Towards a Rigorous Methodology for Measuring Adoption of RPKI Route Validation and Filtering

Andreas Reuter
Freie Universitaet Berlin
andreas.reuter@fu-berlin.de

Randy Bush
IIJ Research Lab / Dragon Research
randy@psg.com

Ítalo Cunha
Universidade Federal de Minas Gerais
cunha@dcc.ufmg.br

Ethan Katz-Bassett
USC / Columbia University
ethan.kb@usc.edu

Thomas C. Schmidt
HAW Hamburg
t.schmidt@haw-hamburg.de

Matthias Wählisch
Freie Universität Berlin
m.waehlisch@fu-berlin.de

ABSTRACT

A proposal to improve routing security—Route Origin Authorization (ROA)—has been standardized. A ROA specifies which network is allowed to announce a set of Internet destinations. While some networks now specify ROAs, little is known about whether other networks check routes they receive against these ROAs, a process known as Route Origin Validation (ROV). Which networks blindly accept invalid routes? Which reject them outright? Which de-preference them if alternatives exist?

Recent analysis attempts to use uncontrolled experiments to characterize ROV adoption by comparing valid routes and invalid routes [5]. However, we argue that gaining a solid understanding of ROV adoption is impossible using currently available data sets and techniques. Our measurements suggest that, although some ISPs are not observed using invalid routes in uncontrolled experiments, they are actually using different routes for (non-security) traffic engineering purposes, without performing ROV. We conclude with a description of a controlled, verifiable methodology for measuring ROV and present three ASes that do implement ROV, confirmed by operators.

1. INTRODUCTION

The Border Gateway Protocol (BGP) [17] is responsible for establishing Internet routes, yet it does not check that routes are valid. An autonomous system (AS) can hijack destinations it does not control by announcing invalid routes to them, either intentionally or unintentionally, as in the well-known accidental announcement of YouTube's address space by Pakistan Telecom [2].

Because this critical aspect of the Internet is vulnerable, there are proposals to improve routing security [7], and one—the RPKI—is standardized and is in early adoption. The Resource Public Key Infrastructure (RPKI) [12] is a specialized PKI to help secure Internet interdomain routing by providing attestation objects for Internet resource holders (*i.e.,* IP prefixes and AS numbers). The RPKI publishes Route Origin Authorization (ROA) objects, each specifying which AS is allowed to announce an IP prefix. Using ROA data, a BGP router can perform RPKI-based origin validation (ROV) verifying whether the AS originating an IP prefix announcement in BGP is authorized to do so [14] and labeling the route as valid or invalid. The validity of a route can be used as part of the router's local BGP policy decisions, *e.g.,* filtering routes that reflect invalid announcements or preferring valid ones. While the RPKI is fairly populated with ROAs and growing [9, 15, 23, 24], adoption of ROV and filtering has been negligible, according to operator gossip. A major reason for this is the lack of economic incentives. Since a significant share of invalid routes are due to misconfiguration [23], adopting ROV and filtering can even have adverse effects such as a loss of connectivity to legitimate network destinations.

A recent paper examined RPKI and ROV adoption from multiple angles, focusing on the slow state of ROV adoption, the security implications of partial adoption, and reasons for slow adoption [5]. To capture the current state of limited adoption, the paper included a measurement study that claimed that most large ASes had not deployed ROV, but that 9 of the 100 largest ASes had. This result was based on observations of existing BGP routes from BGP route collectors, meaning that the experiments were uncontrolled. At a basic level, the approach finds an AS that originates both valid and invalid announcements, then identifies other ASes that appear on paths towards the valid prefix but not on paths towards the invalid prefix. It then assumes these ASes are performing ROV to filter invalid routes.

In this paper, we demonstrate that the above approach to identify ROV adoption, based on passive observation of routes in uncontrolled experiments [5], has three major limitations. First, our measurements show that its characterizations of some networks change depending on which set of BGP collectors is used, inferring ROV adoption in some cases when it definitely has not been deployed and not inferring it in some cases when it may have been deployed. Second, the approach relies on invalid routes that happen to be announced, and so its coverage is limited by their rare nature [8]. Third, we conducted supplemental measurements that suggest that most networks flagged by the approach (and by [5]) as using ROV are actually avoiding invalid routes for unrelated (non-security) traffic engineering purposes, without checking ROV status, meaning that adoption is likely even lower than suggested by the earlier study. In fact, with only uncontrolled measurements of existing routes—the status quo for Internet research—it is impossible to differentiate between multiple feasible explanations.

To overcome challenges of measuring route origin validation (§ 2) and the limitations of uncontrolled experiments (§ 3), we propose a method to accurately infer ROV policies using controlled experiments (§ 4) that manipulate both BGP announcements and the ROAs that apply to them. We provide initial results using our method, verified by ground truth. Although ROV adoption is low and slow, our proposed method allows accurate, longitudinal observation of ROV adoption across the Internet.

2. THE CHALLENGES OF MEASURING ROUTE ORIGIN VALIDATION

Limited visibility. Measuring the deployment of ROV is challenging because of very limited visibility of routing decisions, which has multiple causes. First, an AS does not propagate every path it knows, instead selecting a best path to each destination prefix and then choosing for each neighbor whether to export that best path. So, BGP hides information by only forwarding a subset of available paths to a subset of neighbors. Second, an AS can use arbitrary policy to select a best path and to decide which neighbors to forward it to, and this policy is opaque. The policy may reflect concerns such as business relationships and traffic engineering, as well as route origin validity, and so it can be very difficult to discern the cause of any observed decision. Third, the interactions of these policies can influence the decisions of seemingly uninvolved ASes, meaning that it is not enough to observe a path before and after a change to understand which AS caused the change [10]. Fourth, as researchers, we typically have a limited view of the Internet, with projects such as RIPE RIS [18] and RouteViews [22] collecting routes from a small number of ASes, many of which only provide their routes to a limited set of destinations [16]. This makes it hard to locate where routes diverge or whether differences are due to actual filtering or simply lack of visibility.

Lack of controlled experiments. We distinguish between two experimental methods, controlled and uncontrolled. In a controlled experiment, researchers vary one factor of interest (whether a route is valid) while fixing other factors, then measure the outcome (which route an AS uses), observing how networks route under different scenarios of interest to the current research question [10, 21]. In an uncontrolled experiment, the factor of interest varies outside the control of the researchers and independent of the current research question (ASes on the Internet happen to announce a mix of valid and invalid routes), and researchers measure outcomes.

Our classification of controlled versus uncontrolled describes experiments (*how to test a hypothesis*). It is orthogonal to the classification of passive versus active measurements (*how data are collected*), and passive versus active measurements are orthogonal to control plane versus data plane measurements (*what data are collected*). With uncontrolled experiments, inferring root causes of routing decisions is challenging because pinpointing the reason for the decision (*e.g.,* RPKI policy or traffic engineering) is difficult when path attributes and RPKI data cannot be manipulated independently to observe their impact on decisions.

Implementation variations. Uncontrolled experiments are most challenged when the baseline of the system is unclear or complex. The deployment of ROV introduces additional variations in implementation and configuration (*e.g.,* ROA propagation delay [13], route revalidation because of ROA change) that have not yet been explored but likely affect measurement outcome.

3. REVISITING UNCONTROLLED EXPERIMENTS

A previous approach for detecting ROV deployment used uncontrolled passive measurements [5]. This study did not release the code and data sources needed to reproduce it, and unfortunately neither could be obtained after requesting it. In this section, we try to replicate some of the results and analyze how reliably the method leads to the conclusions. We show that the limited view provided by vantage points can lead to incorrect identification of ROV non-adoption and ROV adoption. Our analysis shows that differences in AS paths towards valid and invalid announcements are mainly a measurement artifact, instead of evidence for filtering.

To be clear on terminology, a *route collector* is a BGP router that peers with border routers of various ASes, each of which we refer to as a *vantage point* [20].

3.1 Uncontrolled Method

The previous approach uses available BGP dumps and RPKI data to estimate a lower bound for ROV non-adoption and identify ROV filtering [5]. It compares AS paths taken by known ROV valid and known ROV invalid announcements from a single AS to a single vantage point. If the paths differ, it assumes that the invalid announcement was filtered by ROV on the path taken by the valid announcement, causing the divergence. This approach does not distinguish between a single router or an entire AS using ROV-based filtering, since it makes inferences based on the ASes that appear on AS paths to vantage points. The method analyzes routes exported by vantage points as follows:

Exclude ASes observed to use invalid routes. First, any AS that is found on a path of an invalid announcement is flagged as *non ROV enforcing.* This assumes that any AS that accepts any invalid route accepts all invalid routes; *i.e.,* ASes do not implement selective filtering or use other policies that can accept some invalid routes while filtering others. An exception is made for invalid announcements originated by the vantage point's AS or by one of its customers [5], as an AS may make exceptions for its customers.

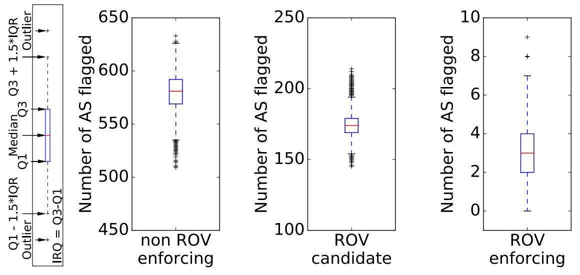
Identify ASes that may be performing ROV filtering. For each vantage point, the approach identifies all ASes observed to originate at least one non-invalid (either valid or not in the RPKI database) prefix announcement and at least one invalid announcement. It then compares each non-invalid path (from the origin to the vantage point) to each invalid path. If there is exactly one AS that (*i*) appears on the non-invalid path but not the invalid path, and (*ii*) has not been flagged as *non ROV enforcing*, the approach marks it as an *ROV candidate* for announcements from that origin.

For example, the vantage point V might observe the following paths for the non-invalid prefix announcements P_{1-2} and invalid prefix announcements P_{3-4} advertised from origin O:

$P_1 : O \rightarrow A \rightarrow C \rightarrow V$	`not found`
$P_2 : O \rightarrow A \rightarrow E \rightarrow V$	`valid`
$P_3 : O \rightarrow A \rightarrow D \rightarrow V$	`invalid`
$P_4 : O \rightarrow A \rightarrow D \rightarrow V$	`invalid`

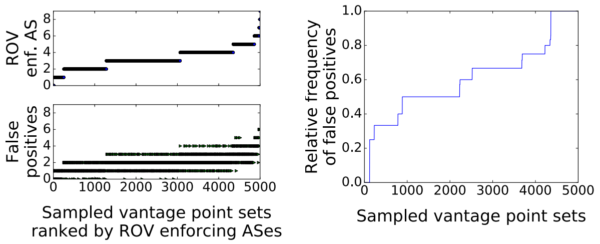
In this case, AS C and AS E are marked as *ROV candidates* for origin O, unless they have been previously marked as *non ROV enforcing.*

Select filtering ASes. The approach then counts the number of origins for which it marked an AS as an *ROV*



(a) (b) (c)

Figure 1: Uncontrolled, passive measurements: Statistical impact of vantage points on the number of identified ASes (5,000 samples of 44 randomly selected vantage points).



(a) AS flagged as ROV enforcing and number of false positives

(b) Relative frequency of false positives

Figure 2: Number of misclassified *ROV enforcing* ASes in different vantage point sets.

candidate and, following previous work [5], classifies an AS marked for at least 3 origins as *ROV enforcing*.

3.2 Data Set and Comparison with Current Findings

Data Set. The previous study [5] specifies the data set they have used to be from July 2016, collected from 44 Routeviews vantage points. It does not mention which vantage points explicitly and this information could not be obtained by us after multiple requests. Our analysis is based on BGP RIB dumps gathered from all route collectors of the RIPE RIS and Routeviews projects from October 25th 2016, 16:00 UTC. This data set includes 27GB of exported routes from 960 vantage points, a larger data set than the previous study has used.

Reproducing existing methodology. We have reproduced the methodology from the description in the previous study [5], since we could not obtain the original code. Analyzing our complete data set using the uncontrolled method, it classifies the following ASes as *ROV enforcing*:

$$AS8100\ AS25761\ AS17819\ AS262150$$

None of these ASes is among the top-100 ASes, based on the CAIDA AS rank [4]. This result differs from previous measurements [5], which used a different set of RIB dumps to conclude that 9 of the top 100 ASes enforce ROV. We want to better understand the validity of the method and why results vary significantly.

3.3 Impact of Limited Vantage Point Sets

When we run the same analysis on a subset of our data, such as data from a single route collector, the results differ. For example, the `routeviews-equix` collector has a feed from 34 vantage points, yet running the same analysis just on this feed results in zero ASes marked as *ROV enforcing*. In contrast, the `routeviews-wide` collector has feeds from only 4 vantage points, but shows the following ASes as *ROV enforcing*:

$$AS48237\ AS262150\ AS3786$$

Out of those 3 AS, AS48237 and AS3786 are both found on the AS paths of invalid routes when considering data from the `route-views4` collector, classifiying them as *non ROV enforcing*, contradicting the previous *ROV enforcing* classification. This shows that using the uncontrolled methodology some ASes might be (mis)classified as *ROV enforcing* if the invalid announcements they propagate are not visible in the data set, leading to false positives. On the other hand, some ASes that are classified as *ROV enforcing* in a more complete data set might not be visible enough in a smaller data set to result in a classification as *ROV enforcing*, leading to false negatives compared to the full data set. For example, using the complete data set, the approach marks AS8100 as an *ROV candidate* for origins AS6921, AS46562, and AS46261. It is thus flagged as *ROV enforcing.* When looking only at the data from the `routeviews-wide` collector, AS8100 is only marked as *ROV candidate* for a single origin, AS46562, and thus it is not classified as *ROV enforcing.*

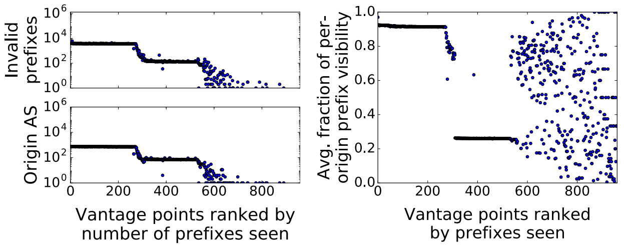
This shows that results vary significantly depending on which set of vantage points the data is taken from. To quantify the impact of this we select 44 Routeviews vantage points (the number used in previous work [5]) and calculate the number of ASes identified in each step of the method (see § 3.1). Figure 1 summarizes statistical properties (quartiles, extreme non-outliers, and outliers) of 5,000 random samples of 44 vantage points, showing that, even for a fixed number of vantage points, results can vary widely depending on which vantage points are used. Results for a single selection of vantage points may not reliably determine a lower bound of either deployment or non-deployment. Figure 2 depicts the number of false positives of *ROV enforcing* ASes, those classified as enforcing given sampled subsets of 44 vantage points but non-enforcing based on the global data set. For 82% of the samples, the ratio of false positives is 50% or more.

Conclusion. Using BGP RIB dumps as a basis for uncontrolled measurements of ROV filtering (or non-filtering) is problematic. It makes inferences based on routes visible in the selected dumps, but lacks complete visibility of the global Internet, leading to misclassification.

3.4 Impact of Limited Prefix Visibility at VPs

Recall that the existing approach to identify ROV filtering compares paths for invalid announcements with paths for non-invalid (*i.e.,* valid or unknown) announcements. We have shown that the selection of vantage points has a major impact on classification using this approach. As the approach uses pairs of non-invalid and invalid announcements, it relies on vantage points receiving such announcements from enough origins to reveal their policies.

Combining all dumps from the RIPE RIS and Routeviews



(a) Invalid prefix announcements

(b) Relative prefix completeness seen per vantage point

Figure 3: Number of prefixes and origin ASes observed by RIPE and Routeviews.

projects, we have data from 960 vantage points. But, not all vantage points provide routes to the same set of prefixes. Some vantage points have a near global view, while some have routes for only a very limited number of prefixes.

For each vantage point, Figure 3(a) shows the number of prefixes received via invalid announcements (top) and the number of distinct origins originating these announcements (bottom). Though some vantage points provide routes for invalid prefix announcements to nearly 1000 distinct origin ASes, more than 36% of the vantage points see less than the needed 3 ASes originating invalid prefix announcements. This observation is independent of the RPKI deployment state. Figure 3(b) shows the relative ratio of visible prefixes per origin and vantage point. Those vantage points that see many prefixes lack a complete view with respect to all prefixes per origin.

Conclusion. Assuming one applies the method with only a subset of VPs as in the previous work [5], selecting vantage points with very limited prefix visibility misses a significant portion of origin ASes, and thus underestimates the set of *ROV candidates* and can lead to misclassification.

3.5 Impact of Limited Control

Just because a vantage point uses different routes to reach a non-invalid and an invalid prefix from the same origin does not imply that the difference is caused by ROV-based filtering, as invalid and non-invalid advertisements might differ in attributes other than RPKI validity. We now investigate traffic engineering as another possible explanation (unrelated to BGP security) for observed differences. For a multi-homed AS, a common technique to influence inbound traffic is to announce different prefixes to different upstreams. These prefixes often overlap, *e.g.,* an AS may announce a more specific prefix (a /24) via upstream A and the covering prefix (a /16) via upstream B to shift traffic to A. Studies comparing current ROAs to announced prefixes have shown that the major cause for invalid BGP announcements is issuing a ROA only for a prefix and then announcing subprefixes [6, 9, 23] which are not covered by the ROA. Announcing the /16 and /24 to separate providers then results in two routes, one valid and one invalid, diverging on the first hop of the AS path.

For each vantage point, Figure 4 shows the fraction of prefixes from invalid announcements that are covered by a prefix from a non-invalid announcement from the same origin. An invalid prefix only counts as *covered* if the vantage point sees both the route to the invalid prefix and a route for the covering non-invalid prefix. For the vantage points between $x = [0,\ 275]$, roughly 80% of prefixes from invalid announcements are covered by a non-invalid from the same origin. This strongly suggests that the prefixes are invalid because of incorrect ROA configuration and the announcements perhaps subject to traffic engineering.

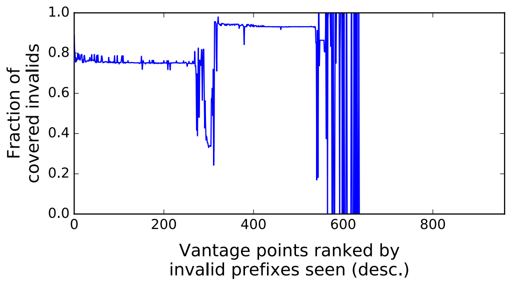


Figure 4: Fraction of invalid prefixes covered by a valid less-specific prefix from the same origin.

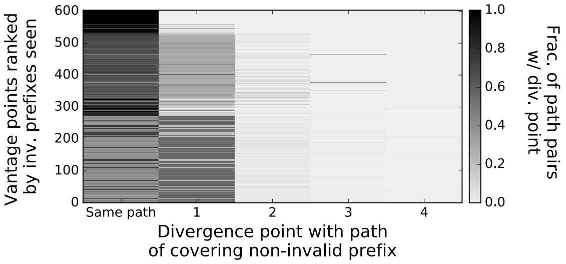


Figure 5: Divergence point distribution of invalid prefixes with covering non-invalid prefixes of same origin

Next, we investigate where the vantage points' paths to these invalid prefixes diverge from their paths to the covering prefixes. Figure 5 shows the distribution of divergence points for all vantage points. The y-axis sorts the vantage points by the number of invalid prefixes they provide routes to. The coloring of the x-axis depicts the fraction of these paths that diverged a given number of hops from the origin. The majority of AS paths of invalid routes either share the AS path of the covering non-invalid (x="Same path") or diverge at the first hop, as would occur with traffic engineering.

Conclusion. ROV-based filtering is not the only plausible explanation for instances of vantage points using different routes to reach non-invalid and invalid prefixes from the same origin. We found that most instances display signatures of traffic engineering, and, during our study, we also observed a router selecting different routes from the same origin AS due to route age (a BGP tiebreaker).

4. CONTROLLED EXPERIMENTS

With uncontrolled, passive experiments, it can be impossible to determine whether an AS is actually filtering or whether it is not using invalid advertisements because of other attributes. Further, the AS where the divergence occurred need not be the one that made a different decision, as it could have been presented with different options for its decisions. To overcome misclassification, experiments must clearly establish whether decisions stem from ROA status.

Controlled experiments provide a means to establish this causation despite our limited visiblity into routing decisions. Based on the challenges in measuring ROV adoption (§2) and our experiences evaluating the existing approach (§3), we arrive at the following requirements for a more reliable methodology.

Experiments must be long-lived. Adoption is likely to be slow, and may be bursty, driven by various initiatives and technologies. Measurements must be rerun periodically.

Experiments must be active and controlled. Passive observations of existing announcements are insufficient to determine a policy since we do not know precisely how the announcements are being made (*i.e.,* traffic engineering or not) and route colletors may not provide the right vantage to locate filtering. Furthermore, we need to coordinate announcements with ROA changes to precisely expose policies.

Experiments require rich BGP connectivity. From a single vantage point, it is difficult to infer which network along a path is filtering an announcement.

4.1 Basic Approach

We describe an approach based on active, controlled manipulation of BGP announcements and RPKI ROAs. We use the PEERING testbed, which allows us to make BGP announcements for prefixes we control from PEERING sites around the world to the hundreds of networks it peers with [21]. We use multiple /24 prefixes from the same /16 block. These prefixes share the same route object in the Internet Routing Registry. To control ROAs, we run a grandchild RPKI Certificate Authority (CA) in the RIPE region, enabling us to programmatically issue and revoke Resource Certificates and ROAs. To guard against uncommonly long ROA propagation delays, we conservatively keep every configuration (set of BGP announcements and ROA states) in place for eight hours.

In our basic approach, an AS must fulfill two assumptions to allow us to unambiguously determine whether the AS is using ROV-based filtering: (*i*) *connected-assumption.* The network peers with PEERING, either directly or using a route server. (*ii*) *visibility-assumption.* The network offers some means to check the BGP route is uses to reach an Internet destination, either via a Looking Glass or via a vantage point. While the connected assumption is limiting, it is necessary to maintain accuracy, relaxing it to allow networks that are not peers of PEERING introduces ambiguity. We discuss the possibility of relaxing the connected assumption, as well as the visibility assumption, in section 4.3.

We announce two prefixes via PEERING (AS47065), a *reference prefix* P_R and an *experiment prefix* P_E. We periodically change RPKI state for the experiment prefix, using an additional origin AS O to alternate between the following configurations:

(C1) ROA specifies AS47065 is `valid` for P_R and P_E, so both announcements are valid.
(C2) ROA specifies AS47065 is `valid` for P_R, AS O is `valid` for P_E. AS47065's announcement of P_E is invalid.

We check the routes a vantage point chooses to both prefixes during both configurations. The reference prefix always has a *valid* RPKI state so should not be filtered via ROV, and so we omit any vantage points at which P_R is not visible. We expect both prefixes to be treated the same as long as both announcements are valid, and so we omit a vantage point if it uses different routes during configuration C1. Analysing only data from vantage points that pass both these requirements eliminates the problem of **limited visibility**, since there is no missing data anymore. We then check the routes a vantage point has chosen after the announcement of the expermiment prefix becomes *invalid.* Three observations might occur: *(O1)* V has the same route for both prefixes P_E and P_R. *(O2)* V has a different route for prefix P_E. *(O3)* V has no route to P_E.

In the cases of O2 and O3, we know that this route change *must* be because of the RPKI status change. Had it been for another reason we would expect a change in route for the reference prefix as well. The reference prefix combined with the ROA changes thus all but eliminates the problem of **limited control**. The experiments are repeated continuously to confirm the behaviour is consistent.

Experiment Reach. The experiments were conducted using PEERING BGP routers in Amsterdam and Seattle. The device in Amsterdam peers with 589 different AS, either directly or via a routeserver at AMS-IX. The device in Seattle peers with 179 different AS either directly or via a routeserver at SIX. In total, via these two location PEERING peers with 730 AS. Out of these 730 AS, only 138 AS peer with a RIPE RIS or Routeviews route collector. Out of those 138 vantage points, 68 actually export direct routes for prefixes announced by PEERING.

Results. These experiments were performed Febuary 20-27, May 11-17, and August 1-7, 2017. In our experiments in February and May 2017, we found AS8283, AS50300, and AS59715 were using ROV to filter invalid announcements. AS8283 and AS50300 comply with both of our assumptions. The experiments in August show AS50300 and AS59715 to be filtering, but not AS8283.

AS8283 was identified based on observation (O3), and AS50300 based on (O2). It is worth noting that AS50300 only filtered routes learned via a route server at the Amsterdam exchange (AMSIX). This contradicts one of the assumption in the methodology studied in section 3, whereas it is assumed that an AS found on the AS path of an invalid route does not use ROV based filtering.

AS59715 was not directly connected to PEERING but lead to (O3). For all three AS we contacted the operators via email and they confirmed that they used ROV based filtering. In the case of AS8283, they confirmed that they had shut off ROV based filtering for technical reasons in July 2017. This confirms our findings from August 2017.

Relaxing the connected-assumption in the basic approach lead to ambiguity since multiple ASes can be on the path between PEERING and a vantage point. To deal with cases such as AS59715 precisely, we propose a roadmap for a more general approach in the next section.

4.2 Operational Concerns

ROA Propagation Time. Analysis of our experiments has shown that the time for some AS to receive newly issued ROAs can be up to 8 hours or more. We have also observed that the propagation time for some AS is inconsistent and varies by up to 2 hours. It is not clear yet whether this is by RPKI cache servers updating infrequently or by BGP routers using excessively long refresh intervals.

Considering implementation variations. Active RPKI experiments require a careful check of router implementations [3]. For a router to perform ROV when an existing route changes from valid to invalid (due to an RPKI change), the BGP implementation must (*i*) receive the new ROA payload and (*ii*) recalculate the best path for this existing entry. We verified that Cisco and Juniper implementations do recalculate the best path upon ROA changes, however there are corner cases where certain Cisco implementations do not re-apply route-maps that change BGP path attributes based on RPKI validation. This might lead to a filtering AS to go unnoticed by our basic approach. To detect such cases we have set up a second set of experiments in which the BGP announcements are withdrawn prior to the ROA changes and then reannounced once the new ROAs have propagated. So far we have found no additional AS to be filtering with these experiments.

Analyzing router implementations in more detail, analyzing the consistency among RPKI cache servers, and measuring ROA propagation time to routers is part of our ongoing work.

4.3 Roadmap for a More General Approach

Going forward, we plan to generalize our approach by conducting additional experiments, but also relaxing our assumptions without sacrificing the precise conclusions enabled by tightly controlled experiments.

Relaxing `connected-assumption`. Suppose the target does not connect directly to PEERING and has no route to P_E. It might check ROAs or might not receive a route from any neighbor. To narrow our policy inferences, we use two techniques. First, we iteratively target networks in a breadth-first search outwards from a PEERING site, similar to an approach that we used to uncover (non-security-related) routing policies [1]. Second, we will make multiple observations and only consider inferences consistent with all observations. We will make multiple observations both by using vantage points across the Internet and by targeting a network with different announcements. We can vary the announcements by changing which PEERING sites we use, which peers we announce to, and what BGP attributes we use to influence route selection and propagation.

Relaxing `visibility-assumption`. Lacking a BGP feed from a network, we can measure the data plane. This is straightforward if it has a traceroute server or RIPE Atlas probe [19]. If not, we can ping a destination in the target network and check the PEERING site the reply arrives at, or use our Reverse Traceroute [11].

Inferring complex RPKI policies. A network may prefer valid routes over invalid but not drop invalid routes. In order to test for such policies, experiments must fulfil an additional requirement:

Experiments require competing announcements. To identify `prefer-valid` policies, we need multiple simultaneous announcements for the same addresses. Since a single BGP session generally allows only a single announcement, the experiment must include sessions with multiple peers.

In order to test for such policies we announce two prefixes P_R and P_E identically, each from two different locations with two different ASN (61575 and 61576). This means that for both prefixes there exist competing announcements. All announcements for the reference prefix P_R will be valid throughout the experiment, while the announcements for the experiment prefix P_E will vary like this:

(C1) ROA specifies AS61575. Announcement of P_E from AS61575 is `valid`, from AS61576 is `invalid`.

(C2) ROA specifies AS61576. Announcement of P_E from AS61576 is `valid`, from AS61575 is `invalid`.

A vantage point might choose the route to AS61576 for prefix P_R. If the vantage point chooses the route to AS61575 for prefix P_E during configuration C1 this indicates a preference for valid routes over invalid routes. An even stronger indicator of this policy is when the vantage point then switches its route for P_E to AS61576 when configuration C2 begins. This reasoning works the same way if the vantage point chooses the route to AS61575 for prefix P_R.

We can differentiate these policies by configuring announcements from PEERING in such a way that a target network receives different combinations of `valid` and `invalid` prefixes through clients, peers, and providers, then observing its decisions.

There are subtleties in checking `prefer-valid` policies, as a network is "allowed" to use ROA status as one part of checking how preferred a path is, but, for example, it may prefer invalid peer routes over valid provider routes but not over valid peer routes. We will explore how best to capture these policies, building on our work on using PEERING to uncover (non-security-related) routing policies [1].

5. CONCLUSION

In this paper, we discussed steps and results towards a rigorous methodology for measuring adoption of RPKI route validation and filtering. We showed that BGP data sets that are incomplete with respect to peering relations—as are all available public data sets—challenge any method based on passive uncontrolled experiments. We discussed several pitfalls. We identified that traffic engineering, combined with negligent ROA configurations, are largely responsible for the routing differences between invalids and non-invalids. To allow for more solid conclusions, we argue for controlled experiments. In fact, our ongoing measurements revealed three ASes that already deploy RPKI-based filtering, which has been confirmed by the operators.

We will rerun our measurements on a weekly basis, providing a public monitoring platform that uses our methods and reports the ongoing deployment of RPKI-based route origin validation and filtering. First, by controlling our own announcements, we can uncover policies proactively, yielding a richer understanding of adoption and configurations than is possible via passive observation of existing announcements; and potentially uncovering issues before they would otherwise manifest. Second, with a longitudinal rather than one-off study, we can evaluate the impact of efforts that, for example, routing registries, Internet exchange points, vendor updates, and operator organizations make to encourage BGP security adoption. Measurements of the effect of such campaigns may yield a better understanding of how to spur uptake. Third, with an Internet-wide characterization, our data may inform best practices, encourage adoption, and reveal topics worthy of study, and provide the basis for understanding overall coverage and effectiveness.

5.1 Reproducibility

We make all code as well as data used for both our attempt at replication of the uncontrolled methodology of [5] as well as our presented controlled methodology available at `https://github.com/RPKI/rov-measurement-code`.

6. REFERENCES

[1] R. Anwar, H. Niaz, D. Choffnes, Ítalo Cunha, P. Gill, and E. Katz-Bassett. Investigating Interdomain

Routing Policies in the Wild. In *Proc. of ACM IMC*, pages 71–77, New York, NY, USA, 2015. ACM.

[2] M. A. Brown. Pakistan hijacks YouTube – Renesys Blog, February 2008.

[3] R. Bush, R. Austein, K. Patel, H. Gredler, and M. Waehlisch. Resource Public Key Infrastructure (RPKI) Router Implementation Report. RFC 7128, IETF, February 2014.

[4] CAIDA. AS Rank. http://as-rank.caida.org/data/, 2012.

[5] Y. Gilad, A. Cohen, A. Herzberg, M. Schapira, and H. Shulman. Are We There Yet? On RPKI's Deployment and Security. In *Proc. of NDSS*. ISOC, 2017.

[6] Y. Gilad, S. Goldberg, and K. Sriram. The Use of Maxlength in the RPKI. Internet-Draft – work in progress 00, IETF, March 2017.

[7] S. Goldberg. Why is It Taking So Long to Secure Internet Routing? *Commun. ACM*, 57(10):56–63, Sept. 2014.

[8] E. Heilman, D. Cooper, L. Reyzin, and S. Goldberg. From the consent of the routed: Improving the transparency of the RPKI. In *Proc. of ACM SIGCOMM*, pages 51–62, New York, NY, USA, 2014. ACM.

[9] D. Iamartino, C. Pelsser, and R. Bush. Measuring BGP route origin registration validation. In *Proc. of PAM*, LNCS, pages 28–40, Berlin, 2015. Springer.

[10] U. Javed, I. Cunha, D. Choffnes, E. Katz-Bassett, T. Anderson, and A. Krishnamurthy. PoiRoot: Investigating the Root Cause of Interdomain Path Changes. In *Proc. of ACM SIGCOMM*, SIGCOMM '13, pages 183–194, New York, NY, USA, 2013. ACM.

[11] E. Katz-Bassett, H. V. Madhyastha, V. K. Adhikari, J. S. C. Scott, P. van Wesep, T. Anderson, and A. Krishnamurthy. Reverse Traceroute. In *Proc. of NSDI*. USENIX, 2010.

[12] M. Lepinski and S. Kent. An Infrastructure to Support Secure Internet Routing. RFC 6480, IETF, February 2012.

[13] O. Maennel, I. Phillips, D. Perouli, R. Bush, R. Austein, and A. Jaboldinov. Towards a Framework for Evaluating BGP Security. In *Proc. of 5th USENIX Workshop CSET*. USENIX, 2012.

[14] P. Mohapatra, J. Scudder, D. Ward, R. Bush, and R. Austein. BGP Prefix Origin Validation. RFC 6811, IETF, January 2013.

[15] NIST. NIST RPKI Deployment Monitor. `http://rpki-monitor.antd.nist.gov/`, 2015.

[16] R. V. Oliveira, D. Pei, W. Willinger, B. Zhang, and L. Zhang. In Search of the Elusive Ground Truth: The Internet's AS-level Connectivity Structure. In *Proc. of ACM SIGMETRICS*, pages 217–228, New York, NY, USA, 2008. ACM.

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

[18] RIPE. RIPE Routing Information Service (RIS). `http://www.ripe.net/projects/ris/rawdata.html`, 2017.

[19] RIPE NCC. What is RIPE Atlas?

[20] M. Roughan, W. Willinger, O. Maennel, D. Perouli, and R. Bush. 10 Lessons from 10 Years of Measuring and Modeling the Internet's Autonomous Systems. *IEEE Journal on Selected Areas in Communications*, 29(9):1810–1821, 2011.

[21] B. Schlinker, K. Zarifis, I. Cunha, N. Feamster, and E. Katz-Bassett. PEERING: An AS for Us. In *Proceedings of the 13th ACM Workshop on Hot Topics in Networks*, HotNets-XIII, pages 18:1–18:7, New York, NY, USA, 2014. ACM.

[22] University of Oregon. Route Views Project. `http://www.routeviews.org/`, 2017.

[23] M. Wählisch, O. Maennel, and T. C. Schmidt. Towards Detecting BGP Route Hijacking using the RPKI. *ACM CCR*, pages 103–104, October 2012.

[24] M. Wählisch, R. Schmidt, T. C. Schmidt, O. Maennel, S. Uhlig, and G. Tyson. RiPKI: The Tragic Story of RPKI Deployment in the Web Ecosystem. In *Proc. of 14th ACM Workshop on Hot Topics in Networks (HotNets)*, pages 11:1–11:7, New York, Nov. 2015. ACM.