

A Variation of Route Flap Damping to Improve BGP Routing Convergence

Wang Lijun, Wu Jianping, Xu Ke

Department of Computer Science and Technology

Tsinghua University, Beijing, China, 100084

Email: wlj@csnet1.cs.tsinghua.edu.cn, jianping@cernet.edu.cn, xuke@tsinghua.edu.cn

Abstract—Inter-domain routing stability and convergence delay have significant effect on QoS in Internet. RFD is a mechanism to limit route oscillation from spreading wildly and is deemed as a key contributor for Internet routing stability. Recent research discovers that RFD may exacerbate relatively stable routes influenced by path exploration procedure and the interaction between RFD reuse timers. In this paper, a variation of RFD is proposed to deal with the side effect of RFD on routing convergence. Flapping routes are confined by neighboring nodes and invalid routes generated in path exploration are reduced by a RFD-like mechanism more suitable for their characteristics. Simulation results indicate that the modified flap damping mechanism limits persistent flapping routes while causing relatively stable routes converge more quickly.

I. INTRODUCTION

As a dynamic routing protocol, BGP[1] adapts to routing changes and converges to new stable routes. BGP Update messages reflecting route change will propagate to all the core routers in the Internet. BGP routing instability not only increases the processing overhead of border routers, but also impairs the quality of service (QoS) Internet provides to various applications. Network latency and packet loss rate increase during BGP route convergence[2]. Zhang *et al.* [3] found that the false uptime of unreachable destination and false downtime of reachable destination closely match the convergence delay after BGP route *Down* an *Up* event.

RFD[4] is a mechanism to limit the propagation of BGP route instability. A route flap means the changing of a route going down and shortly coming up again. Persistent flaps consume much processing power of routers, hence degrade their performance. RFD maintains a data structure for each route received, which includes *Penalty*, indicating stability degree of the route and used to predict future behavior of the route, and *Reuse timer*, indicating when suppressed route will be released. Each time a route changes, its *Penalty* increase a constant value. If $Penalty > P_{cutoff}$, where P_{cutoff} is a predefined threshold for the maximal tolerance extent, the route is suppressed. In stable state, *Penalty* decays exponentially.

RFD is widely considered as a contributor for Internet routing stability. However, Mao *et al.*[5] found that the convergence of relative stable routes may be exacerbated by RFD. Path exploration after a route *Down* event may induce RFD falsely suppress the route. False suppression makes route unavailable for a long time after the route restores stable.

Zhang *et al.*[6] found that interaction between RFD reuse timers also lengthen the convergence time.

RFD is more effective if used closer to the location of problem. In false suppression [5] and secondary suppression [6], RFD is used by nodes more than one hop away from the location, which causes RFD can not judge accurately just from BGP Update whether the route change reflects a real network change. Wrong judgement causes wrong increment of penalty. In this paper, we design a new flap limiting mechanism to damp persistent routes change while making relative stable route converge rapidly. *Suppression mark* is attached to a route to inform the route receiver whether the route is likely to change persistently. According to suppression mark, routes will undergo different damping mechanisms on routers. Persistent route change is damped by Neighboring Nodes Suppression on several neighboring ASes. Relatively further ASes are responsible for reducing invalid routes by using a new damping mechanism, Invalid Routes Damping, which is more suitable for invalid routes produced in path exploration. With the cooperation of ASes, while damping persistent route instability, the modified route flap damping mechanism reduces route convergence and communication overhead dramatically.

II. NEIGHBORING NODES SUPPRESSION

Route flap is damped more closer to the cause location, more significant effect can be achieved [4]. We first introduce Neighboring Nodes Suppression which suppresses route flap originating from source nodes and then extend it to handle route flap resulted from erroneous BGP session.

A. Intuition of Neighboring Nodes Suppression

We introduce the principle of Neighboring Nodes Suppression using the simple topology in Figure 1, in which node 1 to 9 represent ASes and d is a prefix belonging to node 1. The route to d is r_d . If the connection between d and node 1 flaps up and down, oscillating r_d will propagate to other nodes through neighboring nodes 2, 3, 5. In the first several flaps, on receiving Update message, node 3 will propagate route changes of r_d to node 4 immediately(limited by MRAI), the RFD penalty of r_d on node 4 increases almost synchronously with that of node 3. Route changes will further propagate to node 8, so the penalty of r_d on node 8 also increases. On receiving the fourth Withdrawal of r_d , node 3 suppresses r_d

and sends a Withdrawal to node 4. After that, route changes of r_d in succession can not reach node 4 and 8. No matter node 4 and 8 use RFD or not, the route changes of r_d on these two nodes are the same, *i.e.*, before node 3 suppresses r_d , r_d on node 4 and 8 change with that of node 3 and after node 3 suppresses r_d , r_d is unavailable. So, to limit the propagation of persistently oscillating route, if neighbors of the source node apply RFD to a route, node 4 and 8 are unnecessary to apply RFD to the route.

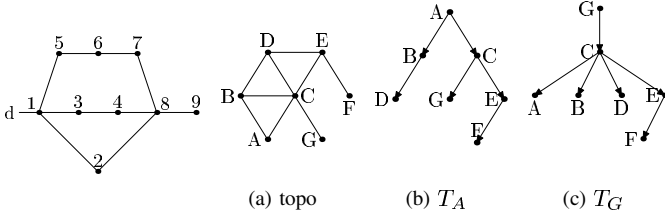


Fig. 1. A topology

Fig. 2. Route propagation tree

B. Route Propagation Tree

Each node selects route to destination node from the routes received from neighbors. The propagation and selection of a BGP route forms a tree rooted at the destination node:

Definition 1: Route Propagation Tree of route r_R , denoted as T_R , is formed as r_R is propagated and selected by nodes in network. The root of T_R is node R which originates r_R . If U selects the r_R from V as the best route, then node V is the parent of node U on T_R .

Figure 2(b) and 2(c) illustrate the route propagation trees of the network topology in Figure 2(a), rooted at node A and G. If route r_G flaps, the RFD mechanism on node C will confine the propagation of r_G and protect node A, B, D, E, and F from being effected by persistent route oscillation. If the nodes close to the root do not enable RFD, the route oscillation will be propagated to downstream nodes. For example, on route propagation tree T_A (Figure 2(b)), if RFD on node C is disabled, all the nodes selecting r_A from C, *i.e.*, E and G, will be affected by route flap of r_A . In this case, node G and E will take the responsibility to damp route flap of r_A . According to the function in limiting flapping route, the nodes on a route propagation tree can be divided into three sets:

$S_1(r_R)$:consists of the nodes closest to root R and RFD is disabled. Route flaps will propagate through the nodes in $S_1(r_R)$ to the nodes in $S_2(r_R)$.

$S_2(r_R)$:consists of the nodes which are direct children of $S_1(r_R) \cup \{R\}$ with RFD enabled. Persistently oscillating route will be limited by nodes in $S_2(r_R)$.

$S_3(r_R)$:consists of the children of the nodes in $S_2(r_R)$. Persistently oscillating route originating from R will not affect the nodes in $S_3(r_R)$.

Thus $S_2(r_R)$ forms a low-pass filter: if r_R is stable or relative stable, it can pass through $S_2(r_R)$; if the changing frequency of r_R exceeds some threshold, the propagation of r_R will be confined by $S_2(r_R)$. With the protection of $S_2(r_R)$, nodes in $S_3(r_R)$ are unnecessary to apply RFD to r_R .

T_R is a static tree matching a converged state of r_R and may be dynamic as r_R propagates, *i.e.*, the relative relation of nodes on T_R and the content of $S_1(r_R)$, $S_2(r_R)$ and $S_3(r_R)$ may change. However, route r_R transmitted on the static route propagation tree is the best route selected by the receiving node. The Update messages to withdraw the former best route in each flap and to announce the new best route in the next flap will be transmitted along the static route propagation tree. So, the nodes in $S_2(r_R)$ can guarantee the persistent oscillation of r_R will not affect the nodes in $S_3(r_R)$.

C. Suppression Mark

To limit route flap originating from the source node R, it is sufficient the neighboring nodes in $S_2(r_R)$ applying RFD to r_R . Our design is to attach an identifier in each BGP route, denoted as *Suppression Mark*, to inform the receiving nodes whether it is necessary to apply RFD to the route. If node V receives route r_R with suppression mark, V does not need to apply RFD to r_R . Otherwise V should apply RFD to r_R if RFD is enabled, *i.e.*, $V \in S_2(r_R)$ and insert suppression mark in r_R before propagating r_R to neighbors. To damp route flap originating from the source node, 1 bit is enough for suppression mark to transmit RFD suppression information. Secondary suppression problem is naturally avoided by discriminating routes with suppression mark.

D. Damping Flap from Link Failure

The interruption of BGP session may also result in route flaps, which can not solved by Neighboring Nodes Suppression. For example in Figure 3, RFD is enabled on node X, Y, so $S_2(r_S) = \{X, Y\}$. Suppose the BGP session between node X and W interrupts repeatedly for such reason as affected by link congestion or wrongly configured timer. Each time the BGP session is down, node W removes all the routes received from X and sends a Withdrawal or Announcement(if W receives r_S from Y) to Z. Moreover, in r_S that W receives from X, suppression mark(represented by *) is attached by X, hence W does not apply RFD to r_S from X and selects it as the best route. Each time the BGP session restores, node W announce r_S newly received from X to Z. Hence, r_S that Z receives from W flaps with the interrupting BGP session(as in Figure 3(a)), which may arose persistent route flap at Z. Therefore, if the erroneous BGP session is inside $S_2(r_S)$, flapping r_S can be limited by RFD mechanism of $S_2(r_S)$. If the location originating route flap is outside $S_2(r_S)$ as the example above, flapping r_S will propagate further.

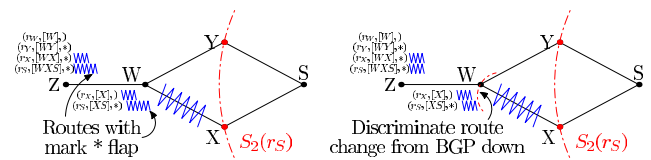


Fig. 3. Suppressing route flap resulted from BGP link failure

For the nodes in $S_3(r)$, route flap of r should be discriminated, *i.e.*, treat route change from Update and route removal from BGP session down differently. In the former case, if a route is attached with suppression mark, no damping mechanism is needed to apply to it. While in the latter case, all the routes should undergo RFD mechanism. In Figure 3(b), node W discriminates route change from BGP session down and apply RFD to route withdrawal no matter suppression mark attached or not. Persistent route flap is limited by the neighboring nodes adjacent to the erroneous links.

Nodes with damping mechanism disabled and with traditional RFD must be taken into consideration. For example, in Figure 3(b), if W is damping mechanism disabled, persistent flapping r_S will pass through W and reach Z . Because r_S is attached with suppression mark by X , according Neighboring Nodes Suppression Z does not apply RFD to r_S . To solve the problem resulted from partial deployment, we extend the content of suppression mark from 1 bit to 16 bits which records the ASN of the last node with Neighboring Nodes Suppression mechanism enabled. Before transmitting the selected route to neighbors, nodes attach its ASN in the route as an optional transitive route attribute. If the suppression mark in received route is not the ASN of the peer sending the route, it can be inferred that the peer is Neighboring Nodes Suppression disabled. All the routes from peers with Neighboring Nodes Suppression disabled should be applied to traditional RFD.

III. INVALID ROUTES DAMPING

Traditional RFD focuses on persistent route flapping. With the default RFD parameters of Cisco routers, if the average interval between successive flaps is less than 15 minutes, the route will be suppressed by RFD, because $P_W \times (1 + 1/2 + 1/4 + \dots) \rightarrow P_{cutoff}$. Research in [2] discovers that the volume of BGP routing Updates is several orders of magnitude more than expected and that the majority of Update is redundant. Neighboring Nodes Suppression can only damp persistent route oscillation resulted from real network changes, but it does not fit to damp invalid routes generated in path exploration.

A. Characteristics of Invalid Routes

Though traditional RFD also has damping effect on invalid routes in path exploration, but due to unfitness for the characteristic of the invalid routes, it can not damp invalid routes very well. For example, the interaction between RFD and path exploration results in false suppression of relative stable routes. The propagation of invalid routes in path exploration have following characteristics:

- 1) The interval between successive invalid routes is MRAI, the default is 30 seconds, which is much less than 15 minutes.
- 2) The maximum continuation time of path exploration is $MRAI * n$, where n is the largest path length in network.
- 3) The routes in path exploration usually changes route attribute, especially AS_PATH, with a Withdrawal

ending the procedure. While in route flap, Withdrawal/Announcement is alternately transmitted.

- 4) The LOCAL_PREF value of successive invalid routes increases monotonously. If nodes selecting route with shortest AS_PATH as the best route, the length of AS_PATH increases monotonously.

Invalid Routes Damping is a mechanism to curtail path exploration while fitting the characteristics of invalid routes. The combination of Neighboring Nodes Suppression and Invalid Routes Damping may reduce Update caused by route flap and path exploration. If a node receives a route attached with suppression mark, the node applies Invalid Routes Damping to it, otherwise, applies Neighboring Nodes Suppression.

B. Processing of Invalid Routes

Neighboring Nodes Suppression and suppression mark provide a method to identify route resulted from protocol behavior. If the suppression mark of a route is the ASN of sending peer, the receiving node will not be affected by persistent flap of the route. Such route changes may be treated differently from the route change resulted from real network change. The processing procedure of Invalid Routes Damping is similar with that of traditional RFD: maintaining a *Penalty* for each route free from RFD, increasing penalty value when the route changes, and suppressing the route if its penalty value exceeds a predefined threshold. However, Invalid Routes Damping has several new properties compared with traditional RFD:

- 1) The penalty increment of attribute change is greater than that of Withdrawal.
- 2) The penalty of routes which are not suppressed decays exponentially, with half life set to MRAI.
- 3) If a suppressed route keeps stable for $k * MRAI$, release the the suppression, where k is a configurable parameter.
- 4) if suppressed routes changes, just reset the reuse timer.

Traditional RFD employs penalty as a indicator of stability history of routes and judges future change basing on the penalty. More frequently a route changes, more great the penalty becomes and longer the route is suppressed. The penalty in Invalid Routes Damping is the indicator of whether path exploration happens, *i.e.*, if a route changes several times and the interval is about MRAI, then the node judges path exploration happens and suppresses the route to reduce invalid routes. The time when the route converges has no relation with how many changes the route have experienced. So, if a route keep stable for $k * MRAI$, the node can judge the path exploration is over and release the suppression.

The comprehensive algorithm to damping route flap combining Neighboring Nodes Suppression and Invalid Routes Damping is shown as Algorithm 1, where P'_{AC} , P'_W and P'_{cutoff} have the same meaning as P_{AC} , P_W and P_{cutoff} in traditional RFD but different values and k is a configurable parameter.

IV. SIMULATION

To validate the effectiveness of our design, we implement the modified route flap damping mechanism in SSFNet [7] and

Algorithm 1 Modified Route Flap Damping : MRFD(Route rt)

```

1: if (peer.Down or rt.supMark != peer.ASN) then
2:   RFD(rt);
3:   rt.supMark = localASN;
4: else
5:   dampInfo d = GetDampInfo(rt);
6:   if d.suppressed != TRUE then
7:     d.Decay(d.lastChgTime, now( ));
8:     if rt.type == Ann then
9:       d.penalty +=  $P'_{AC}$ ;
10:    else if rt.type == Wd then
11:      d.penalty +=  $P'_W$ ;
12:    end if
13:    if rt.penalty >  $P'_{cutoff}$  then
14:      d.Suppress( );
15:      d.SetReuseTimer( $k * MRAI$ );
16:    end if
17:  else
18:    d.ResetReuseTimer( );
19:  end if
20: end if

```

simulate route change on some network topologies.

A. Simulation method

To simplify simulation, each AS in topology consists of only one border router. The transmission delay between neighboring router is 0.01 second. All routers use default MRAI value and WRATE and SSLD are disabled. The parameters of traditional RFD used in Neighboring Nodes Suppression are configured to default values of Cisco router. A node R is randomly selected as source node and the convergence of route r_d is observed. In the initial stage, r_d is stable in all the nodes. Then the link between d and R begin to be Up/Down alternatively and R repeatedly withdraw/announce r_d to neighbors. The procedure that node R sends a Withdrawal and Announcement pair is a flap of r_d . The time interval between Withdrawal and the following Announcement is set to 150 seconds, and the interval between successive flaps is set to 50 seconds. After n flaps, r_d becomes stable again.

B. Synthetic Topology

We first use synthetic AS topologies generated by BRITE [8], [9]. The simulation topology is generated using Waxman probability model, with parameter $\alpha = 0.2$, $\beta = 0.15$. All the nodes enable damping mechanism. We compare the convergence delay of flapping route and the Update number of five different damping mechanisms: *RFD*, traditional RFD mechanism; *PunishLess*, nodes punish less for route attribute change, *i.e.*, $P_{AC} = 250$; *SRFD*, method introduced in [5]; *MRFD-*, modified RFD with only Neighboring Nodes Suppression; *MRFD*, modified RFD implementing Algorithm 1. The simulation results are shown in Figure 4.

The convergence time is obviously reduced by SRFD compared to PunishLess and traditional RFD, but false suppression is not totally avoided because when flap count is less than four times the convergence delay also exceeds 3000 seconds. The expense of SRFD reducing convergence delay is

communication overhead increasing dramatically. MRFD- is the mechanism with best convergence time, but MRFD- has no any prevention on path exploration, so its communication overhead is the greatest. MRFD adds invalid route suppression in MRFD- at the expense of a little more convergence time, still much better than traditional RFD, PunishLess and SRFD, while its communication overhead is the minimal.

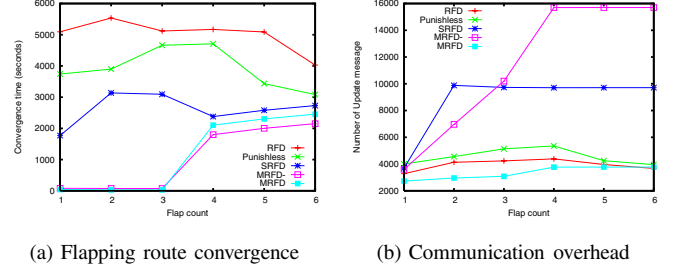


Fig. 4. Simulation result on 100-ASes Synthetic topology

C. Internet Derived Topology

It is difficult to evaluate the representativeness of the synthetic topology, since the fundamental properties of Internet are still open questions. So, we also use the Internet derived, 110 ASes topology [10] to evaluate the effectiveness of MRFD. The configuration of nodes in topology is the same as last simulation. As shown in Figure 5, the lines of convergence time are not as tidy as that of synthetic topology, but the MRFD- and MRFD converge almost immediately when route is relative stable (flap count less than three) and almost converge immediately after the Neighboring Nodes Suppression is released when route flaps (flap count more than four). The communication overhead of MRFD is a little more than that of RFD and PunishLess (as shown in Figure 5(b)) when flap count is more than four, but the difference is negligible and almost remain constant as flap count increases. The relative relationship of MRFD to other methods does not alter, *i.e.*, MRFD is obviously the optimal method with satisfying convergence time and communication overhead at the same time.

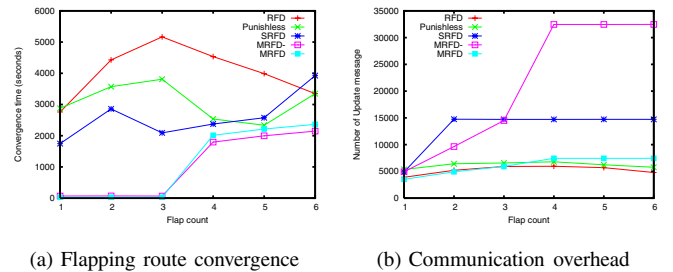


Fig. 5. Simulation result on 110-ASes Internet derived topology

D. Partial Deployment

We also research the effectiveness of various damping mechanisms with partial deployment. In the synthetic topology of Section IV-B, we randomly selected 5% nodes to be damping mechanism disabled. The selected nodes have 3, 4, 4, 9 neighbors respectively.

The percentage of nodes with route converged when route flaps 2 times and 4 times is shown in Figure 6. As shown in Figure 4, when route flaps 2 times, the route of all nodes converge almost immediately. But with 95% deployment of MRFD, route converges on about 25% nodes and the other nodes are affected by false suppression because the convergence delay is more than 1000 seconds. All routes received from nodes damping mechanism disabled will be applied to traditional RFD, the invalid routes transmitted before Invalid Routes Damping take effect bring on false suppression. The convergence delay of SRFD and that of MRFD are comparable, both superior to that of PunishLess and RFD.

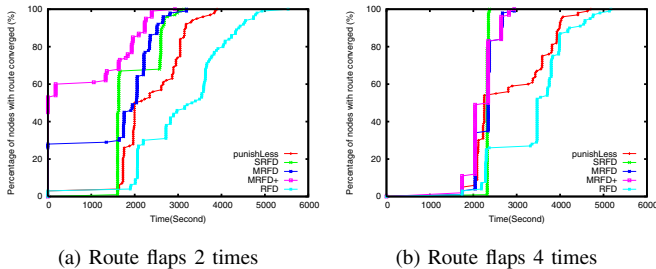


Fig. 6. Route convergence with partial deployment of various damping mechanism, 5% nodes are damping mechanism disabled

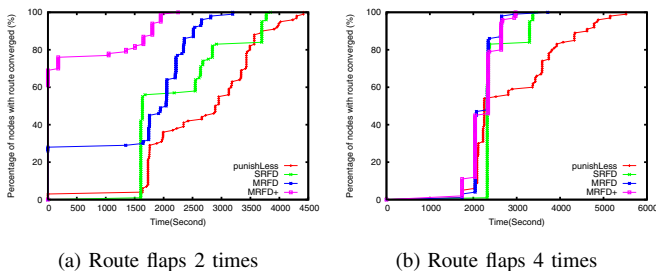


Fig. 7. Route convergence with partial deployment of various damping mechanism, 5% nodes use traditional RFD only

If node V is a direct son of some selected node U and the route V receiving from U is falsely suppressed, V will not reach stable state until the false suppression times out. All the children of V on the route propagation tree will reach stable state immediately after V does. This is exhibited by the step-like shape of the MRFD line. If the U prefer to select routes from neighbor with MRFD enabled, for these routes are more likely to be stable, the route of U and its children will converge more quickly. We use MRFD+ to denote MRFD combining the routing policy that node prefers a route from MRFD-enabled nodes than that from MRFD-disabled nodes. About 50% nodes converge immediately (as shown in Figure 6(a)), which is obviously superior than MRFD and SRFD.

The ideal goal of flapping route suppression is that after the route keeping stable for some time, the route converges as soon as possible on all nodes. As shown in Figure 6(b), after the flapping route is released, SRFD, MRFD, and MRFD+ have comparable performance, *i.e.*, the route converges on all nodes in a short interval. The simulation result of partial deployment with 5% RFD-enabled nodes is shown in Figure 7. Though false suppression can not be avoided totally, MRFD+

has relatively satisfying effect, both for relative stable routes and flapping routes. The simulation result also proves that with proper routing policy MRFD may improve routing convergence performance greatly.

V. CONCLUSION AND FUTURE WORK

In this paper we modify traditional RFD to make it function more accurately to suppress persistently oscillating routes and reduce invalid route in path exploration. Using Neighboring Nodes Suppression and Invalid Routes Damping jointly, the new damping mechanism can achieve satisfied convergence time and communication overhead simultaneously.

Policy conflict [11] among ASes may cause BGP route diverge. Traditional RFD has some mitigation effect on this type divergence. Though the damping mechanism brought forth in this paper can not cure routing divergence, some protocol design methods and routing configuration guidance [12], [13] have been brought out and proven to be effective in reality.

Each AS has a consistent routing policy and RFD is not applied to routes from IBGP, so we abstract each AS as a node and ignore the internal details. IBGP session usually is established using *loopback* interface, which improves the reliability of IBGP. Enhancing the design to solve instability arising from IBGP is our future work.

ACKNOWLEDGMENT

This work is supported by National Key Fundamental Research Plan (973) of China (No.2003CB314801).

REFERENCES

- [1] Y. Rekhter and T. Li, "A Border Gateway Protocol 4(BGP-4)," *RFC 1771*, March 1995.
- [2] C. Labovitz, G. R. Malan, and F. Jahanian, "Internet routing instability," *IEEE/ACM Transactions on Networking*, vol. 6, no. 5, pp. 515 – 527, 1998.
- [3] B. Zhang, D. Massey, and L. Zhang, "Destination reachability and BGP convergence time," in *Proceedings of IEEE Global Telecommunications Conference*, vol. 3, November 2004, pp. 1383 – 1389.
- [4] C. Villamizar, R. Chandra, and R. Govindan, "BGP route flap dampening," *RFC 2439*, November 1998.
- [5] Z. M. Mao, R. Govindan, G. Varghese, and R. H. Katz, "Route flap damping exacerbates Internet routing convergence," in *Proceedings of ACM SIGCOMM*, vol. 32, 2002, pp. 221 – 233.
- [6] B. Zhang, D. Pei, D. Massey, and L. Zhang, "Timer interaction in route flap damping," in *Proceedings of 25th International Conference on Distributed Computing Systems(ICDCS)*, 2005, pp. 393 – 403.
- [7] SSF Research Network, <http://www.ssfnet.org/>.
- [8] BRITE: Boston university Representative Internet Topology generator, <http://www.cs.bu.edu/brite/>.
- [9] A. Medina, A. Lakhina, I. Matta, and J. Byers, "BRITE: An approach to universal topology generation," in *Proceedings of IEEE International Workshop on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems(MASCOT)*, 2001, pp. 346 – 353.
- [10] B. Premore, Multi-as topologies with BGP routing tables, <http://www.ssfnet.org/Exchange/gallery/asgraph/index.html>.
- [11] T. G. Griffin, F. B. Shepherd, and G. Wilfong, "The stable paths problem and interdomain routing," *IEEE/ACM Transactions on Networking*, vol. 10, no. 2, pp. 232 – 243, 2002.
- [12] L. Gao and J. Rexford, "Stable Internet routing without global coordination," *IEEE/ACM Transactions on Networking*, vol. 9, no. 6, pp. 681 – 692, 2001.
- [13] T. G. Griffin and G. Wilfong, "On the correctness of IBGP configuration," in *Proceedings of ACM SIGCOMM*, 2002, pp. 17 – 29.