Freie Universität Berlin

Department of Mathematics and Computer Science
Institute of Computer Science

Bachelor's Thesis

Route Flap Damping in the Wild?!

Clemens Mosig

Matr. 5183740

Supervisor: Prof. Dr. Matthias Wählisch

Institute of Computer Science, Freie Universität Berlin, Germany

June 22, 2020

I hereby declare to have written this thesis on my own. I have used no other literature and resources than the ones referenced. All text passages that are literal or logical copies from other publications have been marked accordingly. All figures and pictures have been created by me or their sources are referenced accordingly. This thesis has not been submitted in the same or a similar version to any other examination board.

Berlin, June 22, 2020

(Clemens Mosig)

Abstract

BGP connects autonomous systems (ASs) on the Internet by announcing (and withdrawing) routes. Route Flap Damping (RFD) withdraws prefixes for which too many route changes are received to prevent frequent, redundant best path selection in routers. Since its introduction in the 90's, recommendations on configuration and use have been a back and forth, primarily because RFD can lead to unreachability of well behaving prefixes. The state of deployment was not measured, yet.

In this study, we introduce an approach to measure RFD using custom BGP Beacons and public data from BGP route collector projects. From seven locations in the Internet core we insert our Beacons based on a custom schedule to trigger a clear RFD signature. We determine RFD usage on AS paths based on the output signal of the three largest route collector projects. Then, we design and apply heuristics to pinpoint RFD ASs. In our study we uncover that 60% of measured ASs are using deprecated, harmful vendor default configurations. And, against the expectation of the Internet community, we find that at least 6% of measured ASs are using RFD.

Acknowledgements

First and foremost, I would like to give special thanks to Matthias Wählisch who not only patiently supervised me, so that I do not break the Internet, but supported me in every possible way throughout the entire project. I am tremendously grateful for all opportunities and look forward to continue to work in his research group!

I would also like to thank Randy Bush, Cristel Pelsser, and Thomas C. Schmidt for their continuous input, useful suggestions, and discussions! And I am also thanking Marcin Nawrocki for never simply turning down any ridiculous thought and brainstorming ideas with me on the whiteboard.

We would like to thank BKNIX, IIJ, Nianet, Registro, RGnet, Seacom, and SpaceNet for hosting our BGP Beacons. We also thank Randy Bush and RIPE who provided us AS numbers and IP prefixes to conduct our experiments. Thanks to the RIPE community and numerous network operators who shared the RFD configurations with us!

Contents

List of Figures

List of Tables

Listings

Glossary

AS Path List of ASs describing the propagation path of a BGP update. The rightmost AS is the origin AS.

Autonomous System Collection of prefixes maintained by a single organization.

Explicit Withdrawal An explicit withdrawal invalidates the preceding route for a specific prefix and does not provide an alternative.

Forwarding Information Base Information Base containing routing information used for forwarding IP packets (subset of Routing Information Base).

Full Feed Vantage Point A VP exporting its complete RIB to the route collector. A common threshold is 700,000 prefixes.

Implicit Withdrawal An implicit withdrawal is an announcement following another announcement. It provides a new route for a given prefix and implicitly withdraws the preceding route.

Peer A routers' neighbor.

Prefix An IP Prefix represents an address range, *e.g.,* `1.1.1.0/24` contains all IP addresses between `1.1.1.0` and `1.1.1.255`.

Propagation time Time delta between an update sent from the Beacon router and received at a route collector peer.

Re-advertisement RFD context: The advertisement that is sent once a route is considered usable again, *i.e.,* accumulated penalty is smaller than the reuse-threshold.

RFD path The announcement pattern at a vantage point containing the respective AS path is RFD-like. See Section 4.2 for details.

Routing Information Base Information Base containing all routing information received from peers.

Vantage Point A route collector peer.

Acronyms

AIMD Additive Increase Multiplicative Decrease.

AS Autonomous System.

ASN Autonomous System Number.

BGP Border Gateway Protocol.

FIB Forwarding Information Base.

IETF Internet Engineering Task Force.

IP Internet Protocol.

IRR Internet Routing Registry.

MRAI Minimum Route Advertisement Interval.

NLRI Network Layer Reachability Information.

RFD Route Flap Damping.

RIB Routing Information Base.

VP Vantage Point.

CHAPTER 1

Introduction

BGP connects autonomous systems (ASs) in the Internet by announcing (and withdrawing) routes. To prevent oscillating routes, Route Flap Damping (RFD) was introduced in 1998, which suppresses repeating BGP updates. Route flaps cause performance problems on routers, but RFD as a mitigation technique might also suppress well-behaved or stable routes [2] and, in the worst case, leads to unreachability of networks. Recommendations on the use of RFD have been revised multiple times in the last three decades.

Since its first implementation and throughout the years, deployment and configuration of RFD has never been measured. Knowing which ASs use RFD in the Internet is crucial, because it affects convergence and reachability and impacts active and passive control plane measurements.

To measure RFD deployment, we utilize 28 BGP Beacons from seven different locations around the globe and observe these controlled signals in BGP Update dumps using the route collector projects Isolario, RIPE RIS, and RouteViews. Beacons are sent in a non-standard update schedule to create the RFD signature which we can observe at route collector peers (vantage points). Based on the signature we can label AS paths with RFD true or false.

We faced multiple challenges. First, BGP routers may have different RFD configurations, which trigger RFD at different flap rates, which we tackle using multiple Beacon frequencies. Second, ASs are often heterogeneously configured, *i.e.,* suppress route updates from a subset of neighbors. The third challenge, determining whether a vantage point or other ASs on the labeled AS path perform RFD is non-trivial, and requires heuristics. Finally, the visibility of BGP updates is per se limited.

In this thesis we make the following key contributions:

1. We develop a measurement infrastructure—RFD Beacons—to precisely identify RFD deployment on AS paths.
2. We perform the first large-scale study on RFD deployment in the wild, based on more than 2 months of routing data. The results suggest that RFD is used by at least 6% of measured ASs.
3. We uncover that RFD is less carefully used than previously believed. Our results, confirmed by ground truth, indicate that $\approx 60\%$ are using harmful vendor defaults.

We examine the history of recommendations on RFD in Chapter 2, followed by related work in Chapter 3. In Chapter 4 we will give a detailed explanation of our methodology to track down damping ASs on the Internet. The infrastructure setup required for our experiment is described, closely inspected, and validated in Chapter 5. We conclude this thesis with the results in Chapter 6 and our tooling in Chapter 7.

CHAPTER 2

Background

2.1 Internet

The Internet is a network of networks. It consists of 60,000 subnetworks [5], which are called Autonomous Systems (ASs) and can span multiple continents. ASs, which are identified by a globally unique 16-bit or 32-bit number, are interconnected with the Internet Protocol (IP). IP provides an end-to-end service between two hosts in two distinct ASs [6]. It enables hosts to be reachable from any other AS and to communicate across ASs. Each host has a globally unique IP address. To exchange data between two hosts, the sender has to specify the destination IP address and its source address. Packets are then forwarded hop-by-hop between routers to the destination. The relevant versions of IP are version 4 and 6. IPv6 is the latest version, but at this time barely adopted [7]. Routing protocols behave identical for both versions, hence we chose to conduct our measurements only within the IPv4 address space.

For the purpose of efficient routing, IP addresses are hierarchically aggregated into IP prefixes with different sizes. The most common IPv4 prefix size is `/24` [5] which contains all IPv4 addresses where the first 24 bits match, *e.g.,* the IPv4 prefix `1.1.1.0/24` contains all addresses between `1.1.1.1` and `1.1.1.255`. Throughout this thesis we will be referring to IPv4 prefixes simply by the word *prefix*.

2.2 Border Gateway Protocol

The Border Gateway Protocol (BGP) is the inter-domain routing protocol of the Internet [1]. BGP acts as the glue between ASs. To achieve reachability across network boundaries, ASs need to exchange routing information. BGP is path-vector protocol where peers exchange their best paths for each IP prefix.

This information is exchanged in the form of BGP updates between the border routers of neighboring ASs. Figure 2.1 depicts a BGP session between the border routers of two ASs, which have Autonomous System Number (ASN) 10 and 20 respectively [8]. The border router of a neighboring AS, to which a BGP session is established and BGP updates are

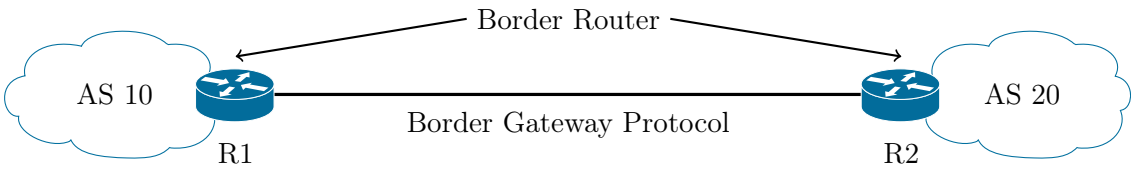


Figure 2.1: Border Gateway Protocol visualization.

```
+-----------------------------------------------------+
|   Withdrawn Routes Length (2 octets)                |
+-----------------------------------------------------+
|   Withdrawn Routes (variable)                       |
+-----------------------------------------------------+
|   Total Path Attribute Length (2 octets)            |
+-----------------------------------------------------+
|   Path Attributes (variable)                        |
+-----------------------------------------------------+
|   Network Layer Reachability Information (variable) |
+-----------------------------------------------------+
```

Figure 2.2: BGP update message format [1].

exchanged with, is called *peer*. In the example above R1 is peer of R2 and vice versa. Both ASs can exchange routes for prefixes using BGP update messages. A *route* is identified by the peer, prefix, and AS path.

There are two types of BGP Updates: announcements and withdrawals. A router sending an announcement shares its best route for a prefix with the neighboring AS. A withdrawal for a prefix informs the neighboring AS that the sender no longer can (or wants to) export a route for the respective prefix. A route contained in a BGP update message always acts as a replacement for the previously sent route regarding one prefix. An announcement following another announcement is an implicit withdrawal, because it overrides the previous route and implicitly withdraws it by providing a new route. An explicit withdrawal invalidates the preceding route and does not provide an alternative.

To able to communicate with the router of a different vendor a standardized message format was specified. A standard BGP update message is shown in Figure 2.2. The second field, *Withdrawn Routes*, contains all prefixes that are withdrawn, *e.g.,* if the update message sent from R1 to R2 contains prefix `1.1.1.0/24` in this field, R1 invalidates any routes for `1.1.1.0/24` that have been sent to R2. The set of announced prefixes is specified in the Network Layer Reachability Information (NLRI) field. The fourth field, *Path Attributes*, contains multiple attributes. Each *Path Attribute* applies to all prefixes present in the NLRI field. Only two attribute fields are relevant for this work:

AS Path Ordered list of ASs through which the route propagated. Rightmost AS is the origin network.

Aggregator ASN and router IP that aggregated routing information present in this message.

Figure 2.3: Route processing within a router.

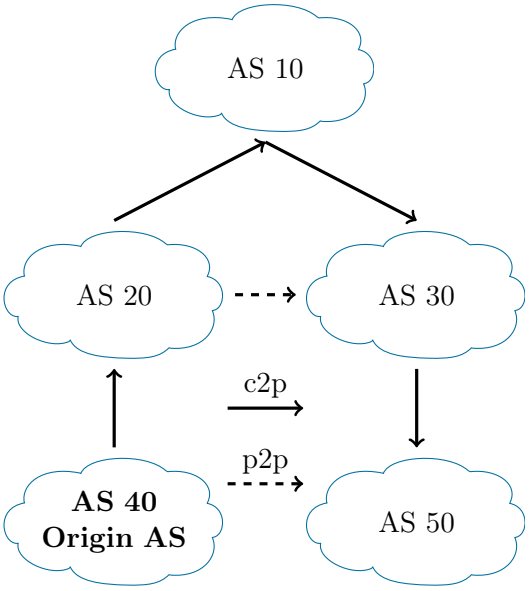
We use this attribute for a different purpose which is described in Section 5.1.

A router maintains three Routing Information Bases (RIBs): *Adj-RIB-In, Loc-RIB,* and *Adj-RIB-Out*. The BGP decision process selects the best route for each prefix to be used for IP packet forwarding and exporting to neighbors. The three processing for routes that have been received from peers are visualized in Figure 2.3. These steps are important to understand, because they directly affect how routes propagate on the Internet.

1. **Calculating Degree of Preference:** Received routes are stored in *Adj-RIB-In*. The router assigns each route a preference degree based on a local policy. Policies are commonly influenced by the relationship between to adjacent ASs, *e.g.,* AS links where traffic costs the least amount of money are preferred.
2. **Route Selection:** Determines the best route (if available) for each prefix, based on the degree of preference. If this value is identical for multiple routes, the following tie-breaking rules will be applied in order:
   a) AS path length
   b) IGP over EGP over Incomplete
   c) *MULTI_EXIT_DISC*
   d) EBGP over IBGP
   e) Lowest Cost to *NEXT_HOP*
   f) Lowest BGP Identifier Value
   g) Lowest Peer IP Address

   The selected routes will be placed in *Loc-RIB* and are used for forwarding packets.
3. **Advertisement to Peers:** Based on a configured policy, routes from *Loc-RIB* will be placed in *Adj-RIB-Out* and announced to peers. Note that only modifications to Adj-RIB-Out will be shared with peers, *i.e.,* if the last sent route does not match the newly elected route in *Adj-RIB-Out*, an update is sent to the peer.

The BGP decision process is performed by each router individually. It finishes by exporting newly learned routes to peers, which perform the same process, apply possibly different policies, and in turn propagate changes in their Adj-RIB-Out to their peers. The applied policies, to select routes for export, are typically based on the relationships between ASs. There are two common types of relationships [9]: peer to peer (p2p) and customer to provider (c2p). The former is the case if both ASs exchange a similar amount of traffic. Therefore, neither pay money to their neighbor. In the latter relationship type the customer pays for the exchanged traffic. The following rule applies in most cases: *All learned routes are propagated to customers and routes learned from customers are propagated to all neighbors.* All other cases usually do not occur, because it neither generates income nor does it benefit the network. After all, most networks exist to generate money directly or indirectly.



Announcement Convergence Process

Step	Updates	
0	AS40: (40)	→ AS20
1	AS20: (20 ← 40)	→ AS30
1	AS20: (20 ← 40)	→ AS10
2	AS10: (10 ← 20 ← 40)	→ AS30
2	AS30: (30 ← 20 ← 40)	→ AS50

Withdrawal Convergence Process

Step	Updates	
3	AS40: (W)	→ AS20
4	AS20: (W)	→ AS30
4	AS20: (W)	→ AS10
5	AS30: (30 ← 10 ← 20 ← 40)	→ AS50
5	AS10: (W)	→ AS30
6	AS30: (W)	→ AS50

Figure 2.4: Convergence process in a complex AS topology with p2p (dashed) and c2p (solid) relationships. The tables on the right side show the updates sent during the convergence process of an announcement and withdrawal originating in AS40. "AS40: (1,2,3) → AS 60" stands for AS40 announcing a route with AS path (1 ← 2 ← 3) to AS60. *W* represents a withdrawal.

2.3 BGP Optimizers to Limit Volatile Routes

Figure 2.4 displays a more complex AS topology compared to Figure 2.1. The real Internet is dynamically changing over time, far complexer, and most importantly the topology one router observes is, in most cases, unique. For the sake of clarity, we omitted the routers in this figure and assume that all ASs only consist of one border router interfacing with the neighboring ASs. Lines between ASs symbolize an established BGP session. The line type indicates the type of relationship, where a solid line stands for c2p and a dashed line represents a p2p relationship. Note that the only p2p relationship in this example is between AS20 and AS30. The time it takes for route update to converge to all routers on the Internet, *convergence time* is one downside of BGP. In the next two paragraphs we will be explaining the *convergence process* of an announcement and a withdrawal. We make the assumptions that (*i*) prefix p belongs to AS40, (*ii*) prefix p has not been announced before, and (*iii*) best path selection is solely based on AS path length.

AS40 wants announces its address space (prefix p) to the Internet via AS20. AS20 receives the first and only route to p and subsequently selects it as the best route for p. As routes from customers are announced to all neighbors, AS20 announces the route with AS path (20 ← 40) to its two neighbors AS30 and AS10. Both ASs now also receive their first route to prefix p and select it as their best route for p. As all known best-routes are announced to customers, both AS10 and AS30 announce their newly received routes for p to their customers AS30 and AS50 respectively. Note that AS30 has now received two routes for prefix p. One route via AS10 with AS path (10 ← 20 ← 40) and one route with AS path

$(20 \leftarrow 40)$ via AS20. The latter route contains a shorter AS path and will therefore remain the best route to prefix p. AS50 now knows a route to p with AS path $(30 \leftarrow 20 \leftarrow 40)$. At this point all AS shared their best routes and the initial update has fully converged. A common synonym for the convergence process is *path hunting*.

AS40 now wants to withdraw its prefix p from the Internet. It subsequently sends a withdraw for p to its upstream AS20. As a result AS20 removes its only route to p and sends a withdraw for p to its neighbors. Recall that AS30 knows two routes for prefix p. One route via AS10 and one route via AS20. As AS30 has only received a withdrawal from AS20 thus announces the alternative route via AS10 to AS50. Now AS10 sends a withdrawal for prefix p, thus withdraws the route that AS30 has just announced to its customer AS50. Consequently, AS30 now also sends a withdrawal for p to AS50.

In Figure 2.4 AS50 receives first an alternative path and then a withdrawal during the convergence process of the withdrawal sent by AS40. This means a single withdraw caused two updates somewhere else in topology. This topology contains 5 ASs, but the real-world Internet consists of over 60,000 ASs with an increasing amount of updates [10]. A common solution is to deploy BGP optimizers to reduce the amount of route updates on the Internet, because each update causes a router to recompute the best route for a given prefix. Therefore BGP needs to be optimized for a complex topology. The BGP optimizers Route Flap Damping (RFD) and Minimum Route Advertisement Interval (MRAI) try to reduce the amount of BGP updates propagating throughout the Internet in order to reduce the CPU load in routers. While they greatly lessen the amount of updates, they also increase the convergence time [2, 11].

2.3.1 Minimum Route Advertisement Interval

Minimum Route Advertisement Interval (MRAI) defines the minimum amount of time that must elapse between two consecutive updates for a specific prefix. It is applied on a per-prefix basis, but can be set on a per-peer basis. MRAI is enabled by default and set to 30 seconds in Cisco routers [12] and is in use today. The actual deployment has not been measured yet and is therefore unknown. MRAI is a rate limiter using a fixed timer, but it does not hinder a router from sending updates at exactly the rate MRAI defines. To accommodate this this problem one could increase the MRAI timer, but that would significantly slow down the convergence process. MRAI suppresses route updates only on a short timescale.

2.3.2 Route Flap Damping

Route Flap Damping (RFD) in turn suppresses prefixes on longer timescale, *i.e.,* short bursts of updates are tolerated, but constantly flapping prefixes will be suppressed. The name for this mechanism is misleading, because RFD does no suppress routes, but peer-prefix pairs. We define a router that has RFD enabled as *RFD router*. ASs which contain at least one RFD router are defined as *RFD ASs*.

For each prefix, an RFD router maintains a penalty value based on the Additive Increase Multiplicative Decrease (AIMD) principle. Based on the type of update the penalty is incremented by 500 or 1000 (not configurable). Withdrawals and, depending on the vendor, also announcements succeeding a withdrawal increment the penalty by 1000 (see Table 2.1).

RFD parameter	Cisco	Juniper	RFC 7454
Withdrawal penalty	1000	1000	1000
Readvertisement penalty	0	1000	0/1000
Attributes change penalty	500	500	500
Suppress-threshold	**2000**	**3000**	**6000**
Half-life (min)	15	15	15
Reuse-threshold	750	750	750
Max suppress time (min)	60	60	60

Table 2.1: RFD default parameters [2, 3, 4].

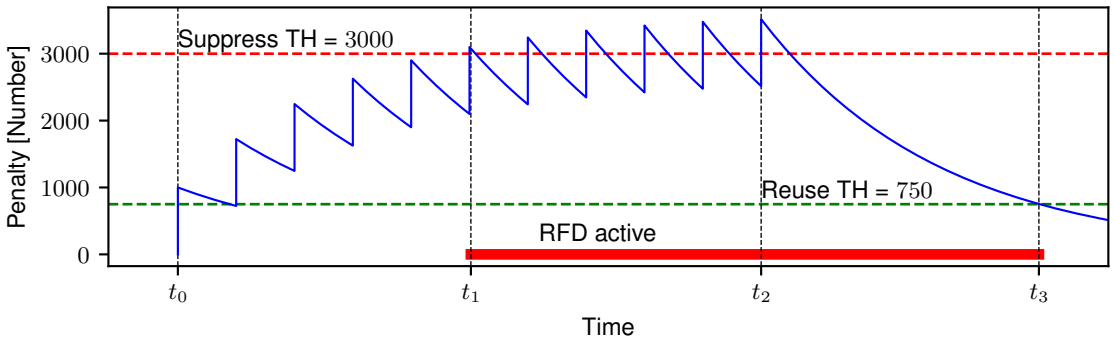


Figure 2.5: Visualization of RFD penalty with thresholds (TH).

An attribute change, *e.g.,* AS path change, increases the penalty by 500. The penalty decreases exponentially over time, based on a configurable *half-life* parameter. Prefixes that surpass the *suppress-threshold* will not be advertised to neighbors until the respective value is smaller than the *reuse-threshold*. The duration a prefix is suppressed after being stable again is limited by the *max-suppress-time* parameter.

Figure 2.5 visualizes a scenario how the penalty for a prefix can behave during route oscillation using Cisco default parameters (2.1). The initial penalty at t_0 is 0, *i.e.,* the router has never received an update for this prefix or the last update has been received too far in the past. The router starts receiving updates for the prefix periodically, thus increases the penalty each time. Between updates, the penalty is decreased based on the *half-life* parameter. The penalty surpasses the *suppress-threshold* at t_1 and is subsequently withdrawn. The prefix stops oscillating at t_2 and therefore the penalty goes below the *reuse-threshold* at t_3. If the router has not received any other updates for the prefix in the mean time, the previously damped prefix will be announced again.

The usage of the AIMD principle in RFD has two key implications. Prefixes which flap once or twice in a short period of time and then stop flapping are not suppressed. It is MRAI's purpose to suppress these short bursts of updates. The second implication is that prefixes which flap slowly over an extended period will be suppressed constantly, because the penalty will not have chance to go below the *reuse-threshold*.

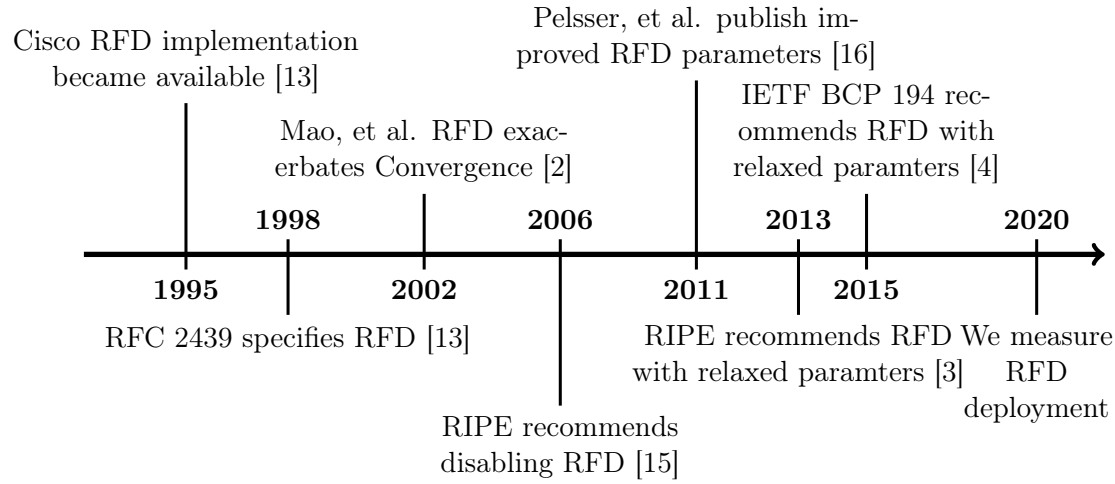


Figure 2.6: History of RFD parameter recommendations.

There is not a definition that defines when a prefix is flapping to fast for too long, thus choosing the correct RFD parameters has been a controversial topic in the past. Major router vendors such as Cisco started implementing RFD in the early '90s, because routers struggled to keep up with the amount of BGP Updates [13]. Recall that each Update triggers the BGP decision process again. It is unclear whether the performance argument still holds true for today's hardware. RIPE published the first recommendation on RFD parameters in February 1998 [14]. The main problem was already apparent back then: RFD can dampen normal behaving prefixes during the route convergence process. Therefore the recommendation was that routes shall not be damped until the 4th flap. The same parameters were recommended in the RFD RFC published shortly after, in November 1998 [13]. These recommendations were based on their observations at the time. As one might expect, the Internet topology kept getting richer and more complex as time went on. Mao, et al. published in 2002 that a single withdrawal could cause 4 or more updates somewhere distant in topology [2]. This means that RFD, with recommended parameters, also dampens normal behaving routes, as sending one withdrawal is not uncommon. As a result RIPE recommended to disable RFD in all routers in 2006 [15]. In 2011 Pelsser, et al. published that, with adjusted parameters, RFD could still be used without causing harm [16]. The main idea was to only dampen "heavy hitters" by increasing the *suppress-threshold*. Based on this paper, RIPE and the Internet Engineering Task Force (IETF) recommended new parameter sets in 2013 and 2015 [3, 4]. An overview of the most important events is visualized in Figure 2.6.

2.4 Securing the Control Plane

When BGP was first introduced, peering relationships were based on trust. As the Internet grew and its overlaying services gained popularity trust is not a viable solution anymore. *Prefix hijacks* now very frequently occur and lack a proper mitigation technique [17]. To create a safer Internet, some routers perform route origin validation, i.e., check if the last

AS in the AS path (origin AS) is authorized to announce the prefix present in a received route.

Initially *Route Objects* maintained in the Internet Routing Registry (IRR) were sometimes used to validate the origin of the route. Route Objects contain, among other information, prefix-ASN pairs. Unfortunately these are not cryptographically signed and could be tampered with.

A seconds, more promising solution it the Resource Public Key Infrastructure (RPKI), which was introduced to solve this problem. An RPKI certificate purely proofs the ownership of an IP prefix or an ASN. A ROA (Route Origin Authorization) proofs to everyone that an AS is allowed to announce a specific prefix. These ROAs are signed by the respective RPKI certificate. ASs can perform BGP update filtering by checking whether a valid ROA (or Route Object) exists for the origin network and the prefix it announces. This means when announcing a prefix, the network operator needs to make sure both Route Objects and RPKI ROAs are properly created. Otherwise the network may experience poor visibility, i.e., only few ASs know routes for the prefix, because they were being filtered on the way. Unfortunately, not every network has created route objects as of yet, but the number is growing [18].

2.5 Measurement Infrastructure

To measure understand the control plane of the Internet, a variety of concepts have been introduced by the network measurements community. Two of them, namely BGP Beacons and route collectors, are relevant for this project and explained in the next two paragraphs.

BGP Beacons [19] are a sequence of announcements and withdrawals for one specific prefix. A *Beacon pattern*, sent by a *Beacon router*, corresponds to a publicly available schedule. IP addresses within prefixes announced by a Beacon are not in commercial use. Beacons are commonly utilized in active controlled BGP measurements [20, 2, 21, 19]. RIPE RIS provides BGP Beacons with an update interval of 2 hours. We are using faster BGP Beacons in this study, because an update interval of 2 hours does not trigger RFD. Later, we compare the RIPE RIS Beacons to our infrastructure for validation (Section 5.3).

To collect BGP updates from different sites on the Internet route collectors are commonly used in active and passive control plane measurements [20, 21, 22]. Route collectors are peering with multiple routers and store all received BGP Updates. Route Collectors peers are called Vantage Points (VPs) and export changes in their Adj-RIB-Out to the collector [23]. In general there two types of VPs: *full-feed* and *partial-feed*. Full-feed VPs export any changes to their Adj-RIB-Out to a route collector, whereas partial-feed VPs export, as the name suggests, only parts of the known address space to a collector. There are multiple public route collector projects, which provide publicly accessible servers from which one can download all available routing data. Two different types of routing data are usually provided: either BGP updates received or the state of the RIB at a given point in time. In this study we are using update dumps from the route collector projects RIPE RIS, Route Views, and Isolario.

CHAPTER 3

Related Work

Negative Impact on Reachability. RFD was initially designed to reduce CPU load in routers, which were not able to handle the amount of new updates triggering the best path selection each time. Mao *et al.* [2] found that RFD can in fact increase the convergence time of routes, which are stable most of the time, *e.g.,* routes that are withdrawn and re-advertised once a week. If a router r withdraws a route, and its peers have selected the route as best route for the given prefix can either need to find a new route. If available an alternative route is advertised, and if not, then the entire prefix is withdrawn. In a richly connected Internet, it is common that many routers have multiple routes at hand for a prefix. This results in all these alternative routes being announced one after another causing multiple, arguably unnecessary, announcements multiple AS hops away from the original withdraw. Mao *et al.* discovered that RFD configured with vendor default values can withdraw the entire prefix during the convergence process in specific topologies, *e.g.,* a clique of 5 nodes. Intuitively one would think this is not an issue because the prefix was initially withdrawn anyway. But if router r announces a route for the same prefix again, then the convergence will be significantly slower, because some routers are still suppressing the prefix, thus not propagating the new route. Mao *et al.* also explains that this phenomenon can occur when sending a single announcement, though this relies on precise timing hence it is much less likely to happen on the real Internet.

Using a real-world sample topology they also tested whether withdrawal triggered suppression can happen in the real world. In the base case, they chose to use Cisco default values and set the MRAI to 30 seconds, which is also default in at least on vendor software. Here, they found that for 22% of the simulations withdrawal triggered suppression occurred. With MRAI set to 5 seconds suppression occurred for 52% of the simulations, showing that the MRAI value influences how well RFD performs. They also note that no damping occurs in the sample topology with a less aggressive RFD configuration, where the penalty increase for an attribute change is set to 250, which is half of the default.

Recommended Parameters. Pelsser *et al.* pick up this observation and perform more detailed analysis to make precise recommendations for parameters to make RFD usable again [16]. Their goal was to filter out "elephants" without causing collateral damage, *i.e.,* reachability problems for innocent networks. They found that for a week in 2010 that

3% of prefixes were responsible for 36% of BGP updates. We reproduced this study for a day in March 2020 using the public router collector projects Routeviews and RIPE RIS and found that 3% of prefixes were responsible for 65% of BGP updates. This means very few prefixes are responsible for for the majority of BGP updates. To find out whether the RFD default configurations were suppressing the correct prefixes, they setup a mirror which received updates from both a large IXP and a Tier 1 provider. This mirror exported all updates to a router which has RFD enabled and does not suppress routes, but records the penalty for it would assign to each prefix instead. Then, they checked for which share of prefixes the penalty rose above specific thresholds. Surprisingly 14% of prefixes reached a accumulated penalty greater than 2000, which is the default Cisco suppress threshold. Subsequently, this router would have suppressed every second update on average. 4.2% and 0.63% of prefixes rose above 4000 and 12000 respectively. Finally, they came to the conclusion that 12000, which suppresses 11.26% of updates, shall be the recommended threshold.

Route Flap Damping Measurements. Mao *et al.* coined the term *BGP Beacon* in 2003, which they define as a publicly documented prefix having global visibility and a published schedule for announcements and withdrawals [19]. In their study they mentioned that it is very important to understand how likely RFD occurs in today's Internet as RFD can be triggered from a single router reboot. To measure this, they setup a Beacon infrastructure similar to the RIPE Beacons with announcements and withdrawals every two hours for testing prefixes. Based on Routeviews data they observed that at least 5% percent of Beacon signals were suppressed across all VPs. Note, that in their measurement individual withdrawals or announcements caused suppression in routers. This happened likely due to the update amplifying effects of the convergence process. They also found that withdrawals were dampened roughly twice as often, likely because announcement converge faster and do not cause as many updates on the way. Mao *et al.* we first and only to attempt to measure RFD deployment, though in a very limited way as their study was mostly focussed around BGP Beacons. We attempt in our study to deepen our understanding about RFD deployment almost two decades later.

Locating ASs Responsible for Route Changes. In the past, there have been multiple studies which attempted to pinpoint the origin of a routing change. Feldmann *et al.* [24] attempted to find the origin of a set of BGP updates from multiple route collector peers based on multiple heuristics. They proposed and used a very generic approach targeted at any routing change. Based on which VP received updates for a given routing event they can infer the root cause down to a set of AS links which are most likely responsible for the route change. Caesar *et al.* [25] we conducting a similar measurement while limiting themselves to reasonably stable prefixes, *i.e.,* prefixes that are not constantly flapping. They are also detecting the type of high-level event that triggered the series of updates, *e.g.,* prefix withdrawal from origin network, route flap, or link failure. The key difference to our study is that both approaches try to measure reason for a set of BGP updates. Although they are pinpointing ASs with specific properties, to measure RFD we need to find the root cause for *missing or delayed* updates. The events they measure are *responsible* for RFD.

A study that comes much closer is a prototype by Chang *et al.* [26], who try to determine the reason for a missing route, *i.e.,* a prefix is unreachable small set of vantage points. Unfortunately their approach does not differentiate between the type of problem that caused

unreachability, but they stated that misconfigured RFD was one of possible reasons.

BGP Zombies and RFD. Fontugne *et al.* measured in 2018 how often *BGP Zombies* appear on the real-world Internet over a year and half [20]. BGP Zombies, are routes that have been withdrawn from the origin network, but still exist in a RIB. Using the RIPE BGP Beacons, which announce and withdraw testing prefixes every two hours, they label and RIB entry as Zombie that has been active 90 minutes after the prefix was withdrawn from the Beacon router. They discovered that BGP Zombies occur daily and affect on average 10% of measured ASs for IPv4. They suspect that RFD is one of the key reasons for BGP Zombies. Later, we show that the maximum time a prefix is suppressed (without new updates) is very rarely longer then 60 minutes. This is much lower than the 90 minute timeout that is used to detect BGP Zombies. Therefore, we believe that RFD is not the reason for BGP Zombies.

CHAPTER 4

Methodology

This study is a controlled, active measurement on the control plane and attempts to measure global deployment of Route Flap Damping (RFD). Even though RFD is configured and applied at router-level [1], we chose to measure deployment on AS-level, because BGP control plane data, which we use to infer RFD usage, is at AS-level granularity (AS path). Methods using active data plane measurements, *e.g.,* traceroute, exist to infer router-level paths, but tend to be inaccurate. Another observation during the study was that RFD is generally configured based on the neighboring AS instead of the neighboring router. Therefore, we decided against inaccurate router-level inference and focussed on RFD deployment on AS-level.

RFD suppresses unstable prefixes, *i.e.,* prefixes that are announced and withdrawn too quickly over an extended period, as explained in Section 2.3.2. The threshold that defines when a prefix is too unstable is set using the RFD parameters. To simulate unstable routes and trigger RFD, we are injecting different sequences of BGP updates for a set of testing prefixes at multiple sites on the Internet using BGP Beacons (Section 4.1). We send multiple BGP Beacons in different locations to increase the number of paths between Beacon routers and VPs (Section 5.2). Each Beacon router oscillates a set of Beacon prefixes, which are simulating unstable prefixes and each target a different RFD configuration. An additional prefix, sent from each Beacon router, acts as an anchor prefix with an *update interval* of two hours and is not supposed to trigger RFD in any router, where the update interval is defined as the duration between consecutive announcements and withdrawals. Section 4.1 goes into detail about the Beacon pattern and the necessary update interval.

RFD-enabled routers suppressing our Beacon prefixes change how the Beacon signal propagates through the Internet. To pick up these modified signals we use the three largest public route collector projects: RIPE RIS, Routeviews, and Isolario. Each VP exports a their best routes for our prefixes to the route collector. In Section 4.2 we analyze the observed update pattern for each Beacon prefix and derive RFD usage on each AS path based on the difference between the Beacon pattern and the observed update pattern at a given VP.

Knowing that one, but not which AS, on an AS path causes the Beacon pattern to change, *i.e.,* uses RFD, is not sufficient. In Section 4.3 we introduce three heuristics to infer the RFD AS on a given path and we later summarize these results to infer RFD deployment

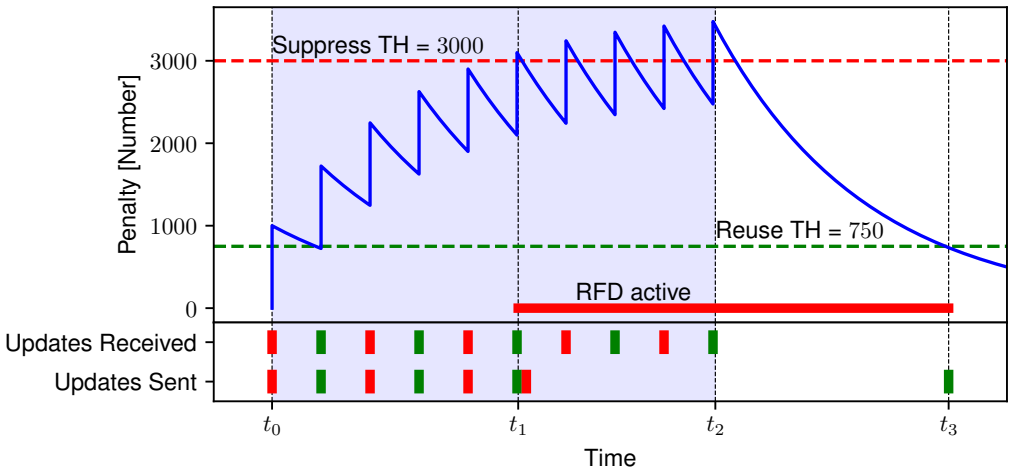


Figure 4.1: Perspective from an RFD router: penalty for a Beacon prefix and the updates, that is announcement (green) and withdrawal (red), the router receives and then sends. Horizontal lines represent suppress and reuse-threshold (TH).

for each AS. The key problem of this approach is that we assume that all ASs are binary, *i.e.,* either use RFD for all neighbors or do not use RFD at all. This assumption does hold for all ASs in the real-world. As a result the heuristics sometimes infer RFD deployment incorrectly. Therefore, in Section 4.4, we apply the same heuristics to ordered AS links for which the above assumption holds true without known exceptions.

4.1 Route Oscillation Generation

We utilize multiple BGP Beacons, which announce and withdraw prefixes in quick succession to trigger RFD. We call the schedule at which a given Beacon oscillates *Beacon pattern*. One might argue that the Beacon pattern does not reflect real-world instabilities, but the goal in this study is not to simulate instabilities, but to precisely measure RFD.

The RFD Signature. Switching between announcements and withdrawals without breaks at a specific frequency would cause RFD ASs to constantly dampen our prefixes. This means after RFD has been triggered once in an AS, we would never observe any routes with AS paths containing this AS at VPs. After a few hours we would simply never receive routes via damping ASs. A solution to this problem is to oscillate prefixes only at specific times and thus giving RFD routers a chance to reset the penalty for our prefixes, which would cause the router to use and announce our prefixes again. The Beacons are always in one of the following two phases:

- **Burst**: Sequence of alternating announcements and withdrawals starting with a withdrawal and ending with an announcement.
- **Break**: No updates.

This Beacon pattern attempts to force a specific RFD penalty behaviour resulting in a very recognizable update pattern sent by RFD routers. This update pattern is later, in

Section 4.2, used to label paths with RFD true. An example is visualized in Figure 4.1, which shows the penalty curve for a Beacon prefix. The axis *Updates Received* shows when the RFD router receives updates for the prefix and the penalty curve (blue) displays its resulting penalty. Finally, the axis *Updates Sent* shows when the prefix is announced or withdrawn from peers.

The light-blue area (t_0 - t_2) denotes the Burst phase and the white area (t_2 - t_0) represents the Break phase, where no updates are sent. The router starts dampening the Beacon prefix at t_1 and therefore sends a withdrawal to its peers. Then the Burst stops and the router no longer receives updates the prefix at t_2. The assigned penalty now decreases and subsequently falls below the *reuse-threshold* at t_3. As the last received update for the Beacon prefix is an announcement (t_2) and the prefix is considered usable again (after t_3), the router *re-advertises* the prefix to its peers.

Effects of MRAI. RFD is not the only BGP mechanism that suppresses prefix updates. In Section 2.3.1 we explained that MRAI also limits BGP updates. Therefore, it is worth investigating whether MRAI can cause a similar pattern to the above and thus influence our results. MRAI, in contrast to RFD, delays updates at most n seconds, where n is a configurable constant. This means any update received by a router which has MRAI configured will be exported to peers at most n seconds after it has been received. RFD on the other hand *(i)* withdraws prefixes instead of delaying their updates and *(ii)* the time until the *re-advertisement* is calculated dynamically during runtime and is not configured statically. Therefore, the update pattern explained above cannot be caused by any MRAI configuration. MRAI would instead lead to a static re-advertisement independent of specific Beacon pattern patterns.

Accuracy Limitations. In Section 2.2, we explained that a single update can cause multiple updates in different places on the Internet. This also means that the update interval of the sent update interval is not necessarily the update interval that arrives at a router multiple hops away from the Beacon router. A single withdraw sent from the Beacon router may cause a router to increase the penalty for a route two or more times. This means its non-trivial to target a specific RFD configuration, as we cannot control the route convergence process.

4.2 Labeling Paths

We send updates from Beacons routers and receive updates at VPs. The update pattern needs to be interpreted in order to derive whether there is an RFD AS between the Beacon and the VP. Figure 4.2 shows an example how updates can propagate differently between a Beacon router and a VP. The leftmost AS contains the Beacon and the rightmost AS contains the VP router. We assume that AS20 is the only AS using RFD in this graph. In this example the updates travel on two different paths from the Beacon AS to the VP AS: (VP←20←10←Beacon) and (VP←30←10←Beacon). Assuming this setup, we cannot simply interpret all the updates received for a Beacon prefix, because then, the Beacon pattern that has been modified by AS20 and the unmodified, clean Beacon pattern via AS30 would be intermixed, meaning we would not be able to clearly see the RFD update pattern. Therefore, we need to separate received updates by their AS path. Unfortunately withdrawals do

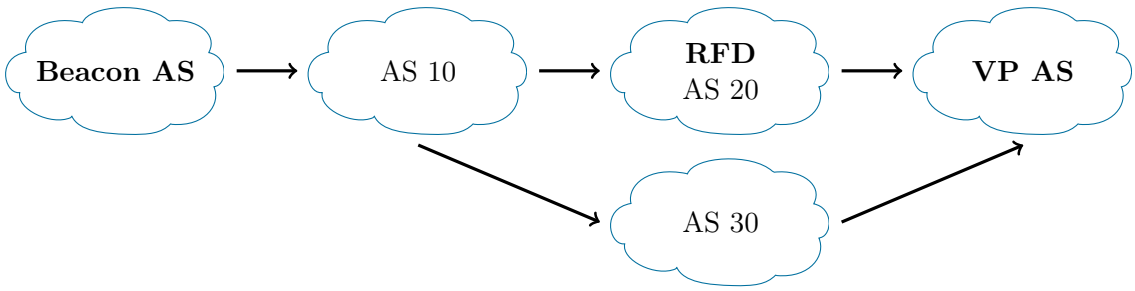


Figure 4.2: Update propagation along two paths from Beacon AS to VP AS.

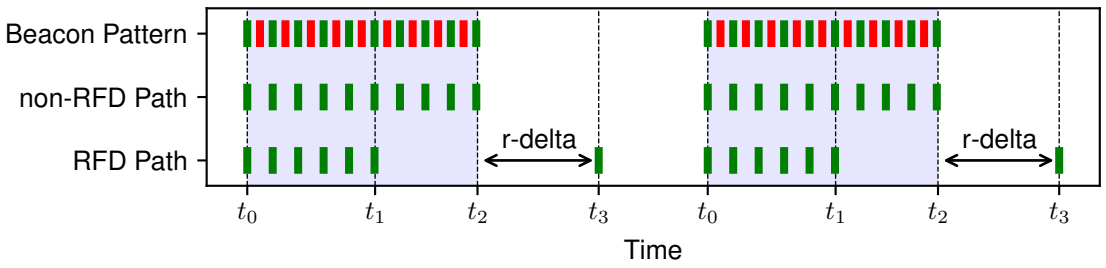


Figure 4.3: Beacon pattern, RFD path, and non-RFD path for two Burst-Break pairs, where *r-delta* is the time delta between the end of the Burst and the re-advertisement.

not contain an AS path attribute, which means only the announcements in the observed update pattern can be used for interpretation.

We interpret the announcement pattern for each AS-path to a Beacon, at each VP individually and refer to these simply as *path*. Paths are labeled with either *RFD path* or *non-RFD path* by analyzing the announcement pattern. We are also analyzing each Burst-Break pair separately, because the Break resets the penalty in RFD routers for our prefixes, which means that we have the same state in all routers for our prefixes at the beginning of each Burst. For each Burst-Break we check whether all of the following three rules apply:

1. The re-advertisement was received in the Break.
2. The re-advertisement is sufficiently delayed, *i.e.,* much greater than the usual propagation time.
3. At least one update during the Burst has not been received.

To define the minimum propagation time for a re-advertisement, both the normal propagation delay for our Beacons and common MRAI configurations need to be considered. The propagation delay of our anchor prefixes is at most 1 minute (see Section 5.3). At the time of this writing, there are no studies measuring the values that are used to configure MRAI on the Internet, but there is at least one vendor defaulting MRAI to 30 seconds. Considering Cisco RFD default parameters, a prefix is suppressed for at least 21 minutes, for Juniper even longer. Given these distinct timescales, we find that setting the minimum propagation time for the re-advertisements to 5 minutes clearly separates the signals.

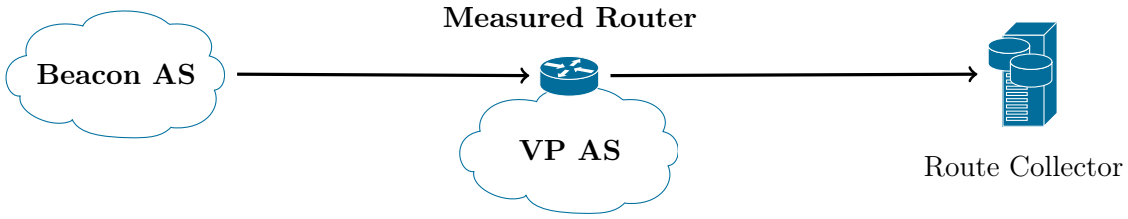


Figure 4.4: Ideal measurement setup.

After analyzing all Burst-Break pairs of each path, we have a set of RFD paths and a set of non-RFD paths. In practise, there is a third group: paths for which we do not receive enough updates at the respective VP and therefore cannot derive RFD usage (Section 6.1). This group of paths is ignored for further analysis. Figure 4.3 [1] shows two consecutive Bursts and Breaks for two paths. The announcement pattern on the middle axis indicates that RFD is not used on the path, because the announcement pattern matches the Beacon pattern of the Beacon. The announcement pattern on the lower path matches all the above RFD rules, hence it is labeled as RFD path. VPs only export their best paths to route collectors, which means we would see only one of the paths announcement pattern.

4.3 Pinpointing ASs

The overall goal is to be able to tell exactly which ASs are using RFD on the Internet. In the previous section we labeled paths with *RFD path* or *non-RFD path*. For damped paths we know that at least one AS uses RFD. Though if an AS is on an RFD path, this does not necessarily mean that it uses RFD. Any other AS on the path could also be the reason for the RFD announcement pattern.

In an ideal setup, the VP would directly peer with the Beacon router as illustrated in Figure 4.4. This way, if the announcement pattern for the path (VP←Beacon) matches the RFD requirements, we know that the VP AS is the damper because it is, beside the Beacon AS, the only AS on the path. This setup would is not possible to achieve in the real-world, because of the overhead that would come with setting up a Beacon and two peering sessions between the Beacon AS, VP AS and route collectors as shown in Figure 4.4 for all 60,000 ASs. Therefore, we need an approach which can identify individual ASs without relying on the ideal measurement setup.

We introduce three metrics which assign each AS a score between 0.0 and 1.0. If the average score of these metrics for a given AS is greater than a threshold, then the AS is declared *RFD AS*. We compute the threshold based on the output of an evaluation function that determines how good a given threshold performs (Section 6.3.1). In the following sections we give a detailed description of each metric.

[1] Note that the announcement pattern shown in Figure 4.3 is unrelated to the topology in Figure 4.2.

4.3.1 RFD Path Ratio

The *RFD Path Ratio* is the simplest metric. Given the set of RFD and non-RFD paths, this metric is the ratio between the number of RFD paths and the total number of paths the respective AS appeared on, *i.e.,* relative occurrence. We compute for each AS:

$$M_1(\text{AS}) = \frac{\#\text{RFD paths of AS}}{(\#\text{RFD paths of AS} + \#\text{non-RFD paths of AS})}$$

Assuming the example in Figure 4.2 with paths (Beacon←10←20←VP) and (Beacon←10←30←VP) this metric would assign ASs 10, 20, and 30 the scores 0.5, 1.0, and 0.0, respectively. This means we can be sure that AS20 uses RFD, we are uncertain about AS10, and we were not able to detect RFD in AS30.

4.3.2 Inferring RFD ASs Based on Alternative Paths

If RFD is triggered in a router then the prefix will be withdrawn from its neighbors. Neighbors that have more than one path for the prefix in their RIB will announce *alternative paths* to their neighbors. We rely on these paths to identify the damping AS on damped paths. For that we need to (*i*) define and detect alternative paths and (*ii*) infer the RFD AS based on the alternative paths.

Alternative path detection. We assume the topology in Figure 4.5 and for simplicity, all ASs consist of only one router. We also assume that the VP AS selects best paths solely based on the length, thus prefers (Beacon←10←20←VP) over (Beacon←10←40←30←VP). The Beacon AS sends a Burst-Break Beacon pattern that will trigger RFD in AS20 during the Burst. When AS20 starts suppressing the Beacon prefix then the VP AS receives a withdrawal. The duration from the beginning of the Burst until this withdrawal is called *time-until-damp*. Since AS10, AS30, and AS40 do not use RFD, the VP AS still receives announcements via AS30 until the end of the Burst. The VP AS selects (Beacon←10←40←30←VP) as new best, *alternative path* for the rest of the Burst. In general, we label any path as an alternative path for an RFD path if it (*i*) matches Beacon and VP AS of the RFD path and (*ii*) does not occur less frequently in the period after the time-until-damp compared to before. For each damped path we determine set of alternative paths, which can also be empty, *e.g.,* if the VP deploys RFD. Depending on whether this set is empty we perform a different analysis to infer the RFD AS on the RFD path.

Non-empty set of alternative paths. With a non-empty set of alternative paths A (where each path is a_i) for an RFD path r, we compute for each a_i the set of ASs C_i, which consists of all ASs that are on the RFD path r but not the alternative path a_i. In the example in Figure 4.5 the RFD path is r = (Beacon←10←20←VP) and the set of alternative paths contains only one path: a_1 = (Beacon←10←40←30←VP). The resulting set $C_1 = r - a_1 = \{20\}$. In different topologies C_i could also be an empty set or a set with more than one AS. To compute a score for each AS, we count how often each AS occurs in all sets C_i normalized by the size of C_i. This assigns each AS on the RFD path r a score between 0.0 and 1.0. In the above example AS20 receives a score of 1 and all other ASs

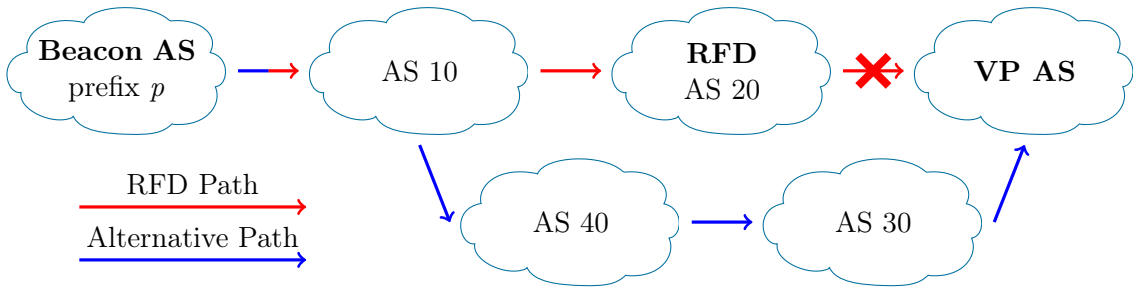


Figure 4.5: RFD path and alternative path for prefix p example.

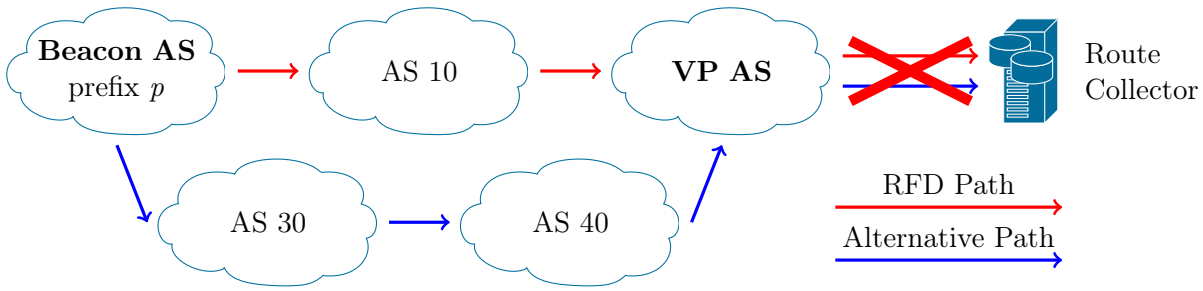
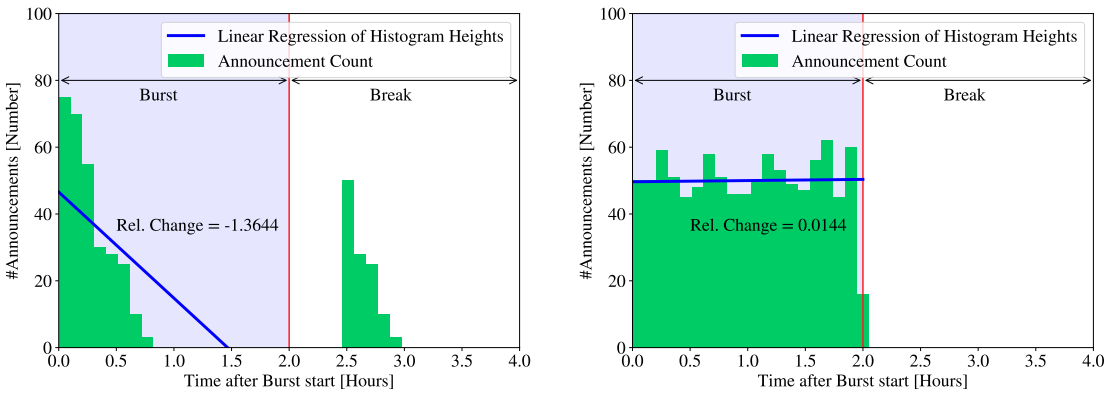


Figure 4.6: RFD AS is the VP AS, because the route collector does not receive any alternative paths.

received score 0 as they all occur on alternative paths.

Empty set of alternative paths. We apply a different method if the set of alternative paths A is empty for an RFD path r. We create a directed graph for each VP AS, which consists of the collected paths to all Beacon ASs. Figure 4.6 shows a simplified example with only one Beacon and two possible paths from the Beacon AS to the VP AS. In practise, this graph would consist of multiple Beacon ASs and more paths between each Beacon AS and VP AS. Recall that each AS on the RFD path, except the Beacon AS, could potentially be the RFD AS. For Figure 4.6 the candidates would be AS10 and the VP AS. As a result of the set of alternative paths being empty we know that there cannot be any path in the graph between the Beacon AS and the route collector if we remove the RFD AS. If AS10 is removed from the graph there is still a path between the Beacon AS and the route collector: (VP←40←30←Beacon). This means the VP AS should have exported this alternative path to the route collector. This is not the case and therefore we know that AS10 cannot be the RFD AS. If instead we remove the VP AS from the graph we find that that there is no path between the route collector and the Beacon AS anymore. This matches our observation that we did not receive alternative paths at the route collector, which means the VP AS is the RFD AS in this case. In general, all ASs on the RFD path for which the above holds true receive score 1. In the above example we would only assign VP AS a score of 1.

The preceding two paragraphs explained how the ASs on each RFD path receive a score. ASs being present on multiple damped paths can have multiple scores. To get the final individual score for an AS we compute the sum of all RFD path specific scores for the AS normalized by the number of RFD paths the respective AS is present on.



(a) RFD AS.

(b) non-RFD AS.

Figure 4.7: Typical distribution of announcements during a Burst-Break pair for an RFD AS and a non-RFD AS.

4.3.3 Announcement Distribution across Bursts

The third metric is very similar to the first method. But instead of analyzing the announcement pattern of individual paths, this method attempts classify the behaviour of each ASs across the Burst from the combined view of all VPs. At beginning of the Burst some VPs may know routes for the Beacon prefixes with AS paths including RFD AS X, because the penalty in all routers should be reset at the beginning of the Burst. One would expect that at the end of the Burst, non of the vantage points should know routes with AS paths that include RFD ASs. In the real-world, updates do not always reliable converge [20], hence a reliable heuristic is needed. This is especially true for large transit ASs that are part of many best path. In general, RFD ASs should occur less often on exported AS paths towards the end of Burst compared to the beginning of the Burst.

Based on this assumption we infer individual RFD ASs based on the distribution of announcements, which contain the AS, during each Burst. Figure 4.7 visualizes how an average case for an RFD AS looks versus a non-RFD AS. The light-blue and white areas, separated by a vertical red line, indicate the Burst and Break phase, respectively. Both plots show a histogram where received announcements are grouped in 40 time intervals. The blue line is the linear regression function $A(t)$ of histogram heights in the Burst, where t is the time difference to the Burst start.

This behaviour needs to be translated into a score between 0.0 and 1.0. We cannot simply use the slope of the linear regression function to derive the score, as a change from 5 to 0 announcements across the Burst is more relevant than a change from 1000 to 995 announcements. Both have a delta of 5 across the Burst length, thus the same slope, meaning the slope is not a good measure. Instead, we compute the relative change across the Burst, which is computed as follows: $relativeChange = \frac{A(\text{Burst End}) - A(\text{Burst Start})}{A(\text{Burst Start})}$. There are edge cases where the relative change is negative but the slope is positive. To account for these cases we take both the slope and the relative change into account when assigning a score to each AS:

$$M_3(relativeChange, slope) = \begin{cases} 0, & \text{if } relativeChange > 0 \\ 0, & \text{if } slope > 0 \\ 1, & \text{if } |relativeChange| > |upperBound| \\ \frac{|relativeChange|}{|upperBound|} & \text{otherwise} \end{cases}$$

The *upper bound* for the relative change is the 5-th percentile of all of negative relative change values.

4.4 Identifying Inconsistent ASs

In conversations with network operators we found that some ASs use RFD inconsistently in routers throughout the AS. We noticed that, without exception, RFD is configured on AS-level rather than router-level. An AS may chose to only apply RFD to one particular neighbor because it expects or has experienced too many routing updates originating in that AS. The three explained metrics work best if ASs use RFD consistently. Since this is not the case RFD deployment needs to be analysed on an AS link basis as well. To accomplish this we applied the three existing methodologies with minor modifications to ordered AS links as well (path $(10 \leftarrow 20 \leftarrow 30)$ consists of links (10,20) and (20,30)). The AS links need to be ordered because RFD usage does not occur symmetrically, *i.e.,* AS X may apply RFD to all updates coming from AS Y, but AS Y may not be applying RFD to any of their neighbors.

CHAPTER 5

Setup

The effort it takes to setup the infrastructure for a distributed measurement is often underestimated. It takes careful design decisions and most importantly validation. In this chapter we first describe the Beacon router configuration in detail with regard update intervals, phase lengths, and the timestamp encoding methodology (Section 5.1). Then, we justify the high number of Beacons required for this study as well as why using all available major route collector projects is pivotal (Section 5.2). Finally, we compare our setup to a similar and also professionally maintained legacy Beacon infrastructure by RIPE where find that convergence of updates and visibility from vantage points is largely identical (Section 5.3).

5.1 Infrastructure Configuration

Our study requires multiple Beacons spread across the world in order to receive numerous paths through links and ASs. We sent 7 Beacons from Europe, South and North America, and Asia. Their locations are listed in Table 5.1. All Beacons are located at a maximum of 2 hops away from a Tier-1 provider. We verified that the respective peer AS of our Beacon routers does not use RFD and therefore does not influence our measurements. IP prefixes and one AS number (AS3130) was provided by Randy Bush and the AS numbers AS58360 to AS58365 were provided by RIPE.

Location	Prefixes	Beacon ASN	Peer ASN
Brasil	`147.28.{48,49,50,51}.0/24`	58364	22548
Denmark	`45.132.{188,189,190,191}.0/24`	58360	31027
Germany	`147.28.{52,53,54,55}.0/24`	58365	5539
Japan	`147.28.{32,33,34,35}.0/24`	58361	2497
USA	`147.28.{36,37,38,39}.0/24`	3130	3130(IGP)
South Africa	`147.28.{44,45,46,47}.0/24`	58363	37100
Thailand	`147.28.{40,41,42,43}.0/24`	58362	63528,63529

Table 5.1: Beacon locations.

Prefix Set	Burst [h]		Break [h]		Interval [min]	
	Nov 2019	Mar 2020	Nov 2019	Mar 2020	Nov 2019	Mar 2020
45.132.188.0/24, 147.28.{32,36,40,44,48,52}.0/24	2	2	2	6	5	1
45.132.189.0/24, 147.28.{33,37,41,45,49,53}.0/24	2	2	2	6	10	2
45.132.190.0/24, 147.28.{34,38,42,46,50,54}.0/24	2	2	2	6	15	3
Anchor: 45.132.191.0/24, 147.28.{35,39,43,47,51,55}.0/24	-	-	-	-	120	120

Table 5.2: Beacon update intervals.

Throughout this study we announced at each Beacon router 4 different prefixes: one anchor prefixes and three prefixes oscillating with different update intervals. All Beacon routers announced its 4 prefixes on the same schedule. In both phases we used anchor prefixes which were announced and withdrawn every 2 hours, *i.e.,* an update interval of 2 hours, identical to the RIPE Beacons, to measure the typical propagation time. To prevent filtering of the prefixes in case of route origin validation deployment, we configured the corresponding route object entries in both the Internet routing registry and the RPKI for all prefixes.

We did not expect RFD configurations more strict than the vendor default values, which already suppress 14% of all prefixes [16]. A Juniper or Cisco router would start damping a prefix that flaps at least every 9 or 8 minutes respectively (Table 2.1). To confirm this, we configured our Beacons with an update interval of 15, 30, and 60 minutes in August 2019. We observed measurable RFD for the fastest Beacon prefix (15 minutes).

After preliminary tests, we conducted two measurement campaigns. In March 2020, we chose 1, 2, and 3 minutes as update intervals during Bursts of two hours, because an update interval of 2 minutes would trigger RFD with the recommended parameters [3, 4]. We set the Break duration to 6 hours to account for very slowly decaying RFD penalties. If a router is configured such that the penalty does not decay during the Break, then the updates from next Burst will increase the penalty again, causing the router to suppress the prefix indefinitely. In April 2020, we chose 5, 10, and 15 minutes as update intervals to cope with RFD parameters that differ more significantly from recommended values—either because vendors ship deprecated default configurations, or manual adjustment by operators. We configure the Break to 2 hours, because the *max-suppress-time* is by default 1 hour and we did not observe suppress phases longer than 1 hour in the Break in March. The Burst length was still 2 hours.

Timestamp encoding. When receiving announcements for the Beacon prefixes in an arbitrary location on the Internet, we want to be able to tell which exact Beacon event caused this announcement. Therefore we decided to encode the current timestamp in BGP updates sent by our Beacons. To accomplish this we required a transitive path attribute field which will not be modified. We took an approach similar to the implementation of the RIPE

Beacons. RIPE uses the *aggregator attribute* to encode both the time of the announcement and a Beacon ID. The aggregator attribute consists of an IP address and an ASN. Contrary to the RIPE Beacon setup [27], we did not encode a Beacon ID in the ASN field, because we could identify each Beacon simply based on the prefix. The time of the announcement is encoded using `10.x.y.z` where the last three octets are a binary count of the number of second from the start of the month to the current time [27].

Upon specification the purpose of the aggregator attribute was to encode the IP and ASN of a router aggregating multiple routes into one. To validate that the aggregator IP we set was not modified, we analyzed received announcements from all VPs for a day (20.11.2019) for our Beacons. We found in 99% of the analyzed announcements an aggregator IP within the prefix `10.0.0.0/8`, indicating this field was set by one of our Beacon routers. For all other announcements, the aggregator IP field was empty. We could not find any specific reason though we noticed that more than half of the announcements with empty aggregator IP fields we sent by AS32097, a peer in the Isolario route collector project. One possible reason could be misconfigured or malfunctioning routers. We decided to not use announcements with a missing aggregator IP, because without the encoded timestamp our metrics would become less accurate.

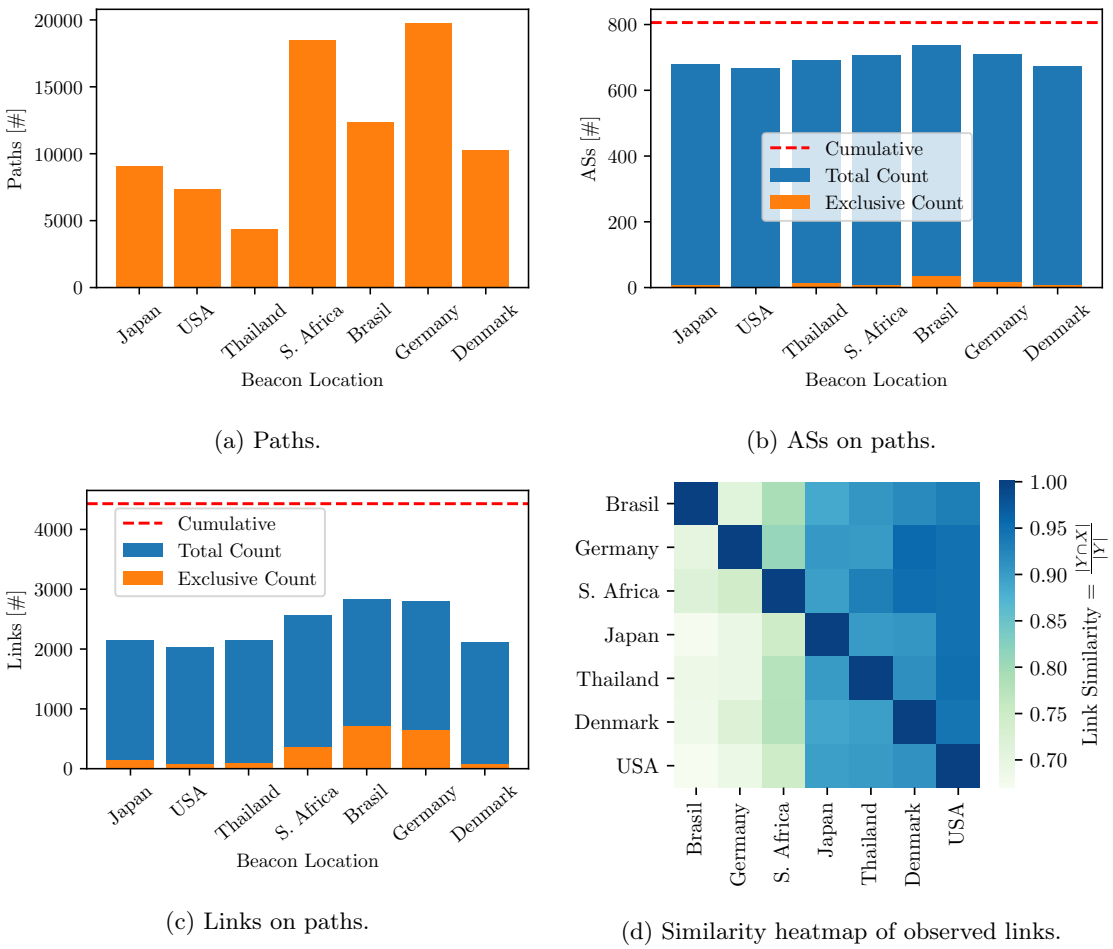
5.2 Benefits of Route Collectors and Beacons

For results of our study to be relevant and representative for networks that we do not observe on paths to our Beacon prefixes, we require large number of ASs and links on the paths from VPs to Beacon prefixes. Therefore, we assessed how much each Beacon benefits our measurements with regard to the number of paths, links, and ASs we observe between each Beacon and all VPs.

Figure 5.1a shows the number of unique, observed paths from all VPs to each Beacon location. Interestingly, we observe more than four times as many paths to the German and African Beacon compared to the Beacon in Thailand. We think the reason for this is that fewer route collector peers are located in the Thailand region.

Figures 5.1b and 5.1c visualize the number of unique ASs and links found on AS paths leading to each Beacon. Since all our Beacons are topologically located close to the Internet core we find only a very small amount of ASs and links exclusively seen on paths to a specific Beacon prefix. The dashed red lines indicate the total number of ASs and links observed on paths to all Beacons. Figure 5.1b shows that with any individual Beacon we can already see at least 85% of all ASs we observe with 7 Beacons. We make slightly different observation for links found on AS paths. Figure 5.1c indicates that each Beacon slightly contributes to the overall seen links, and Beacons located in Brasil, South Africa, and Germany account for the largest amount of links exclusively seen on paths to the respective Beacon.

The heatmap in Figure 5.1d visualizes the similarity of the sets of links observed on paths leading to the respective Beacon. The similarity metric is $\frac{|Y \cap X|}{|Y|}$ where X and Y are the sets on the x-axis and y-axis respectively. As an example 87% of all links found on paths to the USA Beacon can also be found on paths to the Japanese Beacon. The vertical bars in this heatmap suggest that each Beacon is equally similar or dissimilar to all other Beacons. Overall, we found that each new Beacon increased the number of ASs and links



(a) Paths.

(b) ASs on paths.

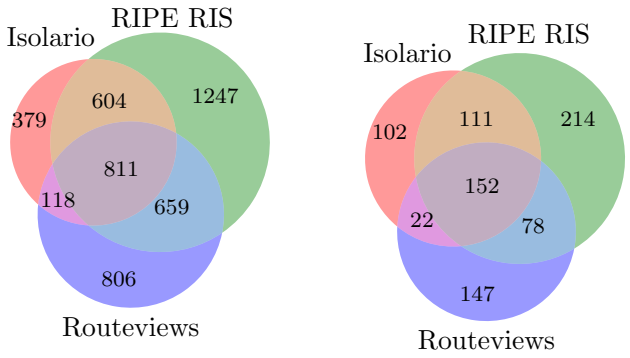
(c) Links on paths.

(d) Similarity heatmap of observed links.

Figure 5.1: ASs on paths, links on paths, and paths leading to the respective Beacon. The orange area denotes ASs, links, and Paths exclusively seen for one Beacon. The dashed red line represents the total number of unique ASs, links, and Paths across all Beacons.

we observe only slightly, because the links and ASs on the paths we see for each Beacon are very similar. Therefore, the question becomes: Why would we need 7 Beacons instead of one? Each Beacon router injects BGP Updates from a different site on the Internet. As a result we find that the median of how often a given link occurs on different paths is 11. Thus, we have much more evidence for each individual link and can draw a safer conclusion about RFD deployment on that link. In contrast, if we were to use any of the 7 Beacons individually than we the median would only be 3 on average.

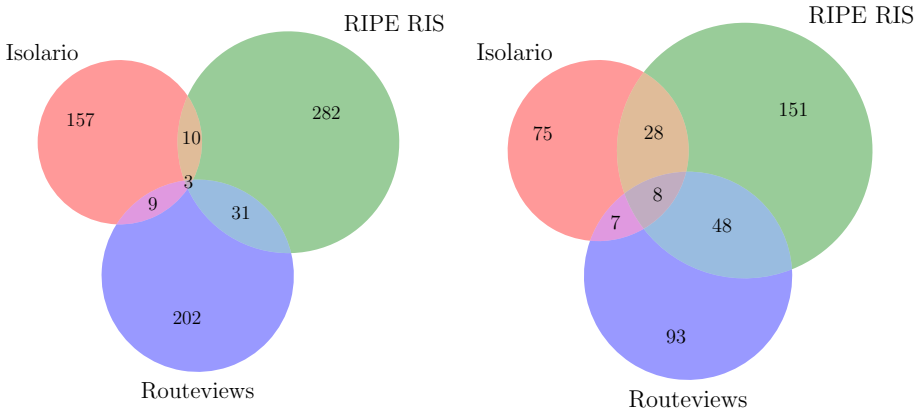
In this study we use the three largest public route collector projects, namely Isolario, RIPE RIS, and Routeviews. Figure 5.2b shows a venn-diagram of how many new ASs each route collector project contributes, *i.e.,* ASs observed on paths from the respective route collector peers to all Beacons. We found that all three projects are very relevant and add a substantial amount of new ASs to our dataset. We also apply our metrics to ordered AS links. The benefit of each route collector project in terms of AS link count is shown in Figure 5.2a. The RIPE RIS project contributes the most unique AS links, whereas Routeviews and Isolario



(a) Ordered AS links on AS paths

(b) ASs on paths

Figure 5.2: Ordered AS links and ASs found on AS path in announcements for our prefixes.

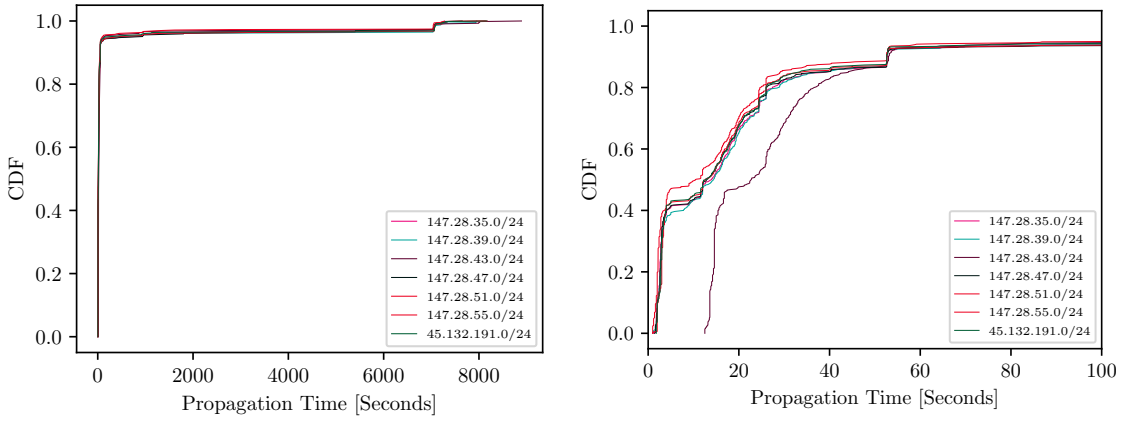


(a) VP IP addresses.

(b) VP ASs.

Figure 5.3: Venn diagram of VP ASs and VP IP addresses for each of the three route collector projects. Integers represent absolute count in the respective subset.

add a similar amount of unique AS links to our dataset. Figure 5.3 visualizes the size of each project in terms of sheer VP IP addresses and number of ASNs. Again, RIPE is the largest project and Routeviews and Isolario are similar in size. Figure 5.3b in particular shows that each project has their own set of ASs they peer with. This means it is beneficial using all three projects, which is also important to know for related research projects based on control plane data from route collectors. Surprisingly we also found that 60% of ASs on all paths between VPs and our Beacons are VPs. The AS paths also include 81% and 26% of the Caida top 100 and top 1000 [28] ASs respectively, where ASs are ranked by customer cone size.



(a) Entire range.

(b) Plot limited at 100 seconds.

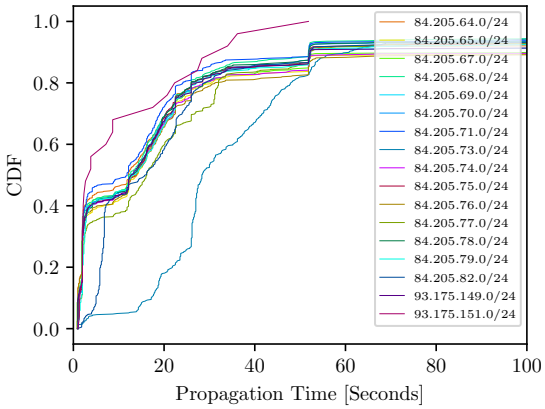
Figure 5.4: Propagation time of anchor prefixes. Average time delta between aggregator IP of first received update of Beacon event for each VP.

5.3 Infrastructure Validation

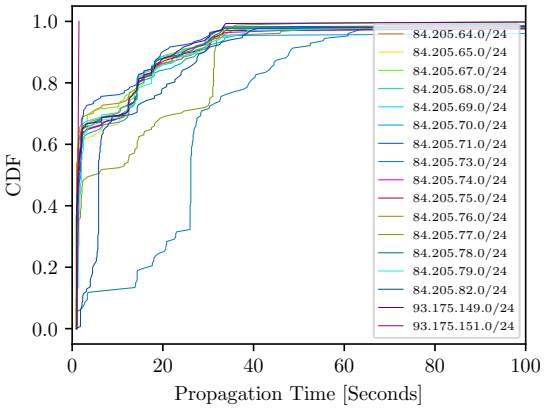
In Section 2.4 we explained that ASs can perform prefix filtering based the existence or validity of route objects or RPKI ROAs. Therefore, we setup both for all Beacon prefixes so that our Beacons are visible from as many points in the Internet as possible. Before oscillating the prefixes we checked whether our prefixes were propagating as expected by announcing them statically. We identified all VPs from RIPE RIS and Routeviews that have a RIB table size greater than 700,000 prefixes. In the beginning of 2020 there were $\sim 800,000$ prefixes present in the Internet routing table. We used the threshold 700,000 to identify routers that see most of the Internets routing table and should therefore also see our prefixes. All prefixes were visible at more than 99% of all VPs. Some ASs were also manually filtering prefixes that are being announced from their customers, but these minor issues which were quickly resolved.

To further validate our infrastructure we measured the time it takes from sending the announcement from the Beacon until a first announcement from at each VP, the *propagation time*. For this we used the seven anchor prefixes that are being announced and withdrawn every 2 hours. Figure 5.4a shows the distribution of propagation times for each anchor prefix across all VPs seeing the respective prefix. Interestingly, we find a small dent in the distribution at around 2 hours, which is the time of the withdrawal. This suggests that some VPs did not export a new route following the announcement, only after withdrawing the prefix. Figure 5.4b shows the same distribution in detail for the first 100 seconds. We notice that prefix `147.28.43.0/24` is shifted by 12 seconds to the right, but we are uncertain why this is the case. Overall, we think that our Beacons propagate normally and did not influence our measurements.

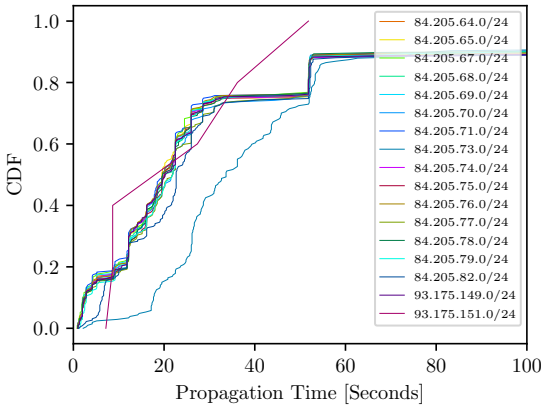
For comparison, Figure 5.5a shows the propagation time for RIPE Beacons, which also oscillate every 2 hours, to all route collector peers. Figure 5.5e shows the median between all prefixes for our anchor prefixes and the RIPE Beacons. Surprisingly, the distributions



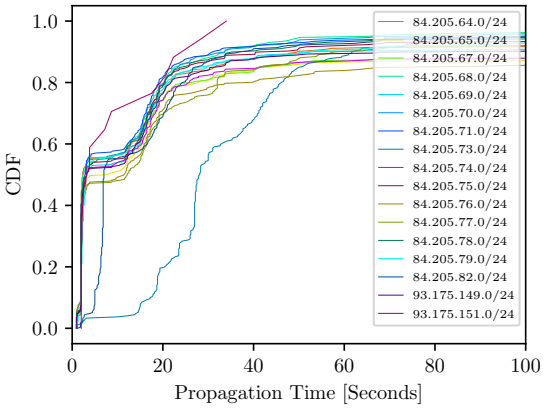
(a) All route collector projects.



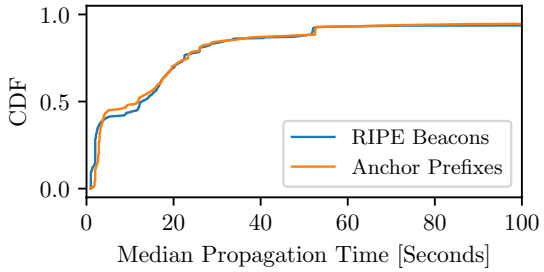
(b) Isolario.



(c) Routeviews.



(d) RIPE RIS.



(e) Median of propagation times for each VP. Anchor Prefixes and RIPE Beacons behave almost identical.

Figure 5.5: Distribution of propagation times of all RIPE Beacons for all route collector projects and individually. All plots are cut off after 100 seconds.

are almost identical between, suggesting that the shape of the distribution depends solely on the VPs.

Figures 5.5d, 5.5c, and 5.5b show the distribution of propagation times from all RIPE Beacons to VPs for each route collector project individually. Each distribution has a very unique shape. Some set of VPs in the Routeviews project seem to export the updates for all Beacons at the exact same time after 50 seconds. Most Isolario VPs export updates for all but two Beacons within 30 seconds of the announcement. Contrary to the Routeviews VPs, the Beacons are not received as similar for the RIPE RIS VPs, maybe suggesting a more diverse set of route collector peers. The Beacon prefix `93.175.151.0/24` stands out regardless of the chosen route collector project. We observed at the Beacon is only seen from 3.5% of all VPs and upon further investigation we found that the RPKI ROA did not exist for the Beacon prefix. All other RIPE Beacon prefixes do have a valid RPKI ROA. We contacted RIPE NCC and they created a valid RPKI ROA for this prefix as well as 3 other Beacon prefixes. This suggests that their provider filters BGP Updates based on the existence of RPKI ROA objects.

When performing active BGP measurements one needs to make sure to not make life more difficult for network operators and impact real-world operations. As we are sending a lot of BGP updates, especially in our second measurement period, we need to make sure the raw amount of BGP Updates does not overwhelm other routers. In the first measurement period, we caused 0.48% of all IPv4 control plane traffic, whereas in the second period our Beacon caused 0.54% of all IPv4 BGP updates. Interestingly, the prefixes oscillating every 1 minute were still causing a lot fewer updates than other prefixes on the Internet. As an example, we picked March 1th an measured how many announcements belonged to each prefix. We found ~ 50 prefixes causing 3 times as many updates as one of our Beacon prefixes and 4 prefixes caused 17 times more updates individually than one of our Beacon prefixes. We also setup a website for network operators which lists the schedule of our Beacons as well as contact information in case routers do get overwhelmed with our measurements. On top of this we create whois entries for all Beacon prefixes and Beacon AS numbers with our contact information.

CHAPTER 6

Results

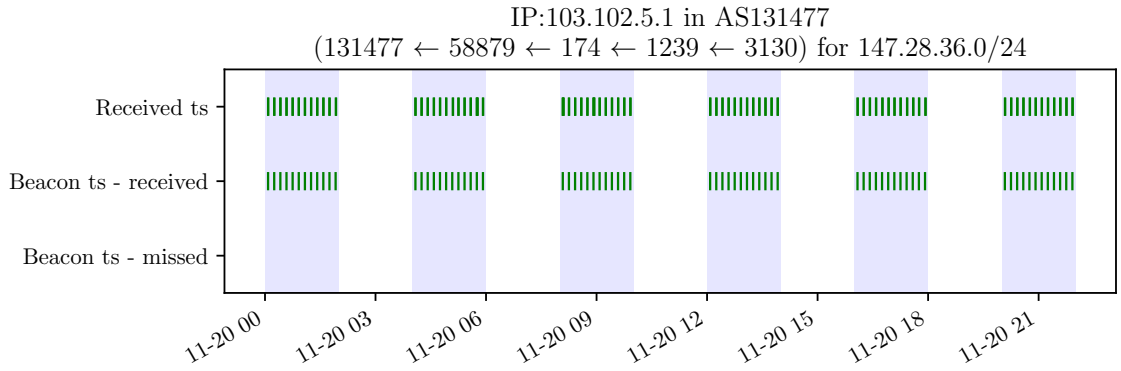
In this chapter we present our results in five sections. First, we will focus path labeling and its caveats (Section 6.1). Then, we analyze how the metrics behave and correlate across different update intervals (Section 6.2). Section 6.3 elaborates on pinpointing RFD ASs and one parameter which we were able to precisely measure. Finally, in Section 6.4 we validate our results and show limitations of the experiment in Section 6.5.

6.1 Path Analysis

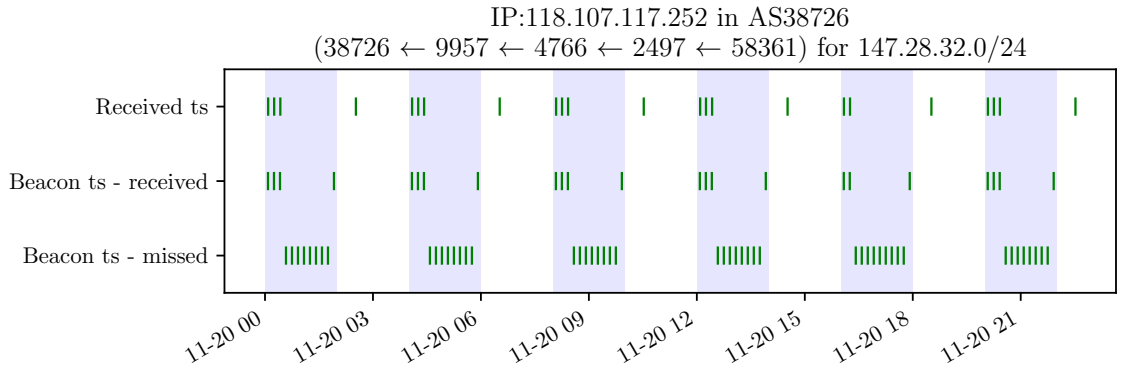
The first step in our analysis is labeling paths, based on the received announcement pattern as described in Section 4.2. Examples for the three general types of announcement patterns are visualized in Figure 6.1. Each plot shows announcements exported by a specific VP for a prefix and path. Each of these plots consists of three axis: *Received ts*, *Beacon ts - received*, and *Beacon ts - missed*. The upmost axis reflects exactly when we receive announcements (green) for the given path from the vantage point. The two axis below depict when updates were sent from the Beacon router and whether they were received or not. Blue areas denote the Burst phases and white areas stand for Breaks.

Figure 6.1a shows that for every announcement sent at the Beacon, a matching announcement was exported by the VP, *i.e.,* no announcements went missing, hence the third axis is empty. This indicates that RFD did not occur and ASs on the path (131477 ← 58879 ← 174 ← 1239 ← 3130) are not using RFD for the respective neighbors, hence labeled *no-RFD*.

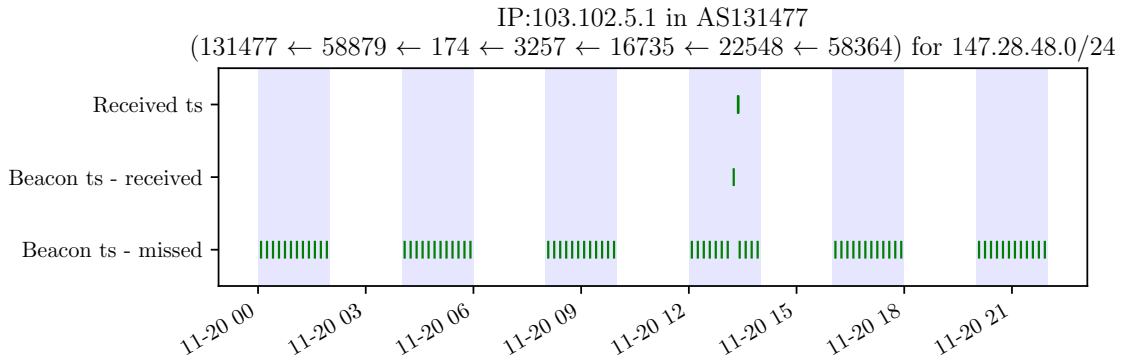
Figure 6.1b shows normally propagating, missing, and delayed announcements. For every Burst-Break pair we first find a sequence of normally propagating announcements, followed by a period, where announcements are suppressed by some AS on the path. In the Break we observe a single announcement, *re-advertisement*, which was originally sent at the end of the Burst. This is a typical RFD announcement pattern, which indicates that one of the ASs on the path is using RFD (see Section 4.2). The *r-delta* explained in Section 4.2 is the amount of delay between the last announcement sent by the Beacon and exported by the VP.



(a) non-RFD announcement pattern.

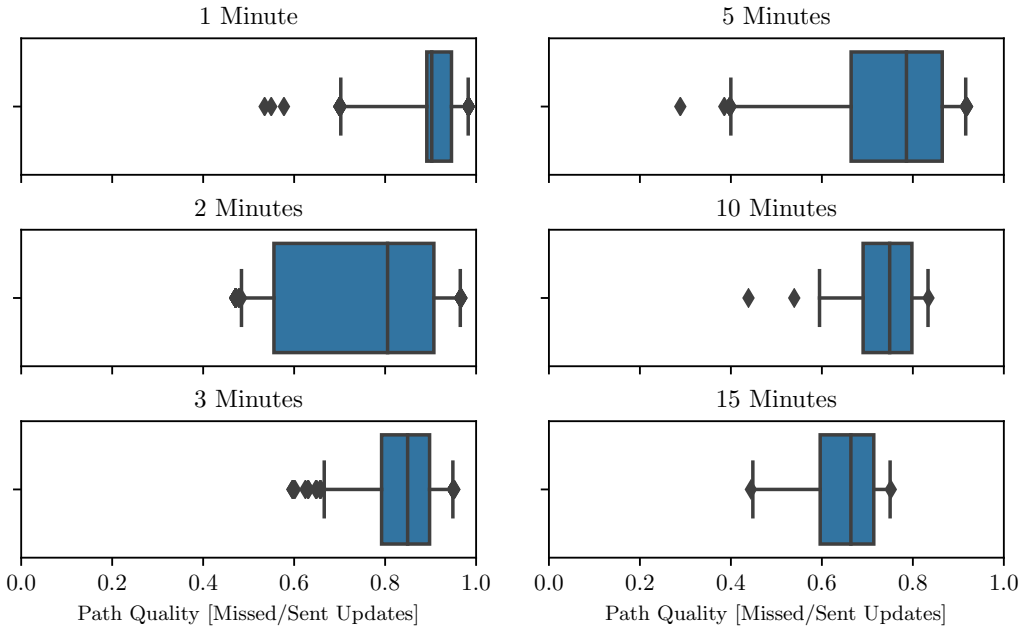


(b) RFD announcement pattern.



(c) Not enough announcements.

Figure 6.1: RFD, non-RFD announcement pattern as well as an example of two few updates to draw a conclusion about whether RFD is disabled.



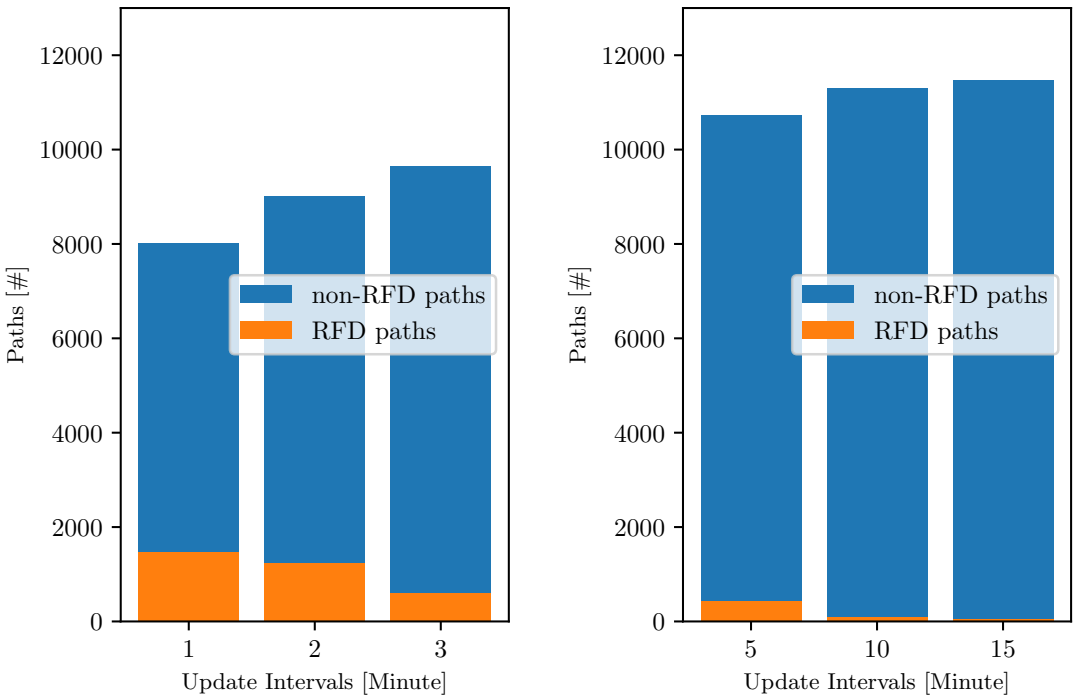
(a) March, 6 hour Break between Bursts. (b) April, 2 hour Break between Bursts.

Figure 6.2: Portion of missed announcements for damped paths for different Beacon update intervals. Whiskers are at the 2nd and 98th percentile.

During the convergence process of a BGP Update we sometimes observe paths for a prefix that are very rarely, *i.e.,* once a day. Figure 6.1c visualizes one of these paths, for which we only received one announcement on the 20th of November. It is impossible to draw a conclusion about RFD usage on the path, because there are simply not enough announcements to analyze. In contrast, one could also label all paths non-RFD if no announcements were missed at all, but it is normal that sometimes announcements are not received because of a changing infrastructure. Hence we introduce a metric representing the quality of a path, that is the ratio between missed announcements and sent announcements. The path in Figure 6.1a receives a score of 0, because all sent announcements have been exported by the VP. The path in Figure 6.1c receives a score of 0.99, because almost all announcements were missed. All paths with a score above a specific threshold are ignored and not labeled. Figure 6.2 shows the distribution of path qualities across RFD paths for different frequencies. Note that the whiskers are at the 2nd and 98th percentile. For simplicity and to be able to compare results we chose to label any path *non-RFD* where the path quality is less than 0.5, regardless of the frequency as the Burst length did not change across frequencies. All other paths were discarded. Notably, we also observe the trend that the median path quality roughly increases with a larger update interval. This suggests that with a larger update interval, routers suppress the Beacon prefixes later in the Break.

After applying the described methods to all paths we have a set of paths labeled with *RFD* and a set of paths labeled with *non-RFD* for each update interval and a third set with insufficient path quality which we ignore.

Figure 6.3 shows the number of RFD and non-RFD paths for each update interval. Intu-



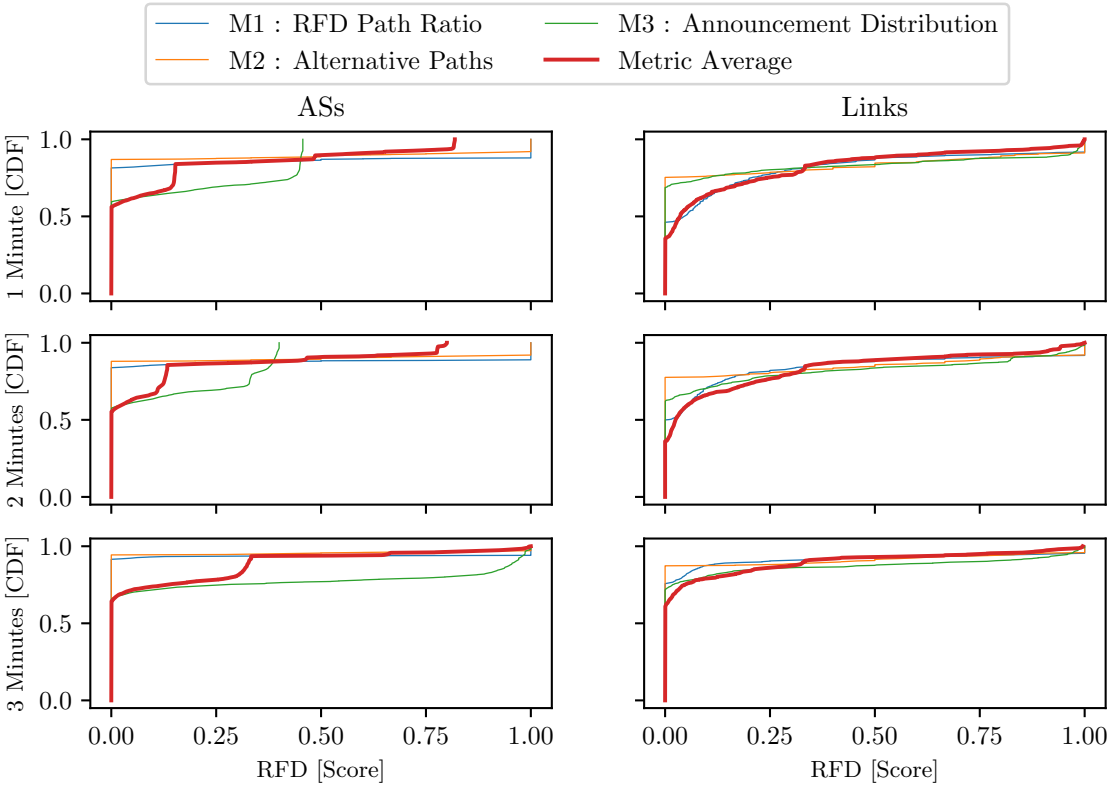
(a) March, 6 hour Break between Bursts.

(b) November, 2 hour Break between Bursts.

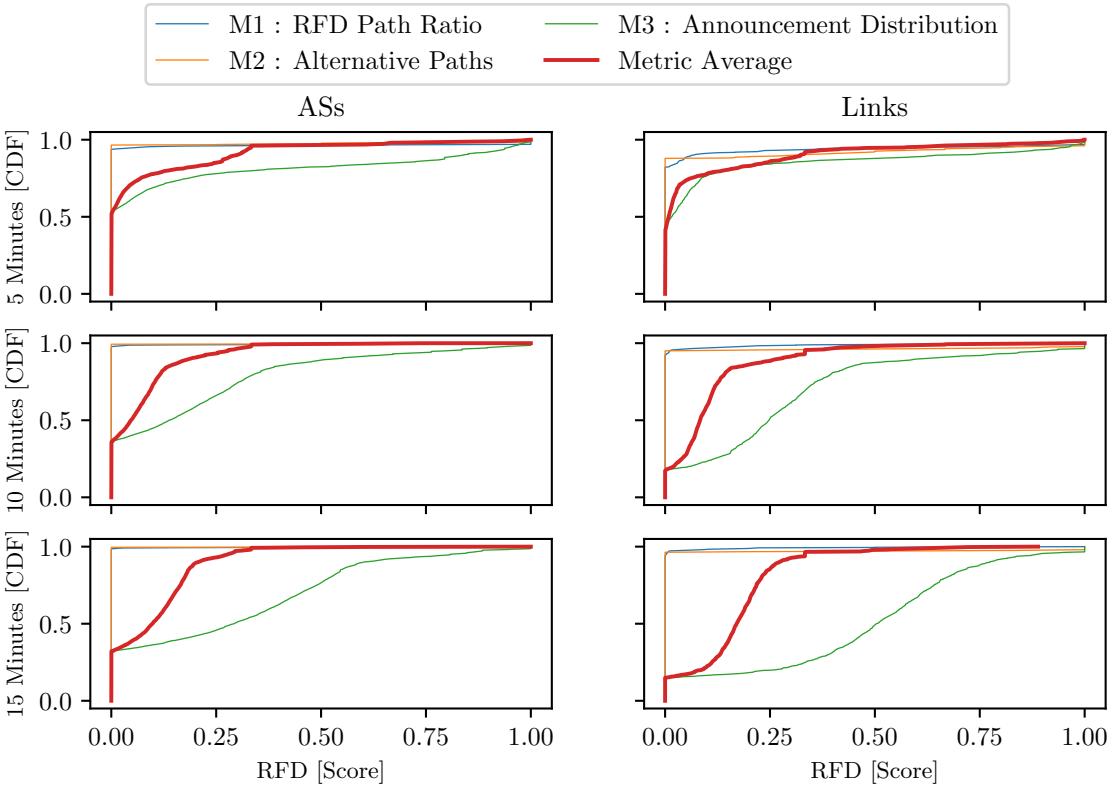
Figure 6.3: Number of RFD and non-RFD paths for different frequencies.

itively, with a decreasing update interval we detect more RFD paths and therefore the share of RFD paths compared to all paths increases. We observed the largest increase of RFD paths from 5 to 10 minutes and 3 to 2 minutes, suggesting that many routers are configured to start dampening prefixes for which they receive updates every 2 minutes or 5 minutes respectively.

For both measurement periods the total number of paths is decreasing with a smaller update interval. We suspect the reason is that the penalty in some routers for our prefixes does not entirely reset during the Break, thus accumulates across multiple Burst-Break pairs. Subsequently, these routers never announce our prefixes, which results in fewer paths to our Beacons observed at VPs. The max-suppress-time, the longest configured timespan a router suppresses a route after the Burst, is set to 60 minutes in Cisco and Juniper default configurations. As we find the total number of paths observed decreasing with a smaller update interval we can derive that at least some but few routers are configured with a max-suppress-time larger than the Break length, 2 and 6 hours, respectively.



(a) March, 6 hour Break between Bursts.



(b) April, 2 hour Break between Bursts.

Figure 6.4: CDF of metric scores assigned to ASs and links for each update interval.

6.2 Metric Analysis

6.2.1 Score Distribution

After labeling paths with RFD and non-RFD we apply the three metrics described in Section 4 to both the labeled paths and raw updates and compute a score for each link and AS. A score of 1.0 suggests RFD usage, 0.0 denotes that RFD was not detected, and a score in between implies that the data is contradicting. Note that, 0.0 does not mean RFD that is not in use, but that we were not able to detect it. Figure 6.4 shows the distribution of scores for all ASs and links across different update intervals. Plots in the left column show the distribution of scores for ASs, whereas plots in the right column show scores for links. The update interval increases from top to bottom.

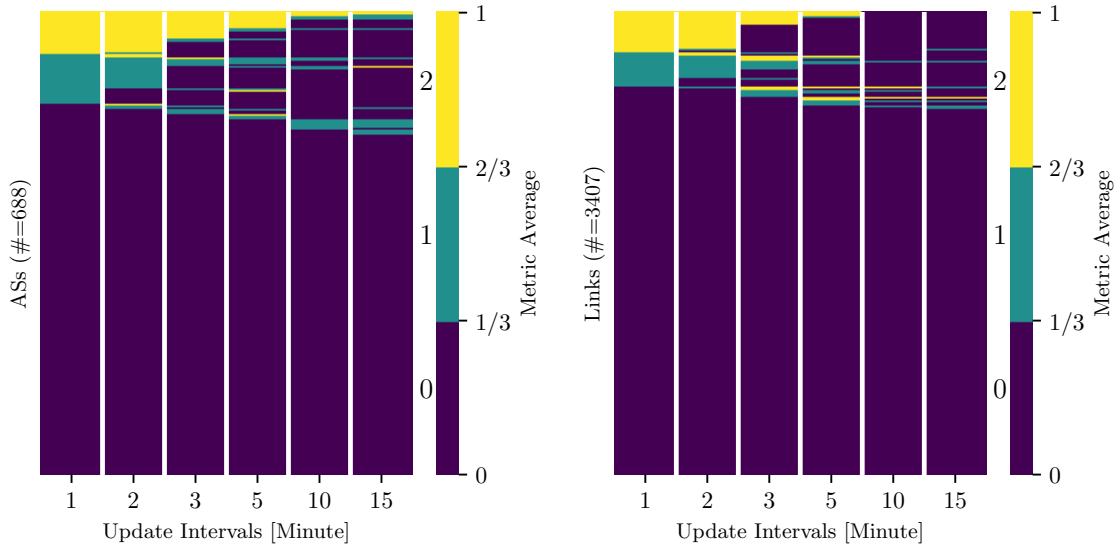
First of, we notice that the distribution of M3(green) does not agree with M1 and M2 for update intervals where little RFD occurs (10 minutes and 15 minutes). In Section 4.3.3 we explained that the calculated score for a given AS or link is the relative change of the announcement distribution normalized by the 5-th percentile of all relative changes. If only very few ASs use RFD the announcement distribution will be rather similar for most ASs. This means the 5-th percentile is very close to the median resulting such a distribution.

Viewing only M1 and M2 for the update intervals 10 and 15 minutes we find that almost most ASs and link receive a score close to 0. This suggests that few routers are configured to dampen Beacon prefixes, for which route updates are received every 10 minutes or rarer. This matches the observation we made about paths (see Figure 6.3), where few paths are labeled RFD for the update intervals. Also, it would be surprising if the results for individual ASs and links were different because M1 is fully and M2 is partially based on the labeled paths. Intuitively, we also find a significant increase in scores for smaller update intervals.

For the large update intervals in Figure 6.4b the score for M2 is distributed similarly for ASs and links, most links and ASs received a score of 0.0. For smaller update intervals shown in Figure 6.4a we find that few ASs received a score of 1.0, but a substantially higher share of links got score 1.0 for M2. More than half of the links for the update intervals one and two minutes received a score greater than 0. This difference between the links and ASs may indicate that ASs are configured inconsistently throughout the entire network, but consistently configured for each link.

6.2.2 Consistency across Update Intervals

Figure 6.5 visualizes the consistency of the average score of ASs and links across all update intervals. Each column shows all ASs (or links) for which we have measurement results in both measurement periods, where each cell is colored using the average between all three metrics for the respective AS (or link). The heatmaps are vertically sorted by so called *metric sequences*, which are the average scores for a given AS (or link) across all update intervals and binned into three spaces 0, 1, and 2, representing $[0, \frac{1}{3}]$, $(\frac{1}{3}, \frac{2}{3}]$, and $(\frac{2}{3}, 1]$. E.g.: (3216 $\leftarrow$ 37100) has sequence 2,2,2,0,0,0, meaning damping occurs for update intervals 1,2, and 3 minutes but does not occur for 5,10, and 15 minutes.



(a) ASs.

(b) Links.

Figure 6.5: Average of all metrics for each AS and link, showing how consistent they behave across multiple update intervals.

Metric Sequence	Share
(0, 0, 0, 0, 0, 0)	73.26%
(1, 1, 0, 0, 0, 0)	4.22%
(1, 0, 0, 0, 0, 0)	2.47%
(2, 2, 0, 0, 0, 0)	2.47%
(2, 2, 2, 2, 0, 0)	1.74%
(0, 0, 0, 0, 1, 1)	1.74%
(2, 2, 2, 0, 0, 0)	1.45%
(0, 0, 0, 0, 0, 1)	1.02%
(0, 0, 0, 1, 0, 0)	0.87%
(1, 1, 1, 1, 0, 0)	0.87%

(a) ASs

Metric Sequence	Share
(0, 0, 0, 0, 0, 0)	78.66%
(2, 2, 0, 0, 0, 0)	4.90%
(1, 1, 0, 0, 0, 0)	2.03%
(1, 0, 0, 0, 0, 0)	1.58%
(2, 2, 2, 0, 0, 0)	1.50%
(2, 2, 2, 2, 0, 0)	1.00%
(0, 0, 1, 0, 0, 0)	0.88%
(1, 1, 1, 0, 0, 0)	0.82%
(0, 0, 0, 2, 0, 0)	0.62%
(0, 0, 0, 1, 0, 0)	0.56%

(b) Links

Table 6.1: Shares of metric sequences for ASs and links.

Table 6.1 shows the shares of each metric sequence for ASs and links. We find that the most frequently seen sequences are decreasing, *i.e.,* if an AS or link received a high score for one update interval, then it receives a lower score for larger update intervals. There are sequence which are not consistent, *e.g.,* (0,0,0,0,1,1), which is ether cause by scores being very close interval borders or measurement artifacts. Metric sequences, which start with three 0's and end with one or more 1's are caused by M3 which is skewed for large update intervals where only very little RFD occurs (see Section 6.2). Overall Table 6.1 shows that sequences are consistent for the most part. By far the majority of links and ASs have the sequence (0,0,0,0,0,0) which means we did not detect RFD for any of these update intervals. Apart from (0,0,0,0,0,0), for both ASs and links the sequences (1,0,0,0,0), (1,1,0,0,0), and (1,1,1,0,0) also appear frequently. This indicates that the underlying data is contradicting for an individual link or AS, as 1 indicates a score in the interval , $(\frac{1}{3}, \frac{2}{3}]$. Though as expected the share of these frequencies is slightly lower for links (7.56%) compared to ASs (4.43%). This, again, suggests that routers are configured on a per neighboring AS basis rather than one consistent configuration for the entire AS. Primarily, this table is a detailed view of the heatmap above (Figure 6.5) showing how consistent ASs and links behave across multiple experiments and update intervals.

One might ask why the resulting scores are not either 0.0 or 1.0, but mostly in between. The exact reason depends on the on the metric. M1, RFD path ratio, solely depends on paths labeled with RFD or not, thus directly inheriting inaccuracies and errors from this process. The primary reason why the results of M1 may be distant from both 0.0 and 1.0 is likely because ASs may have RFD inconsistently configured even on link basis. When we label paths with RFD, we are generally very certain that at least one RFD AS (or link) occurred on the path because of the strict pattern matching and the unique announcement pattern RFD produces for the Beacon pattern (see Section 4.2). Therefore, most inaccuracies will come from labeling paths with no-RFD because this method is threshold based, thus naturally being less accurate. M2, the metric based on analysing alternative paths, is partially based on the labeled paths, thus may also have the same errors as M1. Additionally, if no alternative paths are found for a given RFD path and too few paths exist to build up the graph between VP and the Beacon, M2 may have too little data to infer RFD usage and return 0.0. M3 is implicitly mostly not returning exactly 0.0 or 1.0, because it represents the change of the announcement frequency during the Burst, normalized by the maximum value observed.

6.3 Pinpointing

6.3.1 Score Threshold and Minimum Visibility

We pinpoint ASs and links that use RFD based on the average of all three metrics. All ASs and links above a given threshold are declared damping ASs. It is non-trivial to determine the correct threshold, because we have very limited ground truth at our hands, thus we can never be sure that there are no false positives. Instead, we just aim to optimise the share of damped paths that we cover with the smallest amount of ASs possible. We created an evaluation function that rates a given threshold, where R is the set of RFD ASs (or links)

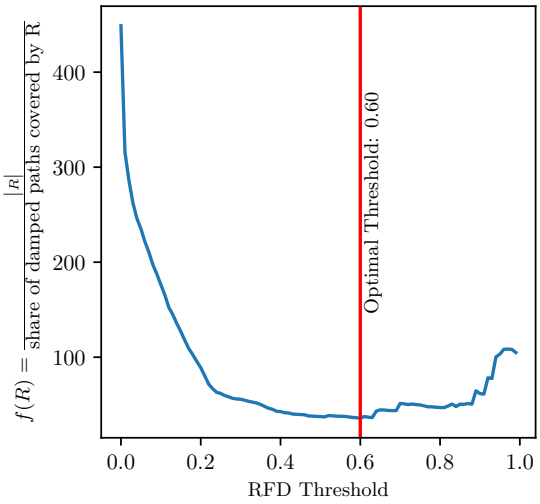


Figure 6.6: Evaluation function for the threshold that defines when an AS is damping or not.

that have been labeled with the respective threshold:

$$f(R) = \frac{|R|}{\text{share of damped paths covered by } R}$$

The results of this function are visualized in Figure 6.6. Based on these figures we set the threshold for AS to 0.60. This threshold will likely change for future measurements, because it is based on just two months and the Internet infrastructure may change in the future. We are also limiting ourselves to ASs that are on paths of at least half of the Beacons prefixes, *i.e.,* at least different Beacon locations. ASs for which we see fewer prefixes are labeled RFD false. We found the data to be too sparse to make inferences about RFD deployment on a link basis. Therefore, we chose to not include detailed analysis about links. For ASs we have validated our results by asking the network operators themselves and encountered one false positive above the threshold (see Section 6.4).

6.3.2 Deployed Parameter Sets

Figure 6.7 shows how many ASs we identified for each update interval. As expected, the number of RFD ASs rises with the update interval decreasing. Intuitively, the increase of RFD ASs and links for a given update interval compared to the next larger one is very similar to Figure 6.3 where the number of RFD paths is shown for each interval.

While Figure 6.7 illustrates how quickly a prefix needs to flap to get damped, we cannot infer the exact value of the *suppress-threshold* in use because one Beacon event may cause multiple updates distant in topology (*e.g.,* path hunting). We assume, however, that many operators use predefined configurations, and try to find confirmation in our data. Currently, there are two sources of parameter sets: (*i*) the recommendations by the IETF and RIPE [3, 4], and (*ii*) vendors that ignore these recommendations and pre-configure a deprecated *suppress-threshold* (see Table 2.1). In Figure 6.7, a single sharp incline is visible at 5 minutes. A

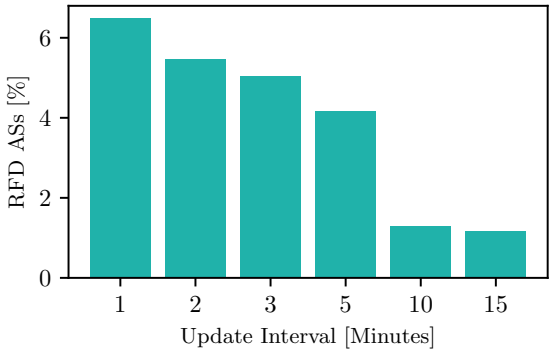


Figure 6.7: Share of RFD ASs (695 total) for each update interval.

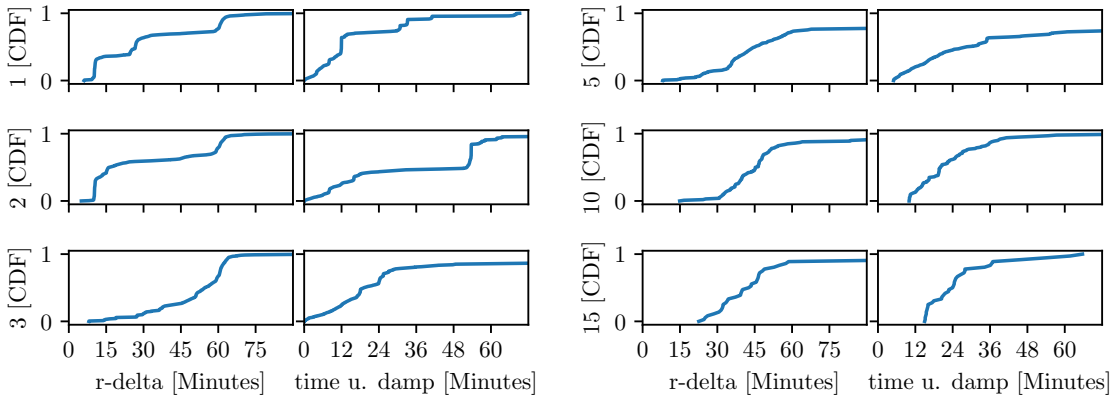


Figure 6.8: Distribution of the average time passed from the end of a Burst until the re-advertisement (left) and the time from the beginning of the Burst until suppression (right) for each RFD path. The numbers on the y-axis indicate the update interval.

router with deprecated default values would start damping at the 5 minutes update interval. We suspect the continuous increase of RFD ASs for the smaller update intervals is caused by some network operators following the current recommendations. The very few damping ASs at 10 or 15 minutes are likely induced by updates amplified by topology properties. Based on feedback from almost 50 network operators we were able to confirm that there is a significant tendency (≈60%) to use vendor default values. As we limit ourselves to ASs for which we have enough data, the results in Figure 6.7 shall be viewed as a lower bound. The actual deployment is likely slightly higher.

6.3.3 Maximum Suppress Time and Time until Damp

Even though it not possible to derive the exact RFD parameters for a given AS or even a specific announcement pattern, we can draw some more general conclusion about about the RFD configuration by calculating the *time until damp* and the *time until re-advertisement*. The time from the beginning of the Burst until the missed announcement, where all following announcements, except the re-advertisement, are also missed is called *time until damp*. We

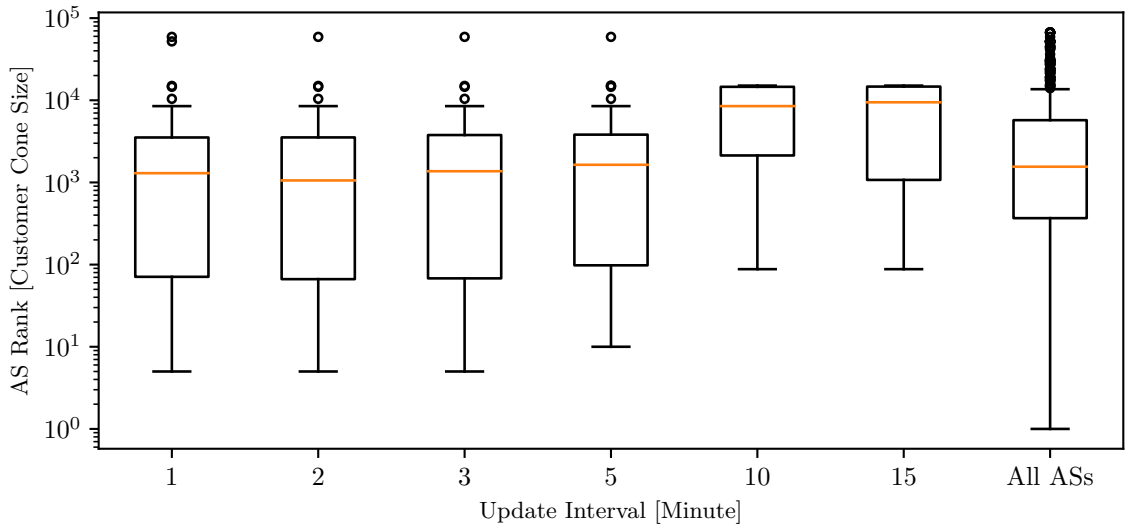


Figure 6.9: Caida AS Rank distribution, where ASs are ranked by customer cone size, for RFD ASs and as comparison, all ASs.

also measure the time from the end of the Burst until the first observed announcement, which call *time until re-advertisement*. Figure 6.8 shows the distribution of these values for each RFD path and the respective update interval. First of, we notice that the time until re-advertisement rarely surpasses 60 minutes, suggesting that it is uncommon to configure the *max-suppress-time* larger than 60 minutes. 60 minutes is also the default and recommended value for the max-suppress-time. For the smallest update interval we find three plateaus, at 10, 30, and 60 minutes, suggesting that these are the most commonly configure values for the max-suppress-time parameter. For larger update intervals we do not find such plateaus, because here, the penalty decreased naturally below the reuse-threshold much quicker than the max-suppress-time. For the time until damp we observe that if RFD is in use, then the prefix also suppressed within 60 minutes for all update intervals. Also, the minimum decreases with the smaller the update interval, with even some router already suppressing the route after the first announcement of the Burst. We suspect that these are routers with very strict RFD parameters, which already suppress prefixes flapping every 15 minutes. Apart from a changing maximum and minimum, we do not observe any notable patterns in the distribution for the time until damp.

6.3.4 AS Rank of Damping ASs

Caida provides a service where ASs are ranked by their customer cone size, representing roughly their importance on the Internet [28]. The highest ranked ASs are not necessarily Tier 1 ASs nor are they responsible for forwarding the most traffic, but this gives an indication of their significance and size. Figure 6.9 shows a standard boxplot of the rank for RFD ASs as well as an overall distribution of ranks of all ASs. First, we notice that our dataset includes ASs from all rank with even the highest ranked AS being represented. This plot also shows that the maximum rank (minimum number) and also median increases

with the update interval decreasing. This shows that higher ranked ASs use less strict RFD configurations, thus likely follow the recommendations and lower ranked ASs are more likely to start dampening at larger update intervals.

6.4 Validation

In this study we attempt to measure real-world usage of RFD. Therefore, we validated our results by asking networking operators as well as parsing *whois* entries for hints about RFD.

In mid-February and again at the end of April we sent out e-mails to about 100 ASs of which most were labeled RFD true. We have collected ground truth from 32 ASs, of which 22 deploy RFD. Of these we have one false positive, namely AS5645, where the direct upstream is also using RFD and thus makes it impossible for the heuristics to draw a correct conclusion. This problem likely occurs for other ASs as well, but we tried to limit these cases by setting a minimum number of Beacons that are required to be seen for a given AS (Section 6.3.1).

All Regional Internet Registries provide a *whois* service, which is a database of resources describing who owns and manages, *e.g.,* ASs and prefixes. These entries sometimes included arbitrary information such as RFD configurations, hence we decided to parse all whois entries for ASs for RFD related information. In total we found 41 entries with information about RFD of which all were found in the RIPE database. First, we should not that whois entries are not guaranteed to be correct and should be taken with a grain of salt. 39 ASs stated that RFD is in use and 2 stated they are not using RFD referencing RIPE deprecated recommendations from 2002 and 2006. Surprisingly, 21 ASs referenced RIPE-229 (2001) where by now deprecated parameters are recommended, which turn out to do more harm than good [2]. We found 9 (7 RFD, 2 non-RFD) of the 39 ASs on paths leading from VPs to our Beacons, meaning we can infer RFD usage about them. Of these there we no false positives and one AS was not detectable for us either caused by poor visibility or an unmaintained whois entry.

6.5 Limitations

Our measurement setup has two main limitations. First, our Beacon routers are located in or close to the Internet core and therefore the Beacons travel mostly from top to bottom in the Internet tree. As a result, ASs that deploy RFD solely towards their customers remain hidden. Based on the feedback from network operators we know that cases like these exist. A solution would be to send Beacons from a large share of stub ASs on the Internet, which is impossible due to the sheer number of stub ASs in existence.

The second limitation is that the number of updates we are sending from the Beacon routers is not necessarily equal to what is received at some far edge of the Internet. This is due to the path hunting process (convergence). Therefore, inferences on the number of damping ASs for a given update interval is unsharp, yet still usable as the comparison to ground truth shows (Section 6.3.2).

CHAPTER 7

Toolchain

7.1 Beacon Configuration

Initially we tried to use Bird [29] to announce prefixes to our peers. Later, we noticed that it does not support setting the aggregator IP field which we require to encode the timestamp. Therefore, we chose to use ExaBGP (Version 4.1.0-2074ac17) [30] instead. We ran ExaBGP as system service on an NTP synchronized Linux (Ubuntu or Debian) virtual machine. An example of how we configured ExaBGP on the Beacon routers is shown in Listing 7.1.

```
neighbor 192.168.10.20 {
router-id 192.168.10.30;
local-address 192.168.10.30;
local-as 58360;
peer-as 44869;
group-updates false;

capability {
    graceful-restart;
    }
}
```

Listing 7.1: Sample ExaBGP configuration file.

To trigger announcements and withdrawals for specific prefixes we created a cron-job running with the respective update interval, which execute a command similar to the one in Listing 7.1.

```
echo "
    neighbor 192.168.10.20 announce
    route 147.28.32.0/24
    aggregator ( 64513:10.0.10.10 )
    next-hop 192.168.10.30
    " > /usr/local/run/exabgp/exabgp.in
```

Listing 7.2: ExaBGP command to trigger BGP updates.

7.2 Receiving and Processing Data

To receive BGP updates from RIPE RIS and Routeviews peers we used BGPStream's [23] BGPReader, which were then saved in ASCII format. Unfortunately the route collector project Isolario is not integrated into BGPReader. Therefore, we had to build a wrapper which download the raw MRT dumps from Isolario's website and fed them into BGPReader. BGPReader does not support displaying the aggregator IP field in its output, hence we extended it. Our modifications can be found here: `https://github.com/cmosig/libbgpstream`.

The output of BGPReader was then primarily processed using the Python data analysis library *pandas*. We also used *joblib* for parallelization, *networkx* for graph analysis, and *seaborn* and *matplotlib* for plotting, as well as numerous other small, helpful libraries.

CHAPTER 8

Conclusion and Outlook

In this thesis we presented a novel approach to measure BGP Route Flap Damping using BGP Beacons and public route collector projects. Using a carefully designed Beacon schedule we trigger RFD in routers on the Internet and create a clear signature which we pick up at route collector peers. This RFD signature is used to label AS paths with RFD deployment and using heuristics we pinpoint ASs that deploy RFD today.

With different Beacon frequencies we target all relevant RFD parameter sets and unveil that at least 6% of ASs use RFD of which most use harmful, deprecated vendor default configurations. We validate our findings with ground truth from network operators which also suggests that approximately 60% use vendor defaults.

In conversations with network operators we found that there is confusion about how recommendations on RFD are correctly applied. The RIPE document RIPE-580 and the IETF BCP 194 are both suggesting to adjust the suppress-threshold without mentioning the other parameters. The RFD mechanism relies on four configurable parameters and three non-configurable values that differ depending on the vendor. The interplay of all parameters determines how RFD acts upon flapping prefixes. Therefore, clear instructions are necessary. The recommendations also rely on measurements that were conducted in 2010, hence may be deprecated, thus need to be revisited. In future work we will address these issues.

Our study is limited to a relatively small, though significant, portion of the Internet and it is up to future researchers to develop a methodology to cover a larger part of the Internet. There two sides to this. First of, our Beacons originate in the top of the Internet "tree" and thus we measure only "downwards"[1]. As a result, ASs damping solely customers (downwards) are not detectable for us. Second, we measure only $\approx 1\%$ of existing ASs on Internet, and are limited by the number and variety of vantage points. One may argue that this small share (though it includes most major networks) may not be representative for the rest of the Internet. Future work can try to address these visibility challenges, but it will be rather difficult, due to the information hiding of BGP, and likely require significantly more measurement infrastructure.

Routers that were overwhelmed from too many BGP updates were the initial motivation

[1] With the assumption that an update, which has travelled downwards once, will not travel up again due to AS relationships, *i.e.,* valley-free assumption.

to create the RFD mechanism. Since then, router performance has significantly improved and one could question the need for this, potentially harmful, mechanism. One network operator from a major AS told us that they do not use RFD and keep it reserve, because their routers were simply not overwhelmed. Future work could examine how much resources typical BGP churn consumes in today's routers and if there is still a need for RFD.

Bibliography

[1] Y. Rekhter, T. Li, and S. Hares. A Border Gateway Protocol 4 (BGP-4). RFC 4271, IETF, January 2006.

[2] Zhuoqing Morley Mao, Ramesh Govindan, George Varghese, and Randy H. Katz. Route Flap Damping Exacerbates Internet Routing Convergence. In *Proc. of ACM SIGCOMM*, pages 221–233, New York, NY, USA, 2002. ACM.

[3] Randy Bush, Cristel Pelsser, Mirjam Kuhne, Olaf Maennel, Pradosh Mohapatra, Keyur Patel, and Rob Evans. RIPE Routing Working Group Recommendations on Route Flap Damping. RIPE Document ripe-580, RIPE, January 2013.

[4] J. Durand, I. Pepelnjak, and G. Doering. BGP Operations and Security. RFC 7454, IETF, February 2015.

[5] Daily Routing Table Report APNIC. `http://thyme.rand.apnic.net/current`, February 2020.

[6] J. Postel. Internet Protocol. RFC 791, IETF, September 1981.

[7] IPv6 Adoption Visualization. `https://www.akamai.com/us/en/resources/our-thinking/state-of-the-internet-report/state-of-the-internet-ipv6-adoption-visualization.jsp`, February 2020.

[8] G. Huston. Autonomous System (AS) Number Reservation for Documentation Use. RFC 5398, IETF, December 2008.

[9] Lixin Gao and J. Rexford. Stable Internet Routing Without Global Coordination. *IEEE/ACM Transactions on Networking*, 9(6):681–692, Dec 2001.

[10] Geoff Huston. The BGP World is Flat. `http://www.potaroo.net/ispcol/2011-12/flat.html`, December 2011.

[11] Alex Fabrikant, Umar Syed, and Jennifer Rexford. There's something about MRAI: Timing diversity can exponentially worsen BGP convergence. In *Proceedings of the IEEE INFOCOM 2011*, pages 2975–2983. IEEE, IEEE, 2011.

[12] Cisco IOS IP Routing: BGP Command Reference. `https://www.cisco.com/c/en/us/td/docs/ios-xml/ios/iproute_bgp/command/irg-cr-book/bgp-m1.html`, December 2019.

[13] C. Villamizar, R. Chandra, and R. Govindan. BGP Route Flap Damping. RFC 2439, IETF, November 1998.

[14] Tony Barber, Sean Doran, Daniel Karrenberg, Christian Panigl, and Joachim Schmitz.

RIPE Routing-WG Recommendation For Coordinated Route-flap Damping Parameters. RIPE Document ripe-178, RIPE, February 1998.

[15] Philip Smith and Christian Panigl. RIPE Routing-WG Recommendation For Coordinated Route-flap Damping Parameters. RIPE Document ripe-378, RIPE, May 2006.

[16] Cristel Pelsser, Olaf Maennel, Pradosh Mohapatra, Randy Bush, and Keyur Patel. Route Flap Damping Made Usable. In *Proc. of PAM Conf.*, volume 6579 of *LNCS*, pages 143–152, Berlin Heidelberg, 2011. Springer.

[17] Pavlos Sermpezis, Vasileios Kotronis, Petros Gigis, Xenofontas Dimitropoulos, Danilo Cicalese, Alistair King, and Alberto Dainotti. ARTEMIS: Neutralizing BGP hijacking within a minute. *IEEE/ACM Transactions on Networking*, 26(6):2471–2486, 2018.

[18] Taejoong Chung, Emile Aben, Tim Bruijnzeels, Balakrishnan Chandrasekaran, David Choffnes, Dave Levin, Bruce M. Maggs, Alan Mislove, Roland van Rijswijk-Deij, John Rula, and Nick Sullivan. RPKI is Coming of Age: A Longitudinal Study of RPKI Deployment and Invalid Route Origins. In *Proc. of ACM IMC*, pages 406–419, New York, NY, USA, 2019. ACM.

[19] Z. Morley Mao, Randy Bush, Timothy G. Griffin, and Matthew Roughan. BGP Beacons. In *Proc. of the 3rd ACM SIGCOMM Conf. on Internet Measurement (IMC'03)*, pages 1–14, New York, NY, USA, 2003. ACM.

[20] Romain Fontugne, Esteban Bautista, Colin Petrie, Yutaro Nomura, Patrice Abry, Paulo Gonçalves, Kensuke Fukuda, and Emile Aben. BGP Zombies: An Analysis of Beacons Stuck Routes. In *Proc. of PAM Conf.*, volume 11419 of *LNCS*, pages 197–209, Berlin Heidelberg, 2019. Springer.

[21] Andreas Reuter, Randy Bush, Italo Cunha, Ethan Katz-Bassett, Thomas C. Schmidt, and Matthias Wählisch. Towards a Rigorous Methodology for Measuring Adoption of RPKI Route Validation and Filtering. *ACM Sigcomm Computer Communication Review*, 48(1):19–27, January 2018.

[22] Florian Streibelt, Franziska Lichtblau, Robert Beverly, Anja Feldmann, Cristel Pelsser, Georgios Smaragdakis, and Randy Bush. BGP Communities: Even More Worms in the Routing Can. In *Proceedings of the Internet Measurement Conference 2018*, pages 279–292, New York, NY, USA, 2018. ACM.

[23] Chiara Orsini, Alistair King, Danilo Giordano, Vasileios Giotsas, and Alberto Dainotti. BGPStream: A Software Framework for Live and Historical BGP Data Analysis. In *Proc. of the 2016 Internet Measurement Conference*, IMC '16, pages 429–444, New York, NY, USA, 2016. ACM.

[24] Anja Feldmann, Olaf Maennel, Z. Morley Mao, Arthur Berger, and Bruce Maggs. Locating Internet Routing Instabilities. In *Proc. of ACM SIGCOMM*, pages 205–218, New York, NY, USA, 2004. ACM.

[25] Matthew Caesar, Lakshminarayanan Subramanian, and Randy H. Katz. Towards Localizing Root Causes of BGP Dynamics. Technical Report UCB/CSD-03-1292, EECS Department, University of California, Berkeley, 2003.

[26] Di-Fa Chang, Ramesh Govindan, and John Heidemann. Locating BGP Missing Routes Using Multiple Perspectives. In *Proc. of the ACM SIGCOMM Workshop on Network*

Troubleshooting: Research, Theory and Operations Practice Meet Malfunctioning Reality (NetT), pages 301–306, New York, NY, USA, 2004. ACM.

[27] RIPE NCC. Current RIS Routing Beacons. `https://www.ripe.net/analyse/internet-measurements/routing-information-service-ris/current-ris-routing-beacons`, 2020.

[28] CAIDA. AS Rank. `https://asrank.caida.org`, 2020.

[29] Ondřej Filip, Martin Mareš, Ondrej Zajíček, Jan Matějka, Libor Forst, and Pavel Machek. Bird Internet Routing Deamon. `https://bird.network.cz/`.

[30] ExaBGP. `https://github.com/Exa-Networks/exabgp`.