Available online at www.sciencedirect.com
ScienceDirect
Computer Communications 30 (2007) 1588–1599
computer communications
www.elsevier.com/locate/comcom

Modified flap damping mechanism to improve inter-domain routing convergence

Wang Lijun *, Wu Jianping, Xu Ke

Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China

Received 13 August 2006; received in revised form 17 January 2007; accepted 20 January 2007
Available online 30 January 2007

Abstract

Inter-domain routing stability and convergence delay have a significant effect on QoS in the Internet. To enhance the stability and reduce the convergence time of the Internet, RFD was introduced to limit route oscillations from spreading throughout the network. However, recent research has shown that RFD may damp relatively stable routes due to its reaction to path exploration and the interaction between RFD reuse timers. In this paper, a variation of RFD is proposed that addresses the negative side effects of RFD with respect to route convergence and stability. Route flapping is confined by neighboring nodes, and invalid routes generated by path exploration are reduced by an RFD-like mechanism. Simulation results indicate that the modified flap damping mechanism limits persistent flapping routes while allowing relatively stable routes to converge more quickly.

Keywords: Inter-domain routing; Border gateway protocol; Route flapping; Routing convergence

1. Introduction

As a dynamic routing protocol, when the link topology of the Internet changes, the Border Gateway Protocol (BGP) [1] adapts to these changes and converges to a new stable routing topology. These changes, which are communicated through BGP Update messages, propagate to all the routers in the core of the Internet. Because of this global effect, BGP routing instability not only increases the processing overhead of all border routers, but also impairs the quality of service (QoS) provided by the Internet. It has been shown that network latency and packet loss rates increase during BGP route convergence [2]. Zhang et al. [3] found that the false uptime of unreachable destinations and false downtime of reachable destinations closely matched the convergence delay seen after BGP route *Down* and *Up* events.

Route Flap Damping (RFD) [4] is a mechanism that limits the propagation of unstable routes. These routes are repeatedly withdrawn and announced on a short time scale, which is referred to as "route flapping". Since its proposal in 1993, RFD has been widely implemented in commercial routers and considered to be a key contributor to Internet routing stability. However, Mao et al. [5] found that the convergence time of relatively stable routes was negatively impacted by RFD. After a route flaps once, many invalid routes are generated in the path exploration and this induces RFD to falsely suppress the route. This false suppression results in the route being unavailable long after the route has become stable again. In addition, Zhang et al. [6] found that the interaction between RFD reuse timers of different routers can also significantly lengthen the convergence time.

Designed to deal with persistent route flapping, RFD is more effective if used closer to the source of the problem. If used by a router more than one hop away from the problem source, RFD cannot accurately determine whether the Update message in fact reflects a topological change.

* Corresponding author. Tel.: +86 10 62785822.
E-mail address: wlj@csnet1.cs.tsinghua.edu.cn (W. Lijun).

0140-3664/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.comcom.2007.01.012

This results in false suppression [5] and secondary suppression [6]. The key reason for the negative effect of RFD on BGP routing convergence is that the misinterpretation of Update messages results in an inappropriate penalty being applied to routes.

In this paper, we propose a variation of RFD that suppresses route flap while improving BGP routing convergence. To achieve this, a *Suppression Mark* is attached to BGP Update messages that allows receiving routers to assess the stability of the advertised route. According to this Suppression Mark, route changes will be subject to different damping mechanisms. With this Suppression Mark, we introduce two damping mechanisms, Neighboring Nodes Suppression that damps persistent route flaps at the source, and Invalid Routes Damping that is more suitable for damping invalid routes produced during path exploration. With the cooperation of ASes, these techniques not only damp persistent route flapping, but also reduce routing convergence and communication overhead.

The rest of the paper is organized as follows. In Section 2, we give an overview of BGP routing and RFD-related convergence issues. Related works are introduced in Section 3. In Section 4, we introduce the design principles and goals of the new damping mechanisms. In the following sections, we present the details of the modified route flap damping mechanisms, including Neighboring Nodes Suppression in (Section 5), and Invalid Routes Damping in (Section 6). Simulation results in Section 7 demonstrate the effectiveness of our design. At last, we conclude the paper in Section 8.

2. Background

The Internet consists of more than 24,000 networks, referred to as Autonomous Systems (ASes), each of which is managed by a different authority and identified by a unique 16-bit AS number (ASN). BGP is the *de facto* inter-domain routing standard that enables communication between ASes and across the internet. AS border routers establish BGP sessions with neighboring ASes to propagate network reachability information, and this information is disseminated across the Internet from AS to AS. BGP is a policy-based routing protocol, and the border routers of an AS have an unified routing policy. Consequently, the entire AS is viewed as a single entity by the outside world. BGP sessions between routers inside a single AS are referred to as Internal BGP (IBGP) sessions, whereas BGP connections between routers in separate ASes are referred to as External BGP (EBGP) sessions. As an AS is viewed as a single entity by the outside world, in this paper we refer to an AS as a node and the term BGP session is taken to mean an EBGP session.

BGP *Update* messages, which are used to transmit routing information, can be of two types; *Announcement* and *Withdrawal*. Route *Announcements* consist of Network Layer Reachability Information (NLRI), which is a single IP prefix, and route attributes. The AS_PATH attribute records the ASNs, in order of traversal, which comprise the route and this complete path is maintained to enable avoidance of routing loops. BGP routes received from each neighbor are stored in the corresponding *Loc-Rib-In*, one of which is maintained for each neighbor. BGP then selects the best route to each prefix from candidates found in all of the *Loc-Rib-In* data structures, and this route is installed in *Loc-Rib*. This selected route will be transmitted to neighbors and recorded in the neighbor-specific *Loc-Rib-Out*. The sending rate of *Announcement* messages is limited by the Minimum Route Advertisement Interval (MRAI), and independent timers are maintained for each prefix. If a node no longer has a route to a prefix, border routers will send *Withdrawal* messages to the neighboring ASes. In this case, MRAI is not applied in order to prevent the "Black hole" effect.

Router software or hardware misconfiguration and malfunction can result in the rapid oscillation of a route between the up and down state, which is referred to as route flapping. Persistent route flapping can consume considerable processing power at border routers, hence degrading their performance. In addition, unchecked route flapping can result in the instability of the Internet routing topology. RFD limits the propagation of flapping routes in order to reduce processing overhead and improve the stability of routing tables.

With RFD, a data structure is maintained for each route received that includes *Penalty*, which indicates the stability of the route and can be used to predict the future behavior of the route, and *Reuse timer*, which indicates when a suppressed route will be released. Each time the state of a route changes, the *Penalty* of the route increases by a constant value. If $Penalty > P_{\text{cutoff}}$, where P_{cutoff} is a predefined threshold, the route is suppressed. When the route stabilizes, *Penalty* decays exponentially. That is, if *Penalty* is $p(t_0)$ at t_0 and become $p(t)$ at t, then $p(t) = p(t_0)e^{-\lambda(t-t_0)}$, $\lambda = \ln 2/H$, where H is the half life of the decay. When *Penalty* decays to below P_{reuse}, which is another predefined threshold, the route is released and reinstated as a candidate for the BGP decision process. Persistent flapping causes *Penalty* to increase rapidly. Based on the default RFD parameters of Cisco routers (Table 1), the change of *Penalty* as a function of time is illustrated in Fig. 1.

Table 1
Default RFD parameters of commercial routers

RFD parameters	Cisco	Juniper
Withdrawal penalty (P_{W})	1000	1000
Re-announcement penalty (P_{RA})	0	1000
Attribute change penalty (P_{AC})	500	500
P_{cutoff}	2000	3000
P_{reuse}	750	750
Half life (H) (min)	15	15
Max hold-down time (min)	60	60

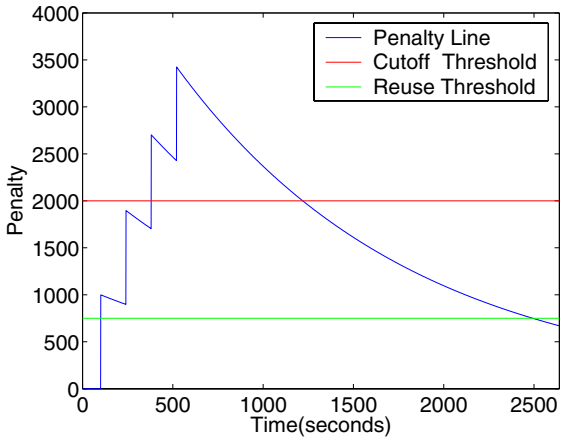


Fig. 1. Changing of damping penalty value.

3. Related work

As BGP is a Path-Vector protocol, when a route is withdrawn a path exploration procedure is triggered. Because information about routes are propagated, and there is no information about links, when a link fails, all routes using this link are not simultaneously withdrawn. Instead, the router withdraws the current route, selects the next best route, and propagates this route to its neighbors. This route, however, may also make use of the failed link and so it is subsequently withdrawn. This process continues until a valid route is found, and this sequence of withdrawals propagates through the entire network. From experiments performed on the Internet, Labovitz et al. [7] found that T_{down} and T_{long} events resulted in average convergence time of 3 min, with the maximal convergence time longer than 15 min. Theoretical analysis indicates that in the worst case, O(n!) routes will be explored, where n is the number of nodes in the network. With the packing effect of MRAI, the convergence time after T_{down} is $t_{MRAI} * N$, where N is the maximal length of AS paths.

Mao et al. [5] first found that invalid routes generated by path exploration may cause RFD to falsely suppress relatively stable routes. One route flap may be amplified by path exploration, and after receiving multiple invalid routes nodes will suppress the route and may have no available route to the destination until *reuse timer* times out. This false suppression delays BGP route convergence. According to the default RFD parameters in Table 1, the convergence delay resulting from false suppression may exceed 20 min.

In [8] RIPE recommends that a route should not be suppressed until it flaps more than four times. Zhang et al. [6] found that when a route flaps more than four times, the convergence behavior is consistent with the design goal of RFD. However, when a route flaps less than four times, the convergence time greatly exceeds expectations. Zhang et al. also found that, in a two-dimensional grid topology, false suppression accounts for 30% additional convergence time. Another reason for delayed route convergence is the interaction between reuse timers of different routers. Update messages resulting from suppression release may increase the penalty on other routers, which generates new suppression.

Mao et al. [5] proposed Selective Route Flap Damping (SRFD) to solve the problems caused by the interaction between path exploration and RFD. Mao et al. found that the LOCAL_PREF values of the routes received from a neighbor after T_{down} decrease monotonically. Based on this observation, SRFD attaches *Comparison bits* in route Announcements, which represents the relative preference value compared to the previous route Announcement. Invalid routes can be detected by comparing with the *Comparison bits* of the received route and the previously received route. If the LOCAL_PREF values of the received routes change from decreasing to increasing, the node can confirm a flap and increase the penalty of the route; otherwise RFD ignores the route change. The blemish of SRFD is that it can not handle Withdrawals appearing in the route sequence timely until the following Announcement arrives when the changing direction of LOCAL_PREF attribute can not be obtained. Duan et al. proposed RFD+ [19] to improve the performance, which inserts a Relative Preference Community Attribute in the Update message. These two methods both need to expose local routing policies to neighbors, while many ASes may be unwilling to do so.

Zhang et al. [6,9] proposed the attachment of a Root Cause Notification (RCN) to each route that triggers an Update. This RCN includes route cause link, link status, and sequence number. Nodes associate with each Update its originating location and event, and differentiate successive identical events by the sequence number. The damping penalty is increased only for Updates caused by route flaps and multiple Updates with the same RCN increase the damping penalty only once. However, RCN requires that each node is capable of detecting the status change of its adjacent links. If the node originating the Update does not support the RCN extension, route convergence will still be poor.

4. Design overview

RFD not only is a mechanism that benefits the deploying nodes, but also contributes to the overall performance of the network. With the current design, a node judges route flap using only local route changes, which is not accurate enough for complicated network topology and protocol behavior. Based on traditional RFD, we design a new flap damping mechanism to damp persistent route changes while assuring that relatively stable route converge rapidly. We classify route changes into two categories: those reflecting real network changes due to software or hardware malfunction, erroneous configuration, link congestion, etc., and those resulting from protocol behavior. Traditional RFD is designed mainly to address the former, while the latter exacerbates the problem, i.e. false suppression and

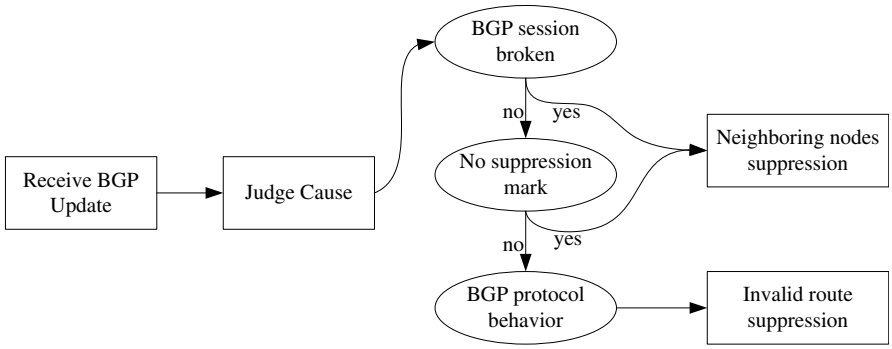


Fig. 2. Architecture of modified damping mechanism.

secondary suppression. In addition, RFD conflicts with itself in some situations, for example, secondary suppression. The design objectives include:

(1) Flapping routes should undergo suppression for a punitive duration, and converge immediately after the suppression is released, i.e. no secondary suppression;
(2) Relatively stable routes that oscillate infrequently or only a few times should converge immediately and, are exempt from path exploration, i.e. no false suppression;
(3) Communication overhead is reduced;
(4) The mechanism is compatible with current RFD mechanisms and lends itself to incremental deployment.

In designing a damping mechanism, we were guided by the following principles: (I) Utilizing cooperation among nodes to dampen route flapping and reduce processing and communication overhead for all nodes; (II) Processing route changes differently according to their different causes. The architecture of the design is illustrated in Fig. 2, where some new designs will be detailed in following sections.

5. Neighboring nodes suppression

The closer to the source a route flap is dampened, the more significant the benefits will be [4]. With this in mind, we first introduce Neighboring Nodes Suppression which suppresses route flap near the source where the route is originated. Then, we extend Neighboring Nodes Suppression to handle route flap resulting from erroneous BGP sessions.

5.1. Introduction of neighboring nodes suppression

We introduce the principle of Neighboring Nodes Suppression using the simple topology in Fig. 3, in which node 1 through 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. For the

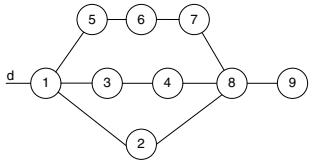


Fig. 3. A simple topology.

first several flaps, on receiving Update messages node 3 will propagate the route changes of r_d to node 4 immediately (limited by MRAI), the RFD penalty of r_d on node 4 increases in step with that of node 3. Route changes will also propagate to node 8, so the penalty of r_d on node 8 also increases. If node 3 uses traditional RFD with default parameters as in Table 1, on receiving the fourth Withdrawal of r_d, node 3 suppresses r_d (shown in Fig. 1) and sends a Withdrawal to node 4. After this happens, route changes of r_d will not reach node 4 and 8. Regardless of node 4 and 8 using RFD, the route state of r_d will remain unchanged even if there are additional flaps, because after node 3 suppresses r_d, r_d is unavailable. So, to limit the propagation of a persistently oscillating route, if the neighbors of the source node apply RFD to the route, it is unnecessary for more distant nodes to do so.

5.2. Route propagation tree

Each node selects a route to a destination node from all routes received from its neighbors. The propagation and selection of a route forms a tree rooted at the destination node, which has the following definition.

Definition 1. *Route Propagation Tree* of route r_R, denoted as T_R, corresponds to a stable state of r_R in the 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.

If node U has route to R, then U must be on tree T_R and its parent is the neighbor who sends the route to U. T_R is formed as r_R is propagated and selected by the nodes in the network. If a node change is selection, the tree is changed. For example, if U selects the route from V as the best

route to R, then on T_R, V is the parent of U. If afterwards, U selects the route from another neighbor, say W, then on T_R the parent of U changes to W.

Fig. 4(b) and (c) illustrate the route propagation trees of a simple network topology (Fig. 4(a)) rooted at node A and G, respectively. 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. Similarly, if route r_A flaps, node B and C protect node D, G, E, and F from being effected. 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 (Fig. 4(b)), if RFD on node C is disabled, all the nodes selecting r_A from C, i.e. E and G, will be effected by the route flap of r_A. In this case, node G and E will be responsible for damping the route flap of r_A.

The oscillation of r_R propagates from the nodes closer to R to the nodes further from R. So, if the nodes closer to R suppress the flapping route, the nodes further from R will not be effected and it is unnecessary for them to apply RFD on r_R. Accordingly, 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 routes will be dampened 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 routes originating from R will not effect the nodes in $S_3(r_R)$.

Thus $S_2(r_R)$ forms a low-pass filter: if r_R is stable or relatively stable, it can pass through $S_2(r_R)$; if the frequency of change of r_R exceeds some threshold, the propagation of r_R will be confined by $S_2(r_R)$. The nodes in $S_2(r_R)$ not only protect themselves from the large processing overhead associated with flapping routes, also protect the nodes in $S_3(r_R)$ which can forego the application of RFD.

The static tree T_R represents a converged state of r_R and it may be dynamic as r_R propagates. This is also true for the nodes on T_R and the content of $S_1(r_R)$, $S_2(r_R)$ and $S_3(r_R)$. In the convergence procedure of r_R, many transient route propagation trees may be formed. However, route r_R transmitted on the static route propagation tree is the best route selected by the receiving node. The Update messages

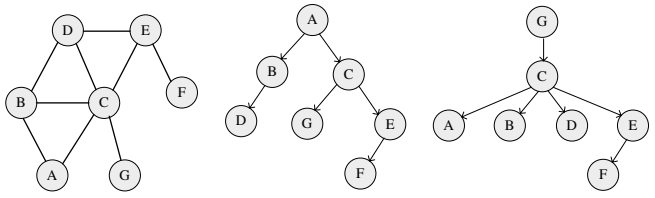


Fig. 4. Route propagation tree examples: left, network topology; center, tree rooted at node A; right, tree rooted at node G.

to withdraw the former best route in each flap and to announce the 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 $S_3(r_R)$.

5.3. Suppression mark

Routers maintain a data structure for RFD for each received route [10] and each time a route changes, its RFD parameters are recalculated. To limit route flap originating from a source node R, it is sufficient that the neighboring nodes in $S_2(r_R)$ apply 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 an attached 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. $V \in S_2(r_R)$. In addition, before propagating r_R to neighbors, V should insert Suppression Mark in r_R. One bit is enough for attaching the *Suppression Mark*. With *Suppression Mark*, secondary suppression may be avoided naturally: after a neighboring node suppressing a route, the downstream nodes do not apply damping mechanism to the route, so secondary suppression will never happen.

5.4. Damping flap from link failure

The interruption of BGP sessions may also result in route flaps, and Neighboring Nodes Suppression is not sufficient for dampening such flaps. For example in Fig. 5, RFD is enabled on node X, Y, so $S_2(r_S) = \{X, Y\}$. Suppose the BGP session between node X and W is interrupted repeatedly due to link congestion or an incorrectly 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 also received r_S from Y) to Z. Moreover, in r_S that W receives from X, a *Suppression Mark* (represented by $*$) is attached by X, hence W does not apply RFD to r_S received from X and selects it as the best route. Each time the BGP session restores, node W announces r_S newly received from X to Z. Hence, r_S that Z receives from W flaps with the interrupted BGP session(as in Fig. 5(a)), which may result in persistent route flap at Z. Therefore, if the erroneous BGP session is inside $S_2(r_S)$, flapping r_S can be limited by the RFD mechanism of $S_2(r_S)$. If the source of the route flap is outside $S_2(r_S)$, as in the example above, flapping r_S will propagate further.

To protect the nodes in $S_3(r)$, route flapping of r should be discriminated according to cause, i.e. routing changes due to Updates and those due to erroneous BGP sessions should be treated differently. In the former case, if a route has an attached *Suppression Mark*, no damping mechanism should be applied. While in the latter case, damping should be applied to all routes. In Fig. 5(b), node W identifies route changes caused by erroneous BGP sessions and

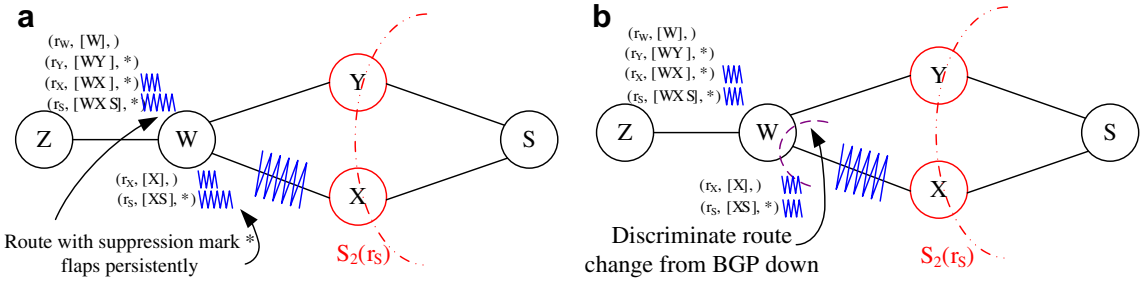


Fig. 5. Suppressing route flap resulting from BGP link failure. (a) Flap of r_S. (b) W damps ap of r_S from BGP down.

applies RFD regardless of the presence of a *Suppression Mark*. Persistent route flap is then completely limited by the neighboring nodes adjacent to the erroneous links.

On some commercial routers, RFD is disabled. Any route flap damping mechanism must take such nodes into consideration, as well as nodes that have implemented other damping mechanisms such as traditional RFD. For example in Fig. 5(b), if W is not using any form of damping, persistent flapping r_S will pass through W and reach Z. If r_S is marked with a *Suppression Mark* by X, according to Neighboring Nodes Suppression, Z will not apply RFD to r_S. To solve such problems resulting from partial deployment, we extend the size of the *Suppression Mark* from one bit to sixteen bits, and record the ASN of the last node with Neighboring Nodes Suppression enabled. Before transmitting the selected route to neighbors, nodes attach their ASN to the route as an optional transitive route attribute. If the *Suppression Mark* of a received route is not the ASN of the peer sending the route, it can be inferred that the peer has not enabled Neighboring Nodes Suppression. So, to protect nodes from route flap arising from link failure, all routes received from peers that have not enabled Neighboring Nodes Suppression should apply traditional RFD, if it is enabled.

6. Invalid routes damping

Traditional RFD focuses on persistent route flapping. With the default RFD parameters of Cisco routers (Table 1), if the average interval between successive flaps is less than 15 min, the route will be suppressed by RFD, because $P_W \times (1 + 1/2 + 1/4 + \cdots) \rightarrow P_{cutoff}$. Neighboring Nodes Suppression can only damp persistent route oscillation resulting actual network changes, but it does not damp invalid routes generated during path exploration.

6.1. Characteristics of invalid routes

Though traditional RFD also has a damping effect on invalid routes generated during path exploration, due to its interaction with path exploration, it does not damp invalid routes very well. For example, the interaction between RFD and path exploration results in the false suppression of relatively stable routes. The propagation of invalid routes in path exploration has the following characteristics:

(1) The interval between successive invalid routes is MRAI, the default value is 30 s, which is much less than 15 min.
(2) The maximum continuation time of path exploration is $MRAI * n$, where n is the largest path length in the network.
(3) The routes seen during path exploration change route attributes, especially AS_PATH, with a Withdrawal ending the procedure. By contrast, during route flap Withdrawal and Announcement messages are alternately transmitted.
(4) The LOCAL_PREF value of successive invalid routes decreases monotonically. If nodes select routes according to the shortest AS_PATH, the length of the AS_PATH increases monotonically.

Invalid Routes Damping is a mechanism to curtail path exploration that better fits the characteristics of invalid routes. The mechanism we proposed in this paper incorporates Neighboring Nodes Suppression and Invalid Routes Damping to reduce the volume of Updates caused by route flapping and path exploration. If a node deploys the modified RFD mechanism, the processing a received route will undergo is determined according to the judgment as illustrated in Fig. 2. If a node receives a route with an attached *Suppression Mark* the node applies Invalid Routes Damping, otherwise, Neighboring Nodes Suppression is applied.

6.2. Processing of invalid routes

Neighboring Nodes Suppression provides a method to identify route flapping due to protocol behavior. If the *Suppression Mark* of a route is the ASN of the sending peer, the receiving node can trust that it will not experience persistent flapping of the route. However, Neighboring Nodes Suppression is not sufficient to reduce the negative effect of path exploration. For this, Invalid Routes Damping is used. The processing procedure of Invalid Routes Damping is similar to that of traditional RFD: maintaining a *Penalty* for each route, increasing the *Penalty* when a route changes, and suppressing the route if its penalty value

exceeds a predefined threshold. However, Invalid Routes Damping has several new properties compared to traditional RFD:

(1) The penalty assessed for attribute changes is greater than that for Withdrawals.
(2) The penalty of routes which are not suppressed decays exponentially, with half life set to the default value of MRAI.
(3) If a suppressed route remains stable for $k * MRAI$, where k is a configurable parameter, the suppression is released.
(4) If a suppressed route changes, the reuse timer is reset.

Traditional RFD employs a penalty as an indicator of the stability of a route, and predicts future changes based on this penalty. The more frequently a route changes, the greater the penalty becomes and the longer the route is suppressed. The penalty in Invalid Routes Damping is an indicator of whether path exploration has taken place, i.e. if a route changes several times and the interval is about MRAI, then the node determines that path exploration has commenced and suppresses the route. The convergence time of the route then has no relation to how many changes the route has experienced. If a route remains stable for $k * \text{MRAI}$, the node determines that path exploration has ended and releases the suppression. To reset the reuse timer of a suppressed route after a change is equivalent to suppress the route until it keeps constant for $k * \text{MRAI}$. It has been pointed out that there are different optimal MRAI values for different topologies [17], but it is difficult for a node to find an optimal MRAI value for each destination node. Since it is impractical to make a comprehensive survey about the MRAI value in practice, we just use the default value for all the nodes.

The complete algorithm for route flap damping combining Neighboring Nodes Suppression and Invalid Routes Damping is shown in 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 their values are different, and k is a configurable parameter.

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
```

7. Simulation

To validate the effectiveness of our design, we have implemented our modified route flap damping mechanism in SSFNet [11] and performed a number of experiments on varying topologies.

7.1. Simulation method

To simplify simulation, each AS in a topology consists of only one border router. The transmission delay between neighboring routers is 0.01 s. All routers use the default MRAI values and WRATE and SSLD are disabled. The parameters of traditional RFD that are used by Neighboring Nodes Suppression are configured to the default values of Cisco routers. A node R is randomly selected as the 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 begins to alternate between the Up and Down states, and R repeatedly withdraws and announces r_d to its neighbors. Each pair of Withdrawal and Announce messages is a flap of r_d. The time interval between the Withdrawal and the following Announcement is represented by *timeAfterWd* and the interval between successive flaps is represented by *timeAfterAnn*. After n flaps, r_d becomes stable again. If not specified purposely, *timeAfterWd* is set to 150 s and *timeAfterAnn* is set to 50 s. Different values of *timeAfterWd* and *timeAfterAnn* are also simulated, influence of different Update intervals is discussed in Section 7.2. In the simulation, we mainly observe two metrics, the convergence delay and the communication overhead. The convergence delay is the interval from between the last Announcement is transmitted and when the route becomes stable in the network. The communication overhead refers to the total number of Update messages transmitted from the first flap to route convergence.

7.2. Synthetic topology

In our first set of experiments, we use a synthetic AS topology generated by BRITE [12,13], which reflects many properties of actual Internet topologies. The simulation topology is generated using the parameters in Table 2. Waxmans probability model is given by $P(u,v) = \alpha e^{-d(\beta L)}$, where $0 < \alpha, \beta <= 1$, d is the Euclidean distance from node

Table 2
Parameters used to generate synthetic topology by BRITE

Parameter	Meaning	Value
HS	Size of main plane	1000
LS	Size of inner plane	10
N	Number of nodes in graph	100
m	Number of neighboring nodes for new nodes	2
Node placement	How nodes are placed in the plane	Heavy tailed
Growth type	How nodes join the topology	Incremental
Bandwidth distr.	Bandwidth assignment to links	Constant
Model	Model type used to generate topology	Waxman

u to node v, and L is the maximum distance between any two nodes.

In the first experiment, we use a 100 AS topology generated by the Waxman model, with parameters $\alpha = 0.2$, $\beta = 0.15$, and all nodes have damping enabled. We compare the convergence delay of flapping routes and the number of Update messages for five different damping mechanisms: *RFD*, the traditional RFD mechanism with parameters of Cisco routers; *Punishless*, where nodes punish less for route attribute changes, i.e. $P_{AC} = 250$; *SRFD*, a method introduced in [5]; MRFD−, a modified RFD with only Neighboring Nodes Suppression; MRFD, newly designed damping mechanism implementing Algorithm 1, with parameters $k = 3$, $P'_{AC} = 1000$, $P'_{W} = 500$, $P'_{cutoff} = 1500$. The results for a varying number of flaps are shown in Fig. 6.

By exploiting the fact that during path exploration only route attributes change, especially AS_PATH, Punishless reduces the penalty increment of P_{AC} to avoid and alleviate false suppression. As can be seen in Fig. 6(a), the convergence time of Punishless is less than that of traditional RFD. With SRFD, the convergence time is reduced compared to PunishLess and traditional RFD, but false suppression is not totally avoided as can be seen when the number of flaps is less than 4. In this case, the convergence delay may be more than 3000 s. In addition, reducing convergence delay using SRFD comes at the cost of increased communication overhead, as can be seen in the figure. MRFD− achieves the best convergence time, but does not prevent path exploration and so its communication overhead is the biggest. MRFD adds invalid route suppression to MRFD−, it also has the lowest overhead. The addition of Invalid Routes Damping results in a slight increase in convergence time, but MRFD still outperforms traditional RFD, Punishless and SRFD.

We use different values of *timeAfterWd* and *timeAfterAnn* to evaluate the influence of Update intervals on the effectiveness of our method. As illustrated in Fig. 7(a), the route almost converges immediately after 1–3 flap(s), and *timeAfterWd* together with *timeAfterAnn* has no influence on convergence delay. For 4–6 flaps, the route is suppressed by MRFD, while as the flapping intensity decreases, i.e. *timeAfterWd* and *timeAfterAnn* increase, the routing convergence delay decreases. The influence of intervals on the number of Update message is illustrated in Fig. 7(b). The number of Update messages increases with the increase of *timeAfterWd* and *timeAfterAnn*. This is because each Announcement/Withdrawal starts a new routing convergence procedure, which breaks down the convergence invoked by the former Withdrawal/Announcement. The interval becomes shorter, so the number of Update messages becomes less.

7.3. 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 [14] to evaluate the effectiveness of MRFD. The configuration of nodes in the topology is the same as that in the last simulation. As shown in Fig. 8(a), 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 relatively stable (flap count less than 3) and almost converge immediately after the Neighboring Nodes Suppression is released when route flaps (flap count more than

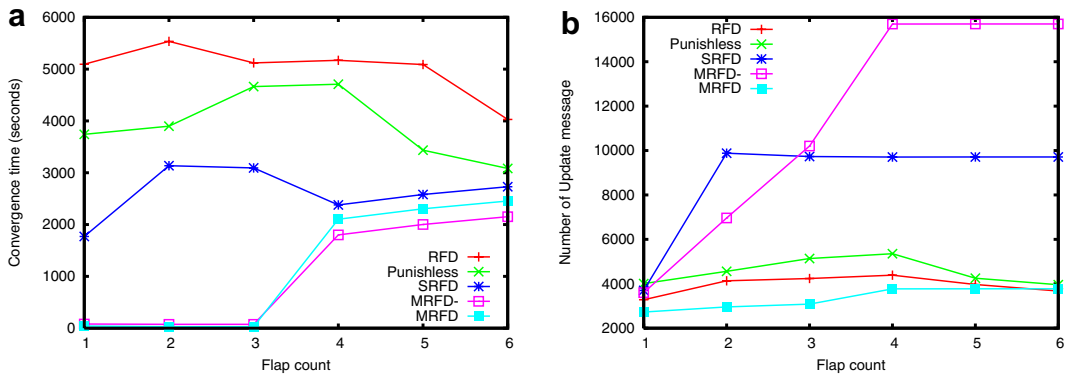


Fig. 6. Simulation result on 100-ASes synthetic topology, damping mechanism is enabled on all nodes. (a) Convergence time of flapping route. (b) Communication overhead.

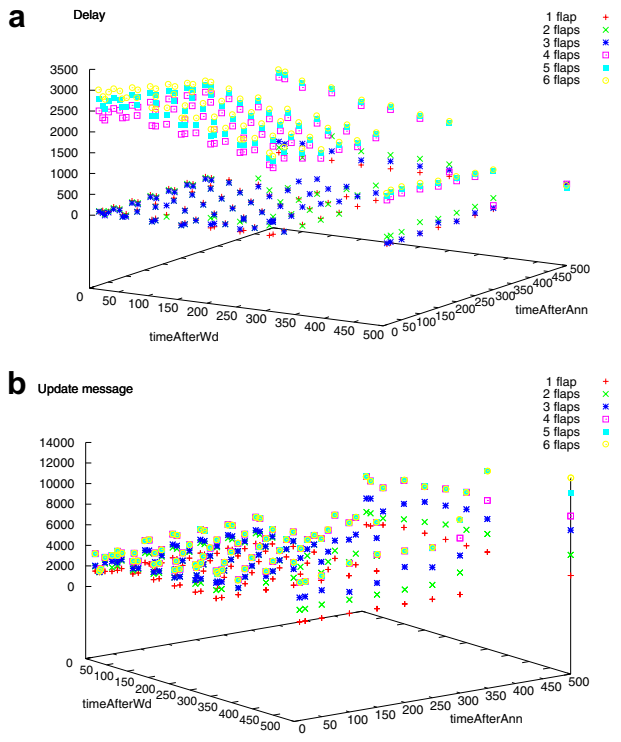


Fig. 7. MRFD convergence with timeAfterWd and timeAfterAnn set to 10, 20, 50, 100, 150, 200, 250, 300, 500 s. (a) Influence of different intervals on convergence delay. (b) Influence of different intervals on communication overhead.

four). The communication overhead of MRFD is a little more than that of RFD and PunishLess (as shown in Fig. 8(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. 9 illustrates the performance of our modified damping mechanism on three Internet-derived topologies, with 29, 110, 208 nodes. The number of Update message in the convergence procedure increases with the topology scale. However, the convergence delays are similar for the three topologies. When the flap count is 1–3, the route converges almost immediately, while when the flap count is larger than 3, the route is suppressed and the length of suppressed period increases with the flap count.

7.4. Partial deployment

We also research the effectiveness of various damping mechanisms with partial deployment. In the synthetic topology of Section 7.2, we randomly selected 5% nodes to be damping mechanism disabled. The percentage of nodes with route converged when route flaps 2 times and 4 times is shown in Fig. 10. We have selected the random 5% nodes multiple times, the experiment results are all like that in Fig. 10. As shown in Fig. 6(a), 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 s. All routes received from nodes whose damping mechanism is disabled will be applied to traditional RFD, and the invalid routes transmitted before Invalid Routes Damping taking effect bring on false suppression. The convergence delay of SRFD and that of MRFD are comparable, both superior to that of PunishLess and RFD.

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 prefers 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 prefers a route from MRFD-enabled nodes than from MRFD-disabled nodes. There are about 50% nodes converge immediately (as shown in Fig. 10(a)), which has obviously superior performance.

The ideal goal of flapping route suppression is that after the route keeping stable for some time, the route converges

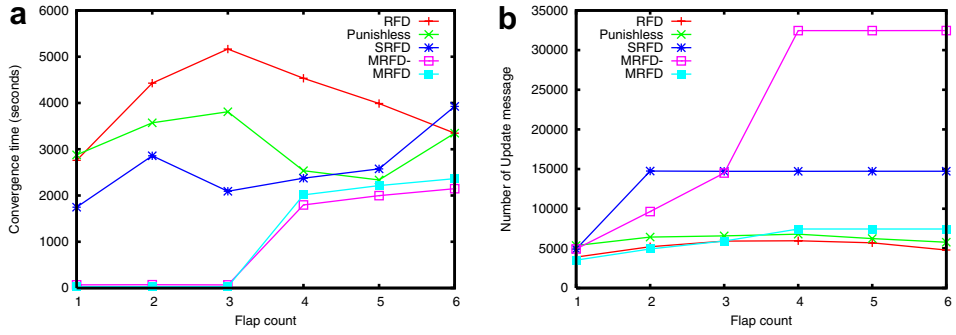


Fig. 8. Simulation result on 110-ASes Internet derived topology, damping mechanism is enabled on all nodes. (a) Convergence time of flapping route. (b) Communication overhead.

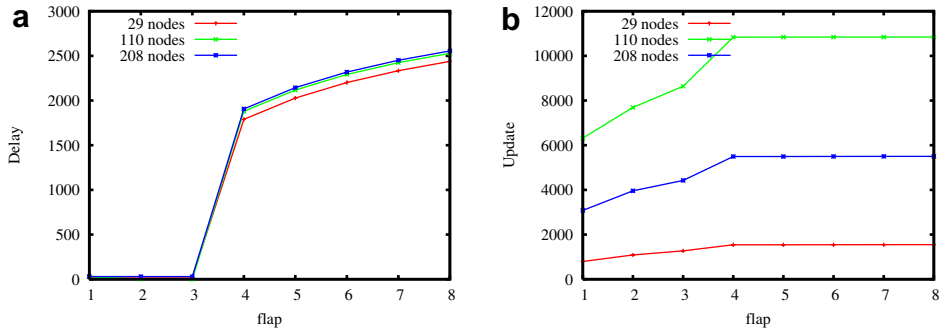


Fig. 9. MRFD performance on topologies of different scale. (a) Convergence time of flapping route. (b) Communication overhead.

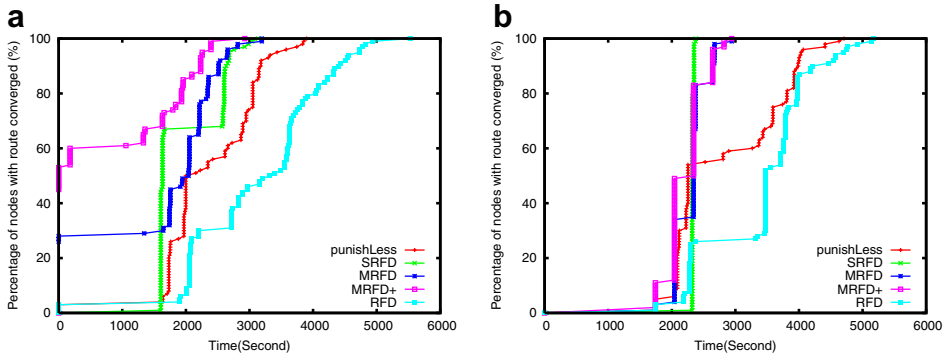


Fig. 10. Route convergence with partial deployment of various damping mechanism, 5% nodes are damping mechanism disabled. (a) Route flaps two times. (b) Route flaps four times.

as soon as possible on all nodes. As shown in Fig. 10(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% only RFD-enabled nodes is shown in Fig. 11. The convergence time for relative stable route and flapping route of MRFD are all superior to that of SRFD. Though false suppression can not be avoided totally, MRFD+ has relatively satisfying effect.

Generally speaking, the performance of partial deployment and coexistence with traditional RFD is reduced severely compared with global deployment of MRFD. However, MRFD is still superior to RFD and other methods, especially combined with routing policy that prefers routes from MRFD-enabled nodes.

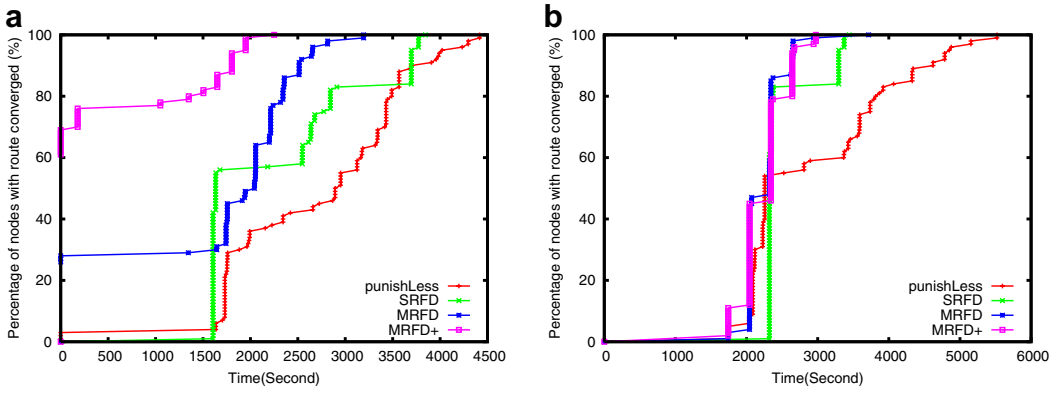


Fig. 11. Route convergence with partial deployment of various damping mechanism, 5% nodes using traditional RFD only. (a) Route flaps two times. (b) Route flaps four times.

8. Conclusion and future work

In this paper, we modify traditional RFD to make it more effective at suppressing persistently flapping routes and reducing invalid routes generated during path exploration. We attribute persistent flapping of BGP routes to two reasons: route instability originating at the source of the route, and instability which is a result of repeated interruption of BGP sessions somewhere along the path of the route propagation. We introduce Neighboring Nodes Suppression to mitigate these two problems. We then introduce Invalid Routes Damping which is designed to reduce invalid routes generated during path exploration. Using Neighboring Nodes Suppression and Invalid Routes Damping jointly, the new damping mechanism, MRFD, can provide significant gains in both convergence time and communication overhead.

Suppression Mark is attached in Update messages to transmit suppression information. Suppression Mark may be implemented as an optional BGP attributes, transitive or nontransitive, depending on the implementation. Thus, MRFD and traditional RFD can co-exist, which allows for the incremental deployment. MRFD is based on the cooperation among ASes by Suppression Mark transmitted between ASes, while traditional RFD judge route flap only according to local observation, which requires that there should be a coordination and trust between ASes.

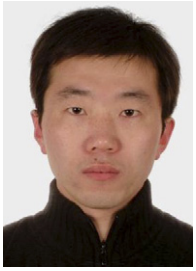
Policy conflicts [15] between ASes may lead BGP routing to divergence. Traditional RFD has some mitigating effect on this problem. Though the modified damping mechanism has no effect on routing divergence resulting from policy conflicts, while this is not the design objective of flap damping mechanism and some protocol design approach [16] and policy configuration guidelines [18] have been proposed. In the design of MRFD, each AS is viewed as a node. However, AS network may has complicate internal structure and interactions. To take the AS internal details into the design is our next step on this work.

Acknowledgement

We appreciate the valuable comments from Michael Buettner of Colorado University.

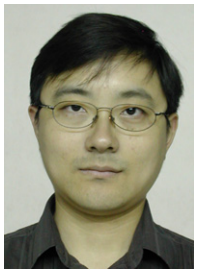
References

[1] Y. Rekhter, T. Li, A Border Gateway Protocol 4(BGP-4), RFC 1771.
[2] C. Labovitz, G.R. Malan, F. Jahanian, Internet routing instability, IEEE/ACM Transactions on Networking 6 (5) (1998) 515–527.
[3] B. Zhang, D. Massey, L. Zhang, Destination reachability and bgp convergence time, in: Proceedings of IEEE Global Telecommunications Conference, vol. 3, 2004, pp. 1383–1389.
[4] C. Villamizar, R. Chandra, R. Govindan, Bgp route flap dampening, RFC 2439.
[5] Z.M. Mao, R. Govindan, G. Varghese, 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, L. Zhang, Timer interaction in route flap damping, in: Proceedings of 25th International Conference on Distributed Computing Systems(ICDCS), Columbus, OH, United States, 2005, pp. 393–403.
[7] C. Labovitz, A. Ahuja, A. Bose, F. Jahanian, Delayed Internet routing convergence, IEEE/ACM Transactions on Networking 9 (3) (2001) 293–306.
[8] C. Panigl, J. Schmitz, P. Smith, C. Vistoli, Ripe routing-wg recommendations for coordinated route-flap damping parameters, RIPE 229.
[9] D. Pei, M. Azuma, D. Massey, L. Zhang, BGP-RCN: improving BGP convergence through root cause notification, Computer Networks 48 (2) (2005) 175–194.
[10] GNU Zebra, <http://www.zebra.org/>.
[11] SSF Research Network, <http://www.ssfnet.org/>.
[12] BRITE: Boston university Representative Internet Topology gEnerator, <http://www.cs.bu.edu/brite/>.
[13] A. Medina, A. Lakhina, I. Matta, 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), Cincinnati, OH, 2001, pp. 346–353.
[14] B. Premore, Multi-AS topologies from BGP routing tables, <http://www.ssfnet.org/Exchange/gallery/asgraph/index.html/>.
[15] K. Varadhan, R. Govindan, D. Estrin, Persistent route oscillations in inter-domain routing, Computer Networks 32 (1) (2000) 1–16.
[16] T.G. Griffin, F.B. Shepherd, G. Wilfong, Policy disputes in path-vector protocols, in: Proceedings of International Conference on Network Protocols, Toronto, Canada, 1999, pp. 21–30.
[17] T.G. Griffin, B.J. Premore, An experimental analysis of BGP convergence time, in: Proceedings of International Conference on Network Protocols, 2001, pp. 53–61.
[18] L. Gao, J. Rexford, Stable Internet routing without global coordination, IEEE/ACM Transactions on Networking 9 (6) (2001) 681–692.
[19] Z. Duan, J. Chandrasheka, J. Krasky, K, Xu, Z, Zhang, Damping BGP Route Flaps, in: Proceedings of IPCCC, Phoenix, AZ, 2004, pp. 131–138.



Wang Lijun was born in Hebei province, P.R.China, in 1978. He received the M.S. from the school of Telecommunication Engineer, Beijing University of Posts and Telecommunications in 2003. Currently he is a Ph.D candidate in the Department of Computer Science and Technology, Tsinghua University. His research interests include Internet architecture and inter-domain routing protocol.

WU Jianping is a full professor of Department of Computer Science, Tsinghua University from 1993. He is also a director of Network Research Center of Tsinghua University. From 1994, he has been in charge of China Education and Research Network (CERNET) which is the largest academic network in China as a director of both Network Center and Technical Board.The major research areas of his group include next generation Internet architecture, terabit IP router, network management and security, P2P and overlay network, QoS management and QoS routing, formal methods and protocol testing. He is vice president of Internet Society of China (ISC), and chairman of APAN. He is a senior member of IEEE.



XU Ke was born in Jiangsu, P.R.China, in 1974. He received the B.S., M.S. and Ph.D. degrees in computer science from Tsinghua University, China in 1996, 1998 and 2001 respectively. Currently he is an Associate Professor in the department of computer science of Tsinghua University. His research interests include next-generation Internet, switch and router architecture, P2P system and overlay network. He is a member of IEEE and IEEE Communication Society.