through a bottleneck link in AS X. So this link becomes congested and the TCP connection used for IBGP message exchange between R2 and R4 is lost. Again the networks that formerly were reachable for AS Y through AS X become unreachable and alternative paths will be provided by AS Z, and the high bandwidth traffic is switched to AS Z. Then the IBGP connectivity of R2 and R4 is recovered and the routes will be switched back to AS X. This problem can recur several times and causes the routing tables of the routers in the AS Y to become unstable.

In this paper we focus on the IBGP scenario and try to show some incidents of that through simulation. In the next section we will describe the simulation scenario in more detail.

4. Simulation Scenario

In our experiments, the network is configured as shown in Figure 4. We configure 16 autonomous systems using SSFnet [6]. Each autonomous system has at least one BGP router, and each BGP router is connected by TCP connection with its peer BGP router.

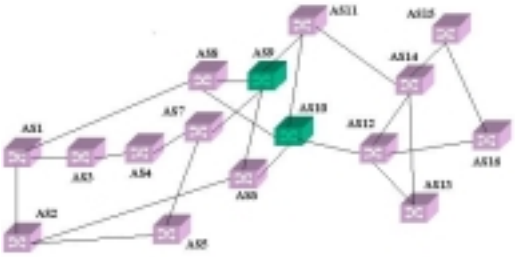

Figure 4: The used topology in AS level

In our experiments, we configure two autonomous systems (AS9 and AS10) to have multi-hop IBGP connections. The internal configuration of these two ASs is shown in Figure 5. OSPF is used for the intra-AS routing protocol. These two ASs are very important to transport traffic from/to the hosts in the ASs in the left hand side, (i.e., AS1 to AS8) to/from the hosts in the right hand side (i.e., AS11 to AS16). So, the congestion and IBGP failure of these two ASs could significantly influence the entire network routing information. In our experiments, we generate congestion in these two ASs, and analyze the BGP behaviors and routing instabilities of the entire network.

We run simulations with three congestion scenarios as follows:

Scenario 1 : Non-IBGP routers are congested in single AS
Scenario 2 : IBGP routers are congested in single AS.
Scenario 3 : Two ASs are congested at the same time.

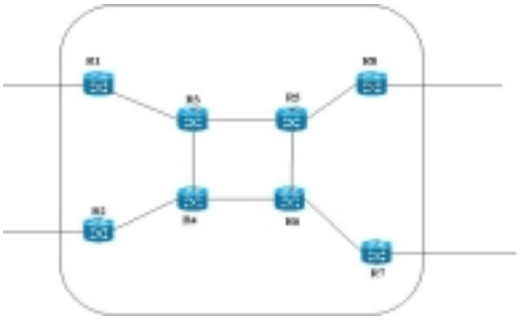

Figure 5: The topology of bottleneck ASs

The primary goal of our experiments is to measure routing instability resulting from IBGP failure caused by internal congestion. In our experiments, network instability is defined as follows: Routers at each AS first build a routing table under the condition that no congestion is present in the network. Then these routing tables are duplicated at each node and remain the same for whole simulation time. If the routing table at each AS is same as the duplicated routing table, then the network is considered to be in a "stable" state; otherwise it is considered to be in an "unstable" state. If the network is unstable, we determine which ASs exhibit the routing instability--i.e., have a routing table different from the duplicated table. For such a table, we determine how many routing table entries (subnet routes) are changed. This information is collected every 5 seconds. Data collection is continued for 500 seconds after the congestion is removed. This is to monitor the routing behavior as the network transitions from congestion conditions back to normal conditions.

In our experiments, we measured the following metrics so as to analyze the BGP routing instability: number of IBGP connections, number of unstable AS, number of unstable routing table entries, and the number of updated routing table information. The latter quantity is the total number of new routes to subnets carried by BGP update messages.

5. Analysis

First, we analyze two IBGP failure scenarios in a single AS. In Scenario 1, we analyze the IBGP failures caused by the congestion of non-IBGP routers. Since IBGP routers in some ASs are connected to each other by multi-hop TCP connections, the congestion of these nodes affects the IBGP connectivity. In our simulations, some IBGP connections are failed by congestion of transit nodes. As shown in Figure 6, up to four IBGP connections are disconnected. We measured how many ASs are affected by this IBGP failure. As shown in Figure 7, up to four ASs are affected by this failure. Although this number is smaller than those of other scenarios, about 30% of the entire network is affected by this failure. This implies that even if the IBGP routers are protected from DDOS attacks, it may not be sufficient to prevent network instability because the failure of transit routers can cause widespread routing instability to occur. Thus, all routers that transit BGP messages need to be protected.

In scenario 2, we monitor the BGP behaviors when IBGP routers are exposed to congestion. As shown in Figures 6 to 8, more ASs and subnets are in unstable status and the instability has longer duration than that of scenario 1. The instability is more severe in scenario 2 because the

congestion of an IBGP router causes failure of all IBGP connections for that router, while only those IBGP connections that pass through the congested transit router fail in scenario 1.

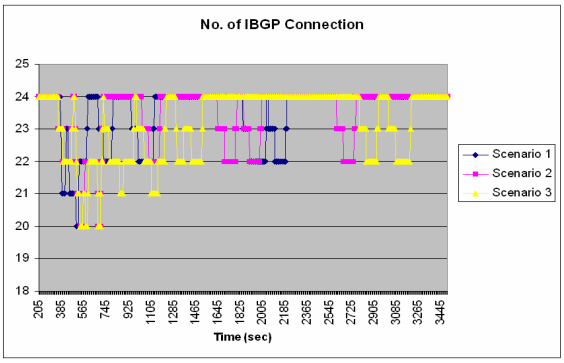

Figure 6. Number of IBGP Connections

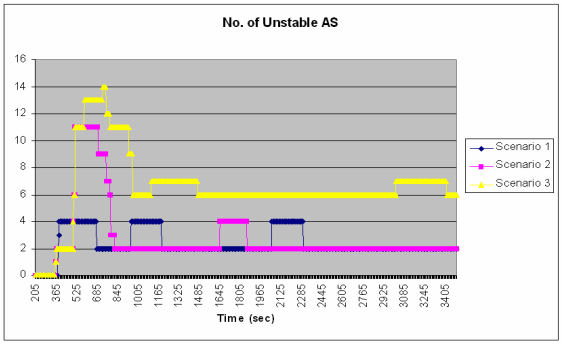

Figure 7. Number of Unstable AS

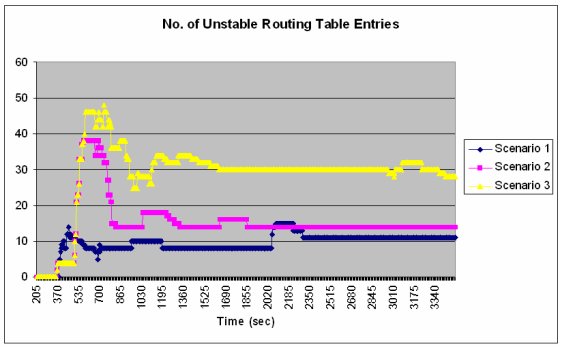

Figure 8. Number of Unstable Routing Table Entries

Finally, in scenario 3 we analyze a situation when multiple ASs are congested. For this, we generate extensive data traffic in AS 9 and AS 10 at the same time. As shown in Figures 6 to 9, we observed more fluctuations in all metrics, but the pattern of the graphs are very similar to those of Scenario 2. Even though two ASs are exposed to congestion simultaneously, the failure time of each AS is different. This results in longer routing table update time, larger routing updated information, and longer convergence time. In addition, all other ASs experience instability resulting from the congestion of both ASs. If congested autonomous systems serve as transit networks for many ASs, then the failure of the internal BGPs causes widespread routing instability.

One of our observations is that even when the congestion is removed, not all the routing entries return to their original values. When the IBGP connections are failed, then each AS recomputes its routing table with updated information. Later, the IBGP connections are recovered and information about the original routes is once again obtained. However, the routing information received from the newly recovered connections will not be reflected in the routing tables if it has the same cost as the existing routing table entries. This can be seen in Figure 8 where the number of unstable (i.e., changed) routing table entries does not return to 0 even after the congestion is removed.

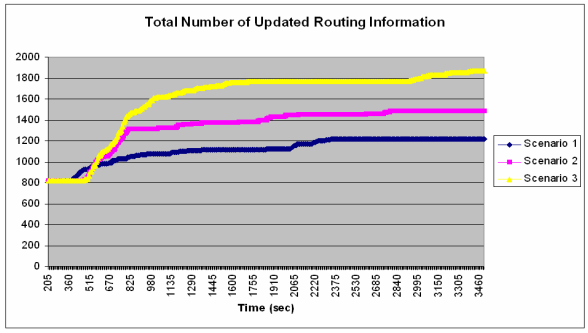

Figure 9. Total Routing Information exchanged by UPDATE message

6. Conclusion and the Future Study

In this paper, we analyzed BGP instability caused by internal congestion. From the simulation analysis, we have several conclusions.

First, even if the congested routers are non-IBGP routers, network instability can be initiated even though the impact of non-IBGP routers is less than that of IBGP routers. This implies that non-IBGP routers are also important for routing instability in multi-hop BGP connections. So, some protection mechanisms are required for non-IBGP routers in order to prevent their failure by DDOS attacks.

Second, IBGP failure caused by internal congestion can influence significantly the routing instability of entire networks. In our simulation, all ASs suffered from routing instabilities in some cases. The scope of this instability is mainly based on the location where the congestion happens.

When IBGP failures occur the first time, then many ASs are affected by this failure and compute another path to reach the destination. However, if congestion is removed or the same IBGP failures occur again, then it does not always cause the BGP routing instability since BGP does not replace the routing table entry unless the cost of the new route is less than that of the existing routing table entry.

In this paper, we analyzed IBGP failure caused by only internal congestion. The primary factor of routing instability is caused by the failure of EBGP connections. So, we need to analyze the effects of external BGP congestion. Also, when internal and external congestion occurs simultaneously, how the network responds and the impacts of it should be analyzed.

Also, the BGP failure in this paper is caused by link congestion. Node failure also causes BGP failure when enough network resources at each node are not available such as CPU time and memory space. The effects can be made more severe by traffic concentration when traffics are re-routed by other BGP failures. The impact of routing instability caused by node failure is left for future study.

References

[1] Christian Huitema, "Routing in the Internet," Second Edition, Prentice hall PTR, 1999.

[2] C. Labovitz, G. R. Malan, and F. Jahanian, "Internet Routing Instability," Proceedings of ACM/SIGCOMM 1997, computer Communication review, Vol 27, No 4, Oct. 1997.

[3] C. Labovitz, G. A. Ahuja, and F. Jahanian, "Experimental Study of Internet Stability and Wide-Area Backbone Failures," Proceedings of INFOCOM'99, March 1999.

[4] R. Barrett, S. Haar, R. Whitestone, "Routing Snafu Causes Internet Outage," Interactive Week, April 25 1997.

[5] L. Wang, X. Zhao, D Pei, R. Bush, D. Massey, A. Mankin, S. F. Wu, and L. Zhang, "Observation and Analysis of BGP Behavior under Stress," Proceedings of the second ACM SIGCOMM Internet Measurement Workshop, November 2002.

[6] SSFnet, http://www.ssfnet.org

[7] T. G. Griffin and G. Wilfing, "An Analysis of BGP Convergence Properties, " Proceedings of SIGCOMM, 1999.

[8] K. Varadhan, R. Govindan and D. Estrin, "Persistent Route Oscillations in Inter-Domain Routing, Computer Networks, March 2000.

[9] R.Mahajan, D.Wetherall, and T.Anderson, "Understanding BGP Misconfiguration," Proceedings of SIGCOMM 2002.

[10] A. Shaikh, L. Kalampoukas, R. Dube and A. Varma, "Routing Stability in Congested Networks:Experimentation and Analysis, " Proceedings of SIGCOMM 2000.

[11] W. Li, " Inter-Domain Routing: Problems and Solutions, , " Available online at: http://citeseer.nj.nec.com/li03interdomain.html