

An Active Approach to Measuring Routing Dynamics Induced by Autonomous Systems

Samantha Lo
Department of Computing
The Hong Kong
Polytechnic University
Kowloon, Hong Kong
cssmlo@comp.polyu.edu.hk

Rocky K. C. Chang
Department of Computing
The Hong Kong
Polytechnic University
Kowloon, Hong Kong
csrchang@comp.polyu.edu.hk

Lorenzo Colitti
Google Inc.
1600 Amphitheatre Pkwy
Mountain View, CA
lorenzo@google.com*

ABSTRACT

* We present an active measurement study of the routing dynamics induced by AS-path prepending, a common method for controlling the inbound traffic of a multi-homed ISP. Unlike other inter-domain inbound traffic engineering methods, AS-path prepending not only provides network resilience but does not increase routing table size. Unfortunately, ISPs often perform prepending on a trail-and-error basis, which can lead to suboptimal results and to a large amount of network churn. We study these effects by actively injecting prepended routes into the Internet routing system using the RIPE NCC RIS route collectors and observing the resulting changes from almost 200 publicly-accessible sources of BGP information. Our results show that our prepending methods are simple and effective and that a small number of ASes is often responsible for large amounts of the route changes caused by prepending. Furthermore, we show that our methods are able to reveal hidden prepending policies to prepending and tie-breaking decisions made by ASes; this is useful for further predicting the behavior of prepending.¹

1. INTRODUCTION

The Internet is a huge network of hundreds of millions of nodes that exchange information with each other. Because of the dynamic nature and sheer size of the network, routing of data within it takes place on two hierarchical levels. At the highest level, the Internet is partitioned into tens of thousands of administrative domains known as Autonomous Systems (ASes). Each AS is identified by an integer number and typically corresponds to an organization, often an Internet Service Provider (ISP). ASes cooperate with their neighbors to ensure global reachability of all destinations on

*Work carried out while at the RIPE NCC.

¹The dataset presented in this paper is available from <http://www.comp.polyu.edu.hk/~cssmlo/active/> and <http://www.datcat.org/>

the network (*interdomain routing*). At a lower level, each AS typically has a complete view of its own network and is responsible for the routing of Internet traffic within its boundaries (*intradomain routing*).

1.1 The BGP protocol

ASes exchange routing information by means of the Border Gateway Protocol (BGP) [21, 12]. BGP operates by propagating information on the reachability of contiguous blocks of Internet Protocol (IP) addresses known as *prefixes*. Reachability information for a given prefix originates from the AS to which the prefix belongs and is selectively propagated from AS to AS by means of BGP messages known as *route updates*, which can be *announcements* or *withdrawals*. When an AS propagates an announcement, it prepends its AS number to the *AS-path* attribute of the announcement itself. Therefore, the AS-path contains an ordered list of the ASes that propagated the announcement; traffic then flows through these ASes in the opposite direction.

BGP provides rich mechanisms for implementing routing policies through the use of route attributes. ASes may attach transitive attributes, such as MED or BGP communities, to a route in order to affect the routing decisions of other ASes. Upon receiving multiple routes for the same prefix, BGP routers will decide which one to use based on the configured routing policy and the routes' attribute values. The decision is made by comparing the attributes in order: first the route with the highest LOCAL-PREF (local preference) value is chosen, then the route with the shortest AS-path is chosen, then EBGP routes are preferred over IBGP routes, and so on [7, 13], until only one route, the *best* route, remains. The best route is used for forwarding and, if the routing policy permits, is propagated to other routers. The other routes are not propagated.

1.2 AS path prepending

If an AS desires, when originating or propagating an announcement it may prepend multiple copies of its own AS number to the AS-path, a practice known as *AS-path prepending*. Since the length of the AS-path is one of the most important factors in route selection, a multihomed AS can use AS-path prepending to influence incoming traffic distribution. Suppose AS 4616 has two links to the rest of the Internet, link 1 and link 2, and prefers to receive incoming traffic for a given prefix over link 1. To achieve this goal, it can announce the prefix to both links, link 1 with an AS-PATH attribute of 4616 and link 2 with an AS-path

attribute of 4616 4616 4616. The latter is said to have a *prepending length* of two. Thus, other ASes will be discouraged from using the route through link 2 as it will have a longer AS-path than the one through link 1.

Compared to other approaches, such as selective announcements, more-specific prefix announcements, and network address translation-based approaches [14], AS path prepending has the advantages of reducing forwarding table size and providing path resilience to the prefix: if the non-prepended link fails, the prepended one will be used for all traffic. On the other hand, its major shortcoming is that the route changes induced by prepending are unpredictable, so operators often apply the method in a trial-and-error basis.

1.3 Contributions of this work

In this paper, we examine the effect of prepending on Internet routing from both a static and a dynamic point of view, overcoming the problems faced by previous work (which we discuss in §2 and §3) by using an active measurement methodology. The basic idea is simple: an AS sends a route update for an unused prefix and then measures the effects of this route update on Internet routing dynamics from a number of *vantage points* (VPs). This active approach of studying routing dynamics is not entirely new; for example, Mao et al. deployed BGP *beacons* [17] to study route convergence issues for route announcements and withdrawals. The route for a BGP beacon is announced and withdrawn according to a published schedule and the BGP dynamics are then studied by analyzing routing updates collected from the public RouteViews routing information archive [3]. This work is not only useful to understand BGP convergence processes; previous work has also shown that route changes caused by BGP updates can result in route convergence problems [26] and degrade end-to-end path performance [24].

Since previous work has not studied the Internet routing dynamics induced by prepending, in this paper we mostly concentrate on measuring those effects. However, measuring the impact of prepending is considerably more complex than those of route announcements and withdrawals. First, the effects of prepending are not deterministic, as they depend on the complex interaction of upstream ASes' routing policies. Second, each link may be prepended to different lengths (for example, from 1 to 10). Third, as the total number of possible prepending combinations increases exponentially, it becomes impractically large for an AS connected to many other ASes.

Moreover, we are interested in the effect on AS-level Internet paths rather than simply quantifying the changes in inbound traffic distribution. We note that if we assume that the behaviour of users of the network is independent of the AS-path used — in most cases, a reasonable assumption, since routing is transparent to applications — the effect on traffic may be predicted by combining path changes for given ASes with expected or extrapolated traffic values from those ASes. Even if traffic values on given paths are not known, we may use cruder approaches: for example, if we are able to identify an AS that is responsible for the majority of route changes, we might be able to predict traffic changes without performing the large number of probes required by previous work [6].

The most important contribution of this work is the active measurement methodology for studying the effects of route updates on the Internet routing dynamics. Although we

discuss only AS path prepending in this paper, the methodology also applies to route advertisements and withdrawals: indeed, a typical set of prepending experiments also involves advertising and withdrawing the route for the beacon prefix. In particular, we describe our deployment of the methodology using the RIPE NCC RIS [22] beacons and close to 200 VPs. Since the methodology and tools can be deployed in other multihomed ASes, it can also be used to perform measurement studies at transit ASes and coordinated measurement studies across different geographical locations.

Although the analysis of the measurement results is still in a preliminary stage, we report non-intuitive results in the course of propagating prepended routes. For example, it is surprising to learn that in some of our experiments, only one AS is responsible for most of the route changes induced by prepending. The behavior of this “high-impact” AS is also responsible for determining what prepending lengths are effective. If such an AS exists, small changes in prepending can have a large effect on routing and traffic; our measurement methodology enables us to discover such an AS, if it exists.

The rest of the paper is organized as follows. In §2, we highlight the previous works that are closely related to this paper's focus. In §3, we motivate the need for active measurement by explaining why passive measurement studies alone are inconclusive. In §4, we describe the active measurement infrastructure we deployed using the RIS route collectors and the measurements we performed. In §5, we report a number of preliminary findings obtained from three sets of experiments. We conclude and discuss possible future work in §6.

2. RELATED WORK

In earlier studies, Feamster et al. [9] and Broido et al. [2] report their findings on prepending based on the AT&T backbone data and RouteViews data, respectively. In a recent passive measurement study, Wang et al. [11] give a more up-to-date report using various statistics compiled from the RouteViews data and proposes a model to study fundamental issues of decentralized traffic engineering in the Internet.

In terms of active approaches, Chang and Lo [6] propose **AutoPrepend**, an automated procedure to determine the best prepending length before effecting the change. A key component of **AutoPrepend** is the use of a beacon prefix to predict the magnitude of the incoming traffic variation due to prepending. Recently, Quoitin et al. [20] conducted similar active measurement experiments to study the prepending method, and evaluated the effect on inbound traffic distribution. Moreover, they built a new simulator to study the prepending method, and reported a number of findings.

However, neither work attempts to analyze the AS-level route changes induced by prepending. In contrast, the goal of our active measurement is to understand the effects of prepending on AS-level routes, for the reasons given in the last section. To do this, we observe route changes from a number of vantage points in the Internet. Although [20] also examines routing tables in some upstream ASes, it is practically infeasible to achieve the scope of measurement described in this paper using that methodology.

By observing and analyzing route changes, we expect to be able to explain some of the results observed in [6] and [20]. Furthermore, the analysis of the results obtained using our active measurement methodology could be used to improve

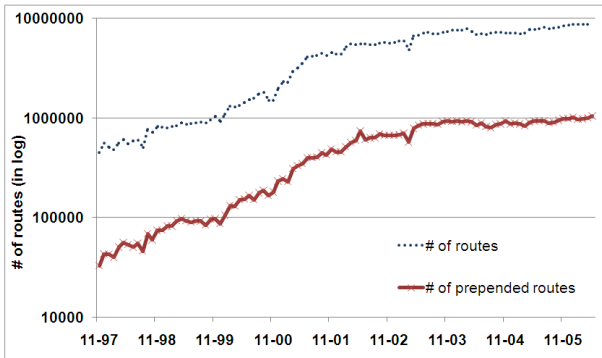


Figure 1: The number of prepended routes observed in the RouteViews archive.

AutoPrepend’s prediction accuracy.

Recent work has focused on the development of methodologies to determine optimal prepending lengths. Gao et al. [10] and Di Battista et al. [4] have proposed algorithms to solve optimization problems for the prepending method. These methods assume that all routers in an AS select the same route and make routing decisions based only on the shortest AS path. However, these assumptions do not hold in practice. As can be see in §5, our results show that this is often not the case.

Finally, we note that prior to this work, we conducted a set of preliminary active experiments for a stub AS [16].

3. WHY PASSIVE MEASUREMENT STUDIES ARE INCONCLUSIVE

In this section we examine measurement results previously derived from RouteViews data [11] to discuss what we can and cannot infer from them. Figure 1 shows the total number of routes and the number of prepended routes observed by the RouteViews server over time. A prepended route is one that contains duplicate AS numbers; these may have been inserted by the origin AS, by another AS on the path, or both. Various factors are responsible for the increase in the number of routes. The most obvious one is the increase in the number of networks connected to the Internet. Another is the increase in the number of RouteViews peers; the new peers provide different views of the network and thus new routes that could not previously be observed. Moreover, many routes are split into multiple routes as a result of selective announcements and traffic engineering[23]. As can be seen in Figure 1, the ratio of prepended routes to total routes has remained fairly constant over time.

However, we note that due to its very nature, this data is not sufficient to quantify the incidence or the effectiveness of prepending. Recall from §1 that the purpose of prepending is to discourage using a given route; therefore, if successful, the prepended route will not be preferred, and since only best routes are announced, will not be propagated and thus not appear in RouteViews data. Thus, we cannot measure how effective prepending is simply based on counting prepended routes observed in passive measurement data: the more effective it is, the fewer times we will observe it.

To probe further, in Figure 2 we show the distribution of prepending lengths over time. We refer to the prepended routes with a prepended length of $i \geq 1$ as i -prepending

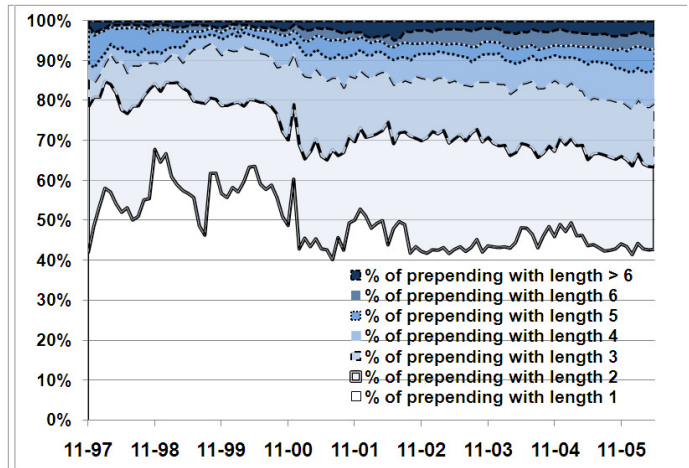


Figure 2: The distribution of prepending lengths observed in the RouteViews archive.

routes. That is, the AS-path attribute contains $i+1$ identical AS numbers. At least 50% are 1-prepended routes, followed by 2-prepended routes and 3-prepended routes. The most noticeable change in recent years is the increase in the share of prepending lengths of four or more.

The presence of a relatively large number of 1-prepended routes is probably due to the fact that a prepending length of one is not sufficient to affect the original routes. As a result, these 1-prepended routes are still preferred. This can apply to other lengths as well; however, a longer prepending length is less likely to be preferred, so we should see fewer visible routes with high prepending lengths. As for the increased share for $i \geq 4$, one possible explanation could be that some operators increased the prepending length beyond three after discovering that a shorter length is not sufficient to affect traffic. Once again, passive measurement data alone cannot help us understand whether a longer prepending lengths are effective or not. As we shall see in the following section, the use of active measurements can overcome this problem.

4. ACTIVE MEASUREMENT

The idea behind our active measurement approach is simple. To evaluate the effects of prepending on one link, we announce a prepended route to the link, observe the resulting routes from a set of well-spread locations, and compare them to the routes seen in the absence of prepending. We define a *link* as a BGP session between two adjacent ASes; in our case, a link is between the AS which announces prepended routes and its upstream AS. Although the basic idea is simple, deploying it in an error-free and fashion is a laborious undertaking that needs to be automated. In the following we describe the three key components in the active measurement methodology: beacon prefixes, route announcer, and vantage points.

1. **Beacon prefixes** To minimize disruption of normal Internet traffic, we send updates for a set *beacon prefixes* not used for production traffic. The length of the prefixes, /24, ensures that upstream routers will not filter them due to prefix length. Since the routing policies currently in use on the Internet are typically based on AS-path, prefix length and point of origin,

we expect the rest of the Internet to treat the beacon prefixes the same way as prefixes used for production traffic. Thus, the effects of prepending observed for the beacon prefixes will be representative of the behavior for prefixes used for production traffic.

2. **Route announcements** There are a few important issues regarding the announcement of the beacon prefixes. First, the scheduled announcements should not interrupt routers' normal operations. Second, the time between two consecutive announcements should be long enough for the newly announced routes to converge before measurement and for route flap damping [17] to expire. Third, we need to ensure that the path changes observed are the results of our active measurements and not of other Internet events or topology changes. For this purpose, we also announce a number of *control prefixes* without any prepending.
3. **Vantage points** We observe the route changes using publicly accessible sources of BGP information, or *vantage points* (VPs), which provide access to their own BGP routes or BGP routes learned from other peers. The main advantage of using VPs is that we can access them without any pre-arranged coordination with other ASes. A similar concept has been used for route convergence studies [17]. For this work we used data from almost 200 VPs.

4.1 Measurement infrastructure

In this section we describe the measurement infrastructure we used to send BGP announcements and observe route changes. An overview of the infrastructure is shown in Figure 3.

4.1.1 Route announcement infrastructure

BGP announcements for the beacon prefixes and control prefixes were made by the the remote route collectors (RRCs) of the RIPE NCC RIS project (AS12654) [22]. After performing preliminary evaluations of all 12 RRCs, we chose three of them (RRC07, RRC10, and RRC14) to conduct full-scale experiments. The choice was made based on their diversity in geographical location and Internet connectivity: RRC07 is in Stockholm, Sweden, RRC10 is in Milan, Italy, and RRC14 in Palo Alto, California. Furthermore, each RRC is connected to several upstream ASes, although in order to reduce the complexity of the measurements and analysis², we use only two of them (i.e., two links) for our measurements. Thus, we are able to perform experiments from different physical locations of the same AS and using different upstream ASes from each physical location.

In order to be able to change the prepending lengths announced on different links without having to modify the configuration of the RRCs at every change, we generate routing announcements by using a Linux PC (`moo.ripe.net`), which maintains IBGP peering sessions with the RRCs using the `announcer` [8] software developed by one of the authors. The PC announces the beacon prefixes with an empty AS-path and encodes the desired prepending configuration by attaching BGP community attributes [5] to the

²The same experiments can be conducted for more than two links, but the number of possible prepending combinations increases exponentially.

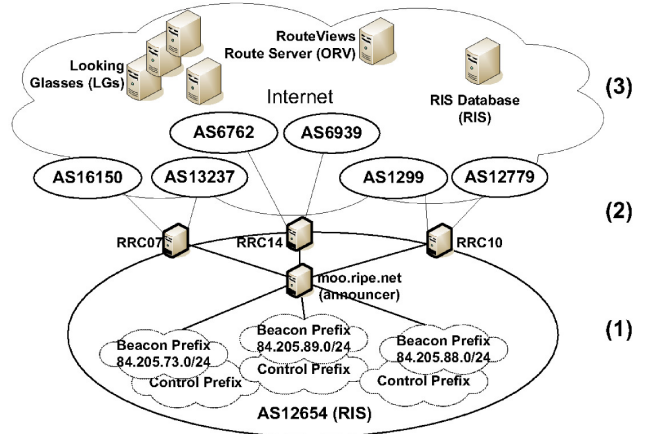


Figure 3: An overview of our active measurement infrastructure. (1) Inside the RIS network (the lower cloud), we use the announcer software on a Linux PC (`moo.ripe.net`) to announce the beacon prefixes to the three RRCs over IBGP with community values encoding the desired prepending. (2) Each RRC decodes the community values using a route-map and announces the prefix with the specified prepending lengths to the upstream ASes. (3) We observe AS-level route changes from the set of VPs (the upper cloud).

announcements. On the RRCs, we configure a `route-map` to decode the community attributes and announce the desired prepending length to each upstream. Since `announcer` is flexible and simple to configure, this allows us to keep configuration overhead to a minimum and avoid configuration errors. Furthermore, it allows us to modify the prepending lengths automatically using simple scripts run on the PC.

Unlike Quoitin et al. [20], we do not restart BGP sessions for each measurement, as this causes brief disruptions in connectivity and convergence problems due to route flap damping [26] and thus cannot be done in an operational network.

4.1.2 Route collection infrastructure

We observe possible route changes due to prepending from a set of almost 200 VPs, including 99 public looking glasses (LGs) [1], the RouteViews (ORV) route server [3], and the RIS database [22]. We have not yet used PlanetLab nodes [19] as VPs, because AS-level traceroute is still not entirely accurate [18]. However, if full accuracy in the AS path tracing is not required, the active measurement methodology can easily include them as additional VPs.

We query the VPs at least 100 minutes after each route announcement, which is long enough for convergence according to [26]. To obtain routes from the LGs, we developed scripts to parse the HTTP responses. For ORV, we obtain routes from its more than 50 peers via a telnet connection. For the RIS, we query the database internally for convenience, but the same information is available via the public web interface. The RIS provides the routes collected by 12 RRCs, which altogether have more than 300 peers, but we can use only 50 of them, because not all of them provide full routing tables. The entire process of collecting the routes from all

RRC	Upstream ASes		Beacon prefix (control prefix)	Announcement period	MPL
RRC07	AS16150 PORT80	AS13237 ⁺ LAMB DANET	84.205.73.0/24 (84.205.88.0/24)	8th - 9th May, 2006 (update every 2 hours)	6
RRC14	AS6762 SEABONE-NET	AS6939 ⁺ HURRICANE	84.205.89.0/24 (84.205.95.0/24)	8th - 9th May, 2006 (update every 2 hours)	6
RRC10	AS1299 TELIANET	AS12779 ⁺ ITGATE	84.205.88.0/24 (84.205.73.0/24)	13th - 15th May, 2006 (update every 3 hours)	10

Table 1: Experiment settings for RRC07, RRC14, and RRC10. We prepend the routes announced to the upstream ASes labeled by ⁺. The prepending length is increased from 1, 2 ..., and up to and including the maximum prepending length.

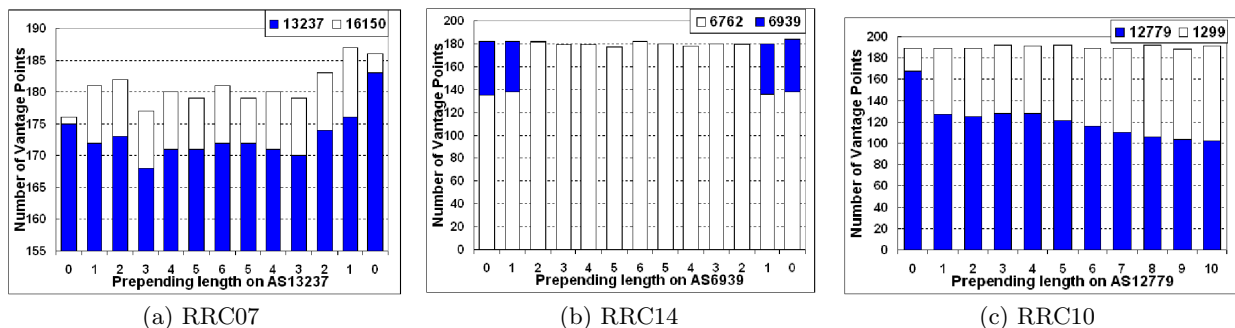


Figure 4: The distributions of incoming link usages from VPs at different prepending lengths.

the VPs is very efficient, taking less than 15 minutes.

Note that there may be multiple VPs residing in the same AS: for example, a router may peer with ORV and another host from the same AS may serve as a LG. For the sake of simplicity, we shall refer to any AS that provides its routing information through LG, ORV, or RIS as a VP.

Another source of routing information available to us is the BGP updates obtained collected and made available by the RIS and RouteViews. The use of BGP updates allows us to track the evolution of a particular VP’s route to the beacon and control prefixes, thus measuring the speed with which changes in prepending take effect. Moreover, for every change in prepending we can study the number of updates caused and how different ASes responded to it. We present an analysis of this data in §5.2.4.

4.2 Route announcement experiments

We conducted measurement experiments from RRC07, RRC14, and RRC10 in May 2006. Table 1 shows the details of our experiments. For each RRC, we prepend only one of the two links, and change the prepending length every two to three hours. We refer to the link that the prepending is applied to as the *prependended link* (PL) and to the other link as the *nonprependended link* (NL). Note that the *maximum prepending length* (MPL) we used for RRC10 is longer than the others, because we could still observe noticeable route changes after prepending six times. As can be seen in Table 1, some of the prefixes were re-used on different RRCs, as the measurements were performed on different days.

5. RESULTS AND ANALYSIS

In this section we first present the overall measurement results obtained for the three RRCs, and then delve into a number of specific findings derived from a more in-depth analysis of the routes observed.

The three RRCs show three different scenarios of prepending behaviour. On RRC07, we prepend on the busy link (i.e. the link with more paths through it) and observe few changes in routes; on RRC14, we prepend on the less busy link and observe a strong effect; and on RRC10, we prepend on the busy link and achieve an intermediate response.

Figures 4(a), 4(b), and 4(c) show how link usage varies with prepending: the X axis shows the prepending changes over time and the Y axis shows the number of VPs that use the PL (in blue) and NL (in white) at different prepending lengths. The measurement results for RRC07, in Figure 4(a), show that increasing the prepending length for AS13237 does not have a large effect on routing. The measurement results for RRC14, in Figure 4(b), show that when the prepending length on AS6939 is increased to two, there is an abrupt change – nearly all VPs switch to the NL (AS6762). When we further increase the prepending length to three, none of the VPs, including AS6939 itself, uses the PL (AS6939). On the other hand, the measurement results for RRC10 are very different from the other two RRCs. Figure 4(c) shows that when the prepending length is increased to one, there is already an abrupt change. But when we further increase the prepending length beyond five and up to ten, there are still noticeable changes in the link usages. Moreover, when the prepending length is increased to 10, there are nearly equal number of VPs using the two links.

Figures 6, 7, and 8 show the AS-level topologies for the beacon prefixes announced by the three RRCs. We will use them to explain discuss interesting effects of prepending later on. Table 2 summarizes the effects of prepending in terms of the number and percentage of VPs and ASes that switch from the PL to the NL. The impact of prepending is very low for RRC07 in spite of prepending on the “busy” link, which is used by the majority of VPs when there is no prepending. In contrast, after prepending on the

	RRC07		RRC14		RRC10	
Level-1 ASes:	AS13237 ⁺	AS16150	AS6939 ⁺	AS6762	AS1299	AS12779 ⁺
No. VPs via the level-1 AS:	184	6	47	139	24	164
No. ASes via the level-1 AS:	203	9	55	151	26	180
No. (%) of VPs switched from the PL to the NL:	8 (4%)	-	47 (100%)	-	-	63 (38%)
No. (%) of ASes switched from the PL to the NL:	12 (6%)	-	55 (100%)	-	-	68 (37%)

Table 2: The measurement results for RRC07, RRC14, and RRC10 after prepending with their MPLs.

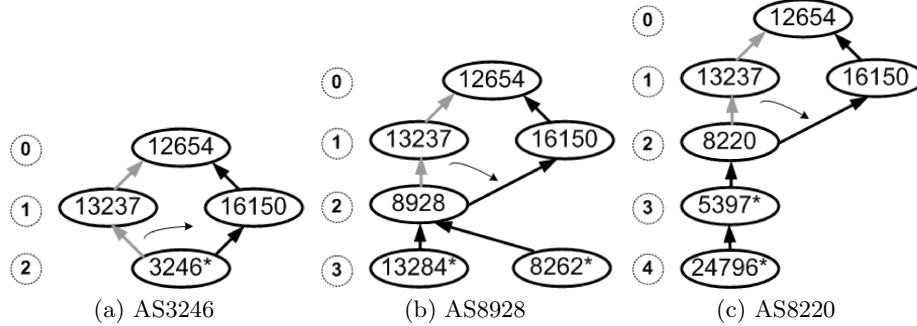


Figure 5: Three examples of switching from the PL to the NL. The gray lines correspond to the old routes before prepending; the black lines correspond to the new routes effected by prepending. The AS with * is a VP.

“non-busy” link of RRC14, all VPs and ASes switch to the NL, leaving the other link empty. In the case of RRC10, prepending on the busy link has a high impact, switching almost 40% of VPs and ASes to the NL.

5.1 Classifying ASes based on their response to prepending

Based on their response to prepending, we classify upstream ASes into *responsive* and *non-responsive* ASes. We further note that sometimes we observe that one or more ASes, which we name *high-impact responsive* ASes, are responsible for most route changes. Identification of these ASes allows greater accuracy in predicting the effects of prepending. In the following, we define the *next-hop AS* of an AS x as the first AS in the AS-path seen by x for the beacon prefix, i.e. the AS to which x will route any traffic addressed to the beacon prefix.

5.1.1 Responsive ASes

Responsive ASes are ASes that switch from the PL to the NL after sufficient prepending. However, not all of them change their next-hop AS. Therefore, we further subdivide responsive ASes into *direct-responsive* ASes (DR-ASes) and *indirect-responsive* ASes (IR-ASes), based on whether they change their next-hop AS. Consider the three examples in Figure 5. AS3246 changes its next-hop AS after receiving a prepended route from AS13237 (Figure 5(a)). Therefore, we say that AS3246 is a direct-responsive AS. However, AS13284, AS8262, and AS24796 (Figures 5(b) and (c)) do not change their next-hop ASes; therefore, their switch to the NL (AS16150) is likely the result of a path change of their upstream ASes. Therefore, we say that these ASes are indirect-responsive ASes. Direct-responsive ASes account for 58%, 13%, and 32%, respectively, of the responsive ASes

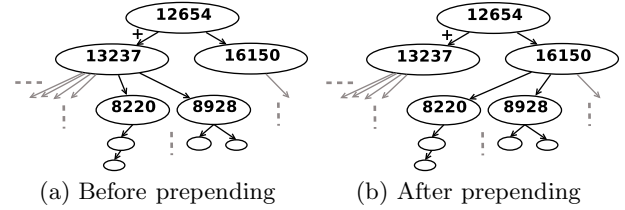


Figure 6: AS-paths to the beacon prefix at RRC07.

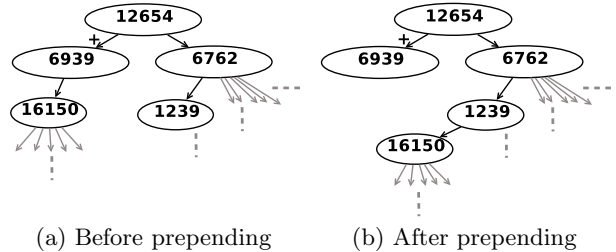


Figure 7: AS-paths to the beacon prefix at RRC14.

for RRC07, RRC14, and RRC10.

5.1.2 High-impact, responsive ASes

In each set of responsive ASes, we can map a direct-responsive AS to a set of indirect-responsive ASes in which it induces link changes. Intuitively, we may think of the DR-AS as being at the root of a “subtree” of its IR-ASes. For RRC07, the DR-ASes we observe (e.g., AS8220 and AS8928 in Figure 6) map to at most two IR-ASes. For RRC14, AS16150 is a *high-impact AS*, responsible for the drastic ef-

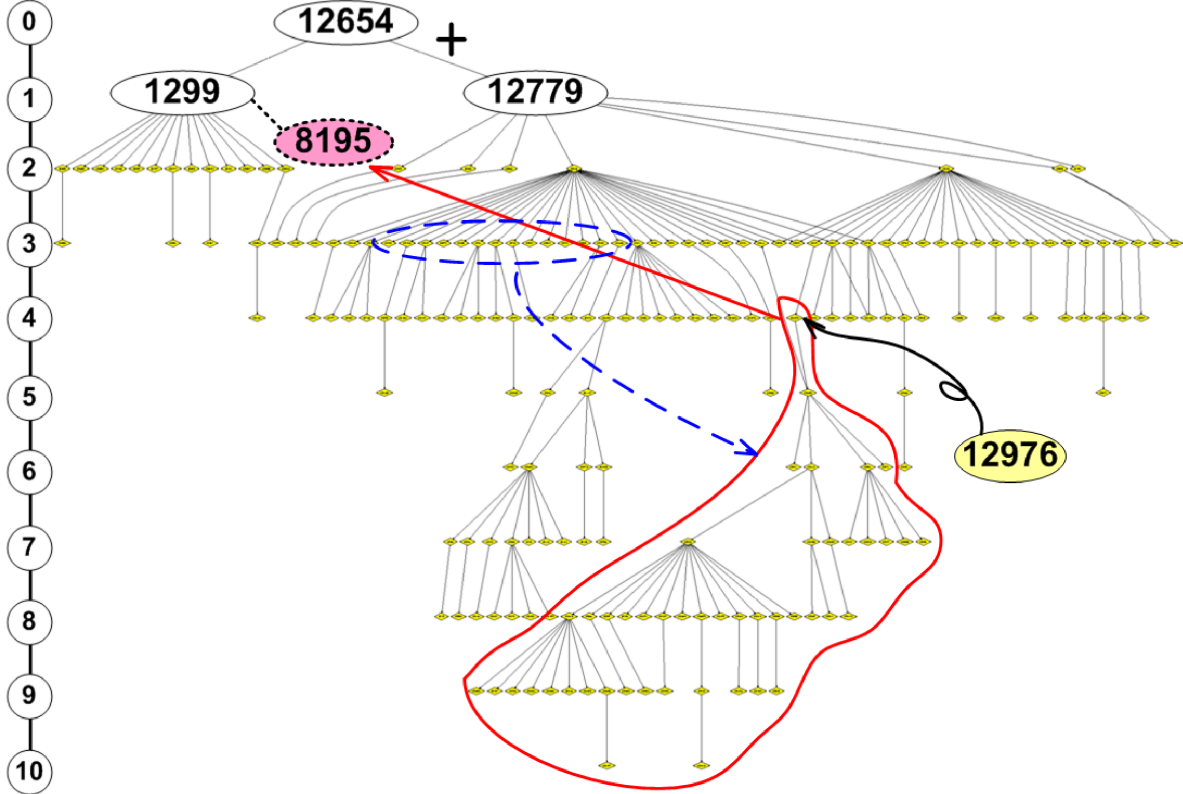


Figure 8: The AS-level routing paths to reach the beacon prefixes without prepending for RRC10. Note that AS8195 is not part of the topology before prepending; it appears only after prepending on the link to AS12779.

fect of the prepending method. AS16150 maps to 46 out of 47 IR-ASes, so, as can be seen in Figure 7, the entire subtree of 46 ASes under AS16150 is affected by prepending and moves to the new location: AS16150’s next-hop change induces link changes in all but one VP. For RRC10, AS12976 is a high-impact AS, as it is responsible for almost 60% of VPs that change to the NL. A more complete topology diagram is shown in Figure 8, in which AS12976 and its IR-ASes are enclosed by a solid red line.

5.1.3 Effect of prepending to high-impact ASes

Past work has observed abrupt changes in inbound traffic distribution at certain prepending lengths (e.g., one in [20] and three in [6]). We have also observed such abrupt changes in terms of the numbers of VPs or ASes for RRC14 and RRC10, respectively at prepending lengths of two and one. These two “special” prepending lengths are in fact not special at all: they are the respective minimum prepending lengths for AS16150 and AS12976 — the high-impact ASes in RRC14 and RRC10, respectively — to change their next-hop ASes. Thus, a high-impact AS dictates the prepending length for which an abrupt route change will occur. Since the topologies of the three RRCs are different, the prepending lengths at which this occurs is different for each RRC.

Moreover, there is a difference between the prepending length distributions for RRC14 and RRC10. For RRC14, we do not observe route changes beyond a prepending length of

three, while for RRC10, route changes are still visible even at a prepending length of ten. The underlying reason is due to a complex interaction of routing decisions. Consider Figure 8. At a prepending length of one, AS12976 changes its next-hop AS to AS8195, thus causing route changes in all its IR-ASes and moving its whole subtree of IR-ASes. As we further increase prepending, a number of ASes (enclosed by the blue dashed line) move to the AS12976 subtree. In other words, the migration of the AS12976 subtree facilitates other route changes. As a result of these route changes, the non-prepended routes seen by some of these ASes are longer, thus requiring further prepending to induce route changes. We note that once again, it would be very difficult to obtain this information without the use of active measurements.

5.2 Inferring commercial relationships, BGP tie-breaking, and routing policies

Besides identifying DR-ASes and high-impact ASes, it is useful to understand why some of the upstream ASes are DR-ASes but others are not. Furthermore, when we prepend to certain length and both upstream paths seen by a DR-AS are the same, it would be useful to predict which path it would pick. In the following we discuss three factors that contribute to these decisions: commercial relationships between ASes, BGP tie-breaking, and hidden prepending policies. Again, our active measurements allow us to understand how these three issues contribute to the effects of prepend-

ing.

5.2.1 Commercial AS relationships and prepending

Since DR-ASes respond to prepending, their higher-ranked BGP attribute values, such as LOCAL-PREF values, must be the same for the old and new next-hop ASes. To explore further, we use the algorithm from [25] to infer the commercial AS relationships of two next-hop ASes. For about 40% of DR-ASes, the result is that both next-hop ASes are inferred to be providers. That is, those DR-ASes switch from one provider to another provider, and they assign the same LOCAL-PREF value to both providers. The algorithm fails to identify the relationships for 30% of DR-ASes; the remaining cases are a mixture of sibling-to-peer, sibling-to-sibling, and so on.

5.2.2 Tie-breaking route decisions

The computer simulation in [20] concludes that it is very common to reach the tie-breaking rules in BGP's route decision process (Where a tie-breaking rule is any rule of lower rank than the AS-path rule, and is usually not visible outside the AS in question.) However, our active measurement can help identify tie-breaking route decisions. As an example, consider two VPs in AS3292 and AS15389. AS15389 uses the route via AS3292. The paths used by AS3292 and AS15389 in the absence of prepending are 3292 1239 12779 12654 and 15389 3292 1239 12779 12654, respectively. When prepending length is increased to six (that means the AS path length from AS15389 increases to 11), AS15389 changes its route to 15389 3292 8342 2118 20483 12976 8195 1299 1299 12654 which has the same AS-path length as the prepended route. However, AS3292 continues to use the route via AS12779 until the prepending length is increased to seven. It is possible that AS3292 uses both paths at the same prepending length. Another possibility is that it announces both routes to AS15389 but under some tie-breaking rules, AS15389 will select the new one when the AS path lengths are the same.

5.2.3 Hidden prepending policies and topology

Our active measurements can also expose hidden prepending policies and hidden ASes. Recall from Figure 8 that prepending on AS12779 reveals AS8195. The prepending also uncovers that AS1299 prepends twice the route sent to AS8195, with the goal of discouraging traffic from AS8195. For example, consider the route from AS16150 to AS12654 before prepending: 16150 8342 2118 20483 12976 1273 1239 12779 12654. After prepending once on the link to AS12779, AS12976 changes its next-hop AS to AS8195 and the new path is 16150 8342 2118 20483 12976 8195 1299 1299 1299 12654. In other words, the change in prepending *cancel out* the effect of the prepending inserted by AS1299. Besides recovering AS8195.

Finally, active measurement can discover other ASes that are not visible from the VPs in the absence of prepending, thus allowing a richer AS-level topology to be discovered. For instance, the prepending announcements for RRC07 increase the number of ASes seen in the topology by 18.7%.

5.2.4 Route convergence

To observe the effect of prepending on route convergence, we analyze the BGP updates observed during our measurements. The updates were collected from the RouteViews

archive and the RIS raw data web site. Whenever a RIS or ORV VP changes its best route to a beacon prefix, it announces the new best to the RIS or ORV collectors, which archive it in raw data files that we download and parse. Therefore, every announcement we observe for a beacon prefix indicates that a VP has changed its best route for that prefix and provides the AS-path of the new route. Similarly, every withdrawal we observe indicates that a VP has no route to the prefix.

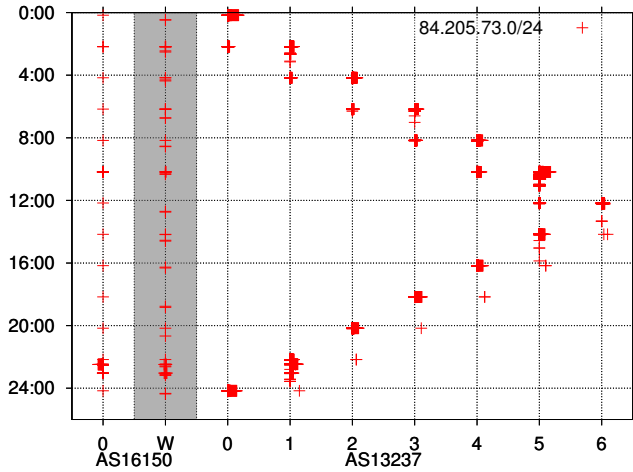
We performed two different experiments. First, in order to examine BGP behaviour both in the case of increasing and decreasing prepending lengths, RRC07 and RRC14 started from a prepending length of zero, incremented the prepending length by one for each successive announcement up to an MPL of 6, and then symmetrically decremented it back to zero. Announcements were made every two hours. Second, in order to determine whether convergence time was longer than the two hours used in the other experiments, and to investigate behaviour with long prepending lengths, RRC10 made announcements every three hours, starting from a prepending length of zero and increasing the length by one with each announcement up to a MPL of 10.

The graphs in Figure 9 plot the updates collected for the beacon prefixes announced by RRC07, RRC14, and RRC10 respectively. Every + symbol represents an update. For every graph, the Y axis shows the time in hours from when the first update was sent, and the X axis shows the link and prepending count seen in the update, as follows. An update in the shaded column, labeled "W", indicates that the VP sent a withdrawal because it did not have any usable route to the beacon prefix; updates to the left of the "W" column are announcements using the non-prepended link; and updates to the right of the "W" column are announcements using the prepended link, with the position on the X axis label indicating the prepending length seen in the update. The labels on the X axis also shows the AS numbers of the non-prepended link and of the prepended link. In order to represent multiple updates observed around the same time for the same prepending length, the updates are slightly offset horizontally from one another in the graph.

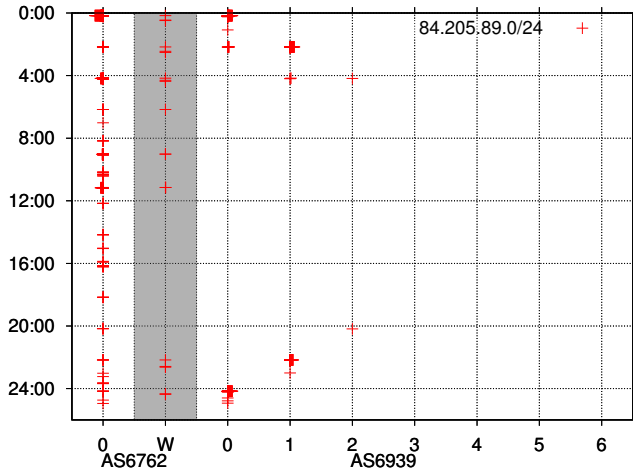
As can be seen from the graphs, when a new prepending length is announced, we observe BGP updates both for the previous prepending length and the new prepending length. For example, in Figure 9(a), at time 2:00, when the prepending length has just been increased to one, we see updates with prepending length 0. We believe this is due to the BGP convergence process, as routers that were using the prepended route switch to alternate routes with the same AS-path length before concluding that the path length has in fact increased by one and accepting the new path; this is a similar phenomenon to that observed in [15] for route withdrawals.

6. CONCLUSIONS AND FUTURE WORK

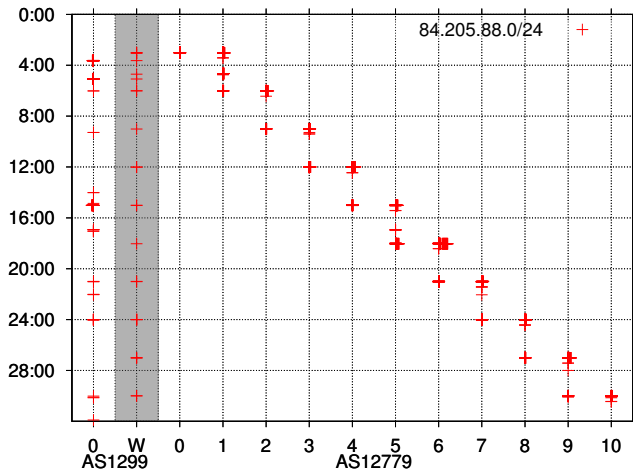
In this paper we propose an active measurement methodology for studying the effectiveness of AS-path prepending and its impact on Internet routing dynamics. We have carefully designed the methodology to minimize disruption to normal Internet services and to make it deployable in a production network. We have so far deployed it to measure the effects of prepending on a stub AS using three of the RRCs of the RIPE NCC RIS project. Our experiments have led us to discover a number of hidden processes in the course of



(a) RRC07



(b) RRC14



(c) RRC10

Figure 9: BGP updates of the three RRCs against different prepending lengths.

propagating prepended routes in the Internet, which have not been discussed before. Furthermore, our methods can help identify tie-breaking route decisions and expose hidden prepending policies and hidden links.

There are several avenues to extend this work. An important issue to study is the stability of the prepending method. The primary metric here is how quickly prepended routes can converge; this can be analyzed by observing route updates. For this purpose, it would also be useful to increase the number of VPs in order to collect more routes and thus have a more representative view of the Internet. On the application side, we are planning to apply the methodology to systematically configure prepending and other traffic engineering methods, such as selective announcement. For example, if we are able to identify a high-impact AS, we know that all routes passing through it will be affected by prepending; thus, if we perform prepending on one of these high-impact ASes (for example, the transit provider of a stub AS), we may obtain more direct control on traffic switching. Finally, we are investigating the relationship between AS path prepending and end-to-end Internet path performance.

Acknowledgments

The work described in this paper was partially supported by a grant from The Hong Kong Polytechnic University (Project no. G-U055) and partially supported the RIPE NCC. It was completed while the third author was at the RIPE NCC.

The authors would like to thank the reviewers and Prof. Dror Feitelson for their useful comments, Prof. Nick Feamster for a long list of editorial comments to improve the readability of this paper, and Ruomei Gao for her insightful suggestions. The authors are especially indebted to the members (Daniel, Grace, Sam, Edmond, and Kathy) of the Internet Infrastructure and Security Group for their continuous inputs and support to this project.

Finally, thanks go to Daniel Karrenberg, Henk Uijterwaal, James Aldridge, and the RIS team at the RIPE NCC for providing the access to the RIS database and infrastructure without which this work would not have been possible.

7. REFERENCES

- [1] traceroute.org. <http://www.traceroute.org/>.
- [2] A. Broido, E. Nemeth, and k. claffy. Internet expansion, refinement and churn. *European Trans. Telecom.*, Jan. 2002.
- [3] Advanced Network Technology Center. University of Oregon Route Views Project. <http://www.routeviews.org>.
- [4] G. Battista, M. Patrignani, M. Pizzonia, and M. Rimondini. Towards optimal prepending for incoming traffic engineering. In *Proc. IPS-MoMe*, 2005.
- [5] R. Chandra, P. Traina, and T. Li. BGP communities attribute. RFC 1997, Aug. 1996.
- [6] R. Chang and M. Lo. Inbound traffic engineering for multihomed AS's using AS path prepending. *IEEE Network*, pages 18–25, March/April 2005.
- [7] Cisco. BGP best path selection algorithm. <http://www.cisco.com/warp/public/459/25.shtml>.

- [8] L. Colitti. announcer software. <http://www.dia.uniroma3.it/~compunet/bgp-probing/announcer>.
- [9] N. Feamster, J. Borcenhagen, and J. Rexford. Controlling the impact of BGP policy changes on IP traffic. Technical Report 011106-02, AT&T Research, Nov. 2001.
- [10] R. Gao, C. Dovrolis, and E. Zegura. Interdomain ingress traffic engineering through optimized AS-path prepending. In *Proc. IFIP Networking conference*, 2005.
- [11] H. Wang, R. Chang, D. Chiu, and J. Lui. Characterizing the performance and stability issues of the AS path prepending method: taxonomy, measurement study and analysis. In *Proc. ACM SIGCOMM Asia*, April 2005.
- [12] J. Stewart. *BGP4: Inter-Domain Routing in the Internet*. Addison-Wesley, 1998.
- [13] Juniper. Examine BGP routes and route selection. <http://www.juniper.net/techpubs/software/nog/nog-baseline/html/verify-bgp9.html>.
- [14] S. Kalyanaraman. Load balancing in BGP environments using online simulation and dynamic NAT. Presented at the Internet Statistic and Metrics Analysis Workshops, available from <http://www.caida.org/outreach/isma/0112/talks/shiv/>, Dec. 2001.
- [15] C. Labovitz, A. Ahuja, A. Bose, and F. Jahanian. Delayed Internet routing convergence. In *Proceedings of ACM SIGCOMM 2000*, pages 175–187, Sept. 2001.
- [16] S. Lo and R. Chang. Active measurement of the AS path prepending method. In *IEEE ICNP '05 (poster paper)*, available from <http://www4.comp.polyu.edu.hk/~csrchang/ICNP05.pdf>. Also presented at NANOG 37, June 2006.
- [17] Z. Mao, R. Bush, T. Griffin, and M. Roughan. BGP beacons. In *Proc. ACM/USNIX Internet Measurement Conf.*, 2003.
- [18] Z. Mao, J. Rexford, J. Wang, and R. Katz. Towards an accurate AS-level traceroute tool. In *Proc. ACM SIGCOMM Conf.*, 2003.
- [19] PlanetLab. home page. <http://www.planet-lab.org/>.
- [20] B. Quoitin, C. Pelsser, O. Bonaventure, and S. Uhlig. A performance evaluation of BGP-based traffic engineering. *Intl. J. Network Management*, (15):177–191, 2005.
- [21] Y. Rekhter, T. Li, and S. Hares. A border gateway protocol 4 (BGP-4). RFC 4271, Jan. 2006.
- [22] RIPE NCC. Routing information service (RIS) project. <http://www.ripe.net/ris/>.
- [23] T. Bu, L. Gao, and D. Towsley. On characterizing BGP routing table growth. In *Proc. IEEE Global Internet Symp.*, Nov. 2002.
- [24] F. Wang, Z. Mao, J. Wang, L. Gao, and R. Bush. A measurement study on the impact of routing events on end-to-end Internet path performance. In *Proc. ACM SIGCOMM*, 2006.
- [25] J. Xia and L. Gao. On the evaluation of AS relationship inferences. In *Proc. IEEE GLOBECOM*, 2004.
- [26] Z. Mao, R. Govindan, G. Varghese and R. Katz. Route flap damping exacerbates Internet routing convergence. In *Proc. ACM SIGCOMM Conf.*, 2002.