Right the Ship: Assessing the Legitimacy of Invalid Routes in RPKI

Andong Chen
Tsinghua University
Beijing, China
chenanton@outlook.com

Yangyang Wang
Tsinghua University; Zhongguancun Laboratory
Beijing, China
wangyy-13@tsinghua.edu.cn

Jia Zhang
Zhongguancun Laboratory
Beijing, China
zhangj@zgclab.edu.cn

Mingwei Xu
Tsinghua University; Zhongguancun Laboratory
Beijing, China
xumw@tsinghua.edu.cn

Abstract

Resource Public Key Infrastructure (RPKI) aims to prevent prefix hijacking by providing secure mappings between IP prefixes and their authorized origin Autonomous Systems (ASes). In recent years, there has been notable growth in the deployment of RPKI and Route Origin Validation (ROV). Nonetheless, over 40% of the routes in the global routing table still lack the protection of RPKI. One of the critical reasons some networks are reluctant to deploy RPKI is the concern that some ROV-invalid routes may be legitimate, and filtering these routes will harm network service quality, especially affecting network connectivity.

In this work, we perform a comprehensive measurement study to assess the legitimacy of ROV-invalid routes in RPKI. We evaluate the impact of filtering all ROV-invalid routes in the global routing table, presenting a view that some ROV-invalid routes are not illegitimate, defined as harmlessly ROV-invalid (h-invalid). We propose five characteristics and design a characteristics-based methodology for identifying h-invalid routes. Based on the methodology, we analyze the magnitude of h-invalid routes present on the Internet each day, revealing that over 91% of ROV-invalid results are h-invalid. Furthermore, we conclude three main reasons for h-invalid routes. Finally, with all our findings, we provide practical recommendations for network operators to help promote RPKI deployment.

CCS Concepts

• **Security and privacy** → **Network security**.

Keywords

RPKI; ROV; BGP

ACM Reference Format:
Andong Chen, Yangyang Wang, Jia Zhang, and Mingwei Xu. 2025. Right the Ship: Assessing the Legitimacy of Invalid Routes in RPKI. In *Proceedings of the 2025 ACM SIGSAC Conference on Computer and Communications Security (CCS '25), October 13–17, 2025, Taipei.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3719027.3744853

CCS '25, Taipei

ACM ISBN 979-8-4007-1525-9/2025/10
https://doi.org/10.1145/3719027.3744853

1 Introduction

The Internet consists of Autonomous Systems (ASes) that exchange routing information with the Border Gateway Protocol (BGP). Since BGP is designed without security considerations, routers cannot authenticate the BGP update messages received from neighbors. This inherent vulnerability makes it fairly easy to launch various attack vectors on the critical routing infrastructure of the Internet, resulting in an AS being disconnected from the Internet or traffic interception. One of the serious threats is known as route hijacking. A common form of this threat [64] [57], prefix hijacking, can occur due to either BGP misconfigurations [62] or malicious attack attempts [57] [64] [59] [1] [42] [17].

Nowadays, Resource Public Key Infrastructure (RPKI) is one of the most prominent mechanisms for validating route origin [44]. In RPKI, Regional Internet Registries (RIRs) dominate the root Certificate Authorities (CAs) issuing resource certificates (RCs) that contain the allocated address resource to subordinate CAs, including National Internet Registry (NIR), Local Internet Registry (LIR), and Internet Service Provider (ISP), etc. CAs use these RCs to issue Route Origin Authorizations (ROAs) to certify ASes and their IP address prefixes authorized to be originated. Based on the ROAs, routers on the Internet validate incoming BGP announcements through Route Origin Validation (ROV), which validates the origin of each route as one of valid, invalid, or unknown [18, 41].

In the current RPKI deployment, approximately 5% (over 66K) of the routes in the global BGP routing table are ROV-invalid each day. It is unreasonable to consider all of the ROV-invalid routes to be route hijacking in terms of the quantity and duration [35, 61]. Some studies [22, 25, 35] indicate that certain ROV-invalid routes may not be malicious hijacks, but rather a result of mistakes in the ROA registry. For example, the customer AS announces sub-allocated prefixes while the corresponding ROAs have not yet been updated, which is also common for DDoS mitigation and traffic engineering [22]. The negative effects of the false positives in ROV-invalid routes could lead to network disconnection and significantly hinder the deployment of RPKI [25, 47]. If all ROV-invalid routes were rejected, a portion of prefixes would disappear from the global routing table, which has a significant impact on routing security.

Our work aims to assess the legitimacy of ROV-invalid routes in the global routing tables, to extract **harmlessly ROV-invalid** (*h-invalid*) routes that have evidence of legitimate announcement or do not pose a threat to other routes, particularly those ROV-valid cases.

The h-invalid routes preclude the possibility of malicious BGP prefix hijacks. Apart from the h-invalid routes within the ROV-invalid routes, the remaining ones are classified as **suspiciously ROV-invalid** (*s-invalid*) routes. These routes lack evidence to support their legitimacy and have a detrimental impact on other routes. We conduct an analysis of BGP route and ROA data over 17 months, innovating in temporal duration, breadth of routing data, and depth of cause analysis. Specifically, we quantitatively analyze several incidents of invalid route surges, summarize the causes of h-invalid routes, and propose recommendations. These contributions help the community better understand the current status of RPKI and promote its deployment.

The measurement relies on BGP route data and ROA retrieved from public platforms, with no effect on the Internet and the privacy data of real users. Consequently, our work does not raise any ethical issues. Our measurement results and dataset are publicly available at https://github.com/H-invalid/H-invalid/.

By assessing ROV-invalid routes in the global routing table, we respond to networks' concerns about RPKI deployment: how many ROV-invalid routes exist on the Internet that are illegitimate? Our fundamental contributions are summarized below:

- A characteristics-based methodology: We delineate five characteristics in terms of AS topology, relationship, and dynamic AS-path, and propose a characteristics-based methodology for the assessment of ROV-invalid routes. The validation of this methodology is demonstrated through the use of two ground-truth datasets and a comparative experiment.
- Measurement of h-invalid routes: Starting in July 2023, we conducted a 17-month measurement of the BGP data published from three public services and the corresponding ROAs, capturing over 1.4 million route pairs daily, of which approximately 5% were identified as ROV-invalid route pairs. We find that at least 91% of ROV-invalid routes each day are h-invalid, with the majority of these routes exhibiting strong evidence of h-invalid status.
- Extending the findings of h-invalid routes: We conducted an in-depth analysis of the temporal evolution and types of h-invalid routes. Based on our findings, we summarized three primary causes, including (1) mismatch between Maxlength and BGP configuration, (2) address transfer within ASes with business relationships, and (3) AS-path re-origination.

2 Background

In this section, we first introduce prefix hijacking and ROV, and then we provide a unified description of the key concepts that will be discussed later to help with understanding (§2.1). Finally, we provide an in-depth analysis of the status of invalid routes. (§2.2).

2.1 Prefix Hijacking and ROV

Prefix hijacking. Prefix hijacking forwards traffic to an illegitimate destination AS, which seriously threatens routing security. Depending on the hijacking target, prefix hijacking is divided into 2 categories: exact-prefix hijacking and sub-prefix hijacking. For instance, if AS65536 announces a route for 192.0.2.0/24, and AS64500 attempts to hijack the matched traffic by a route for the same prefix, it belongs to the exact-prefix hijacking; while if AS64500 announces the route for 192.0.2.0/26, it belongs to the sub-prefix hijacking.

Route Origin Validation (ROV). Network operators use relying party (RP) software to fetch the ROAs from RPKI repositories and cache the Validated ROA Payloads (VRPs) [41] containing (AS, $prefix$, $Maxlength$). The ROV-enforcing router validates incoming BGP announcements by retrieving VRPs from its RP cache. If the prefix of the BGP announcement is not covered by any VRP prefix (exact or sub-prefix), the validation status is ROV-unknown. When at least one VRP prefix covers the prefix of the BGP announcement, the prefix length does not exceed VRP Maxlength, and the origin AS is identical to the VRP AS, it indicates that the BGP announcement is validated as ROV-valid; Otherwise, the validation status is ROV-invalid. It is pertinent to highlight that ROA does not protect against AS-in-the-middle attacks or provide any path validation. It only attempts to validate whether the route origin is legitimate or not [41]. To facilitate a better understanding, we summarize the description of concepts used subsequently:

- **Route pair**: A route pair (origin AS, prefix) is identified by the origin AS and prefix derived from each entry in the global routing table. Entries with the same origin AS and the same prefix are the same route pair, even if they have different AS-paths. We use the route for short to represent a route pair in this paper.
- **ROV status**: To clearly distinguish from h-invalid, we use ROV-valid/invalid/unknown, as stated in §2.1, to refer to the result of ROV (abbreviated as valid/invalid/unknown for the rest), collectively called ROV status.
- **Covering VRP (CVRP)**: We define a CVRP as a VRP that covers the route prefix, whether it matches or not. CVRP is relative to a specific route. One single VRP could be the CVRP of multiple routes, and a single route may also have multiple CVRPs.
- **Route entity**: A route entity is the combination of route (pair), AS-paths, ROV status, and CVRPs, targeted as the input of our characteristics-based methodology.

2.2 Analysis of Invalid Route Status

Not all invalid routes should be considered malicious hijacks. BGP hijacks typically are short-lived compared to legitimate routes. Previous works quantified the discrepancy in duration between hijacks and legitimate routes, such as hijack duration median is 27.25 days versus 264.17 days for legitimate routes [61], and hijacks are mostly short-lived, lasting less than 3 weeks [35]. Differentiating hijacks and legitimate routes precisely by setting a threshold that separates short-lived and long-lived routes is challenging, but we still can summarise the following point from a 275-day measurement of invalid routes' duration, as shown in Fig. 1. Most invalid routes exist for a long period, which is closer to legitimate routes than hijacks. From our duration measurement of invalid routes on October 1, 2023, 99.17% of invalid routes were stable for more than 21 days (time-based resolution threshold in [35]). Furthermore, 98.86% of invalid routes exist for more than 28 days (near the median duration of hijacks in [61]). Given this perspective, most invalid routes should not be considered malicious hijacks.

Rejecting all invalid routes as hijacks would lead to several potential issues. In general, routers may filter invalid routes based on the ROV status, such as rejecting or adjusting preference degree to make correct routing selections in BGP [41]. To better understand the impact of treating all invalid routes as hijacks, we

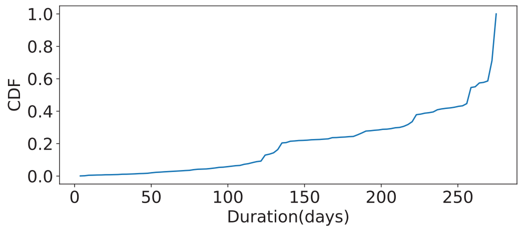


Figure 1: Duration of invalid routes (October 1, 2023) in 275-day measurement, from July 1, 2023, to March 31, 2024.

implement a reject policy for these routes. We measure the effects and present the results of Oct. 1, 2023, in Tab. 1, utilizing global routing table data acquired from RIB (RouteViews [63], RIPE RIS [55]) and snapshots (PCH [36]) files generated at 0:00 (UTC). We find that the origin AS of 40.5% of invalid routes announced a valid parent prefix propagated along the same AS-path. They are also the sole AS in the global routing table that announces the address resource. Rejecting these invalid routes does not block their propagation. Moreover, 19.9% of invalid routes have different propagation paths from the valid parent prefix announced by their origin AS. Rejecting them will block alternative propagation paths that could potentially disrupt traffic engineering. Furthermore, rejecting the remaining invalid routes may cause partial traffic to be forwarded to different destination ASes (7.9%) or the prefix to become unreachable since there were no other valid routes covering the prefix (31.7%). Therefore, rejecting all invalid routes can significantly impact routing connectivity and the implementation of routing policies.

The aforementioned two considerations highlight the necessity to assess the legitimacy of invalid routes.

3 Related Work

Previous studies [16, 60] and real-world hijacking incidents [1, 29] have highlighted the risks associated with route hijacking. There is a growing consensus that the deployment of RPKI/ROV is crucial to address these concerns. And, there has been a surge in research activities focusing on RPKI/ROV topics, including the measurement of RPKI/ROV deployment [21, 22, 25, 34, 39, 43, 48] and studies in the vulnerability or limitation of RPKI/ROV [27, 32, 33, 40]. Next, we put our work against the background of related work on measurements and improvement of ROV.

Measurement of ROV. The first measurement of ROV was conducted in 2017 by Gilad et al. [25], which found 100 ROV-enforcing ASes. Subsequently, the study in 2018 [31] found 296 ROV-enforcing ASes, and the study in 2021 [53] got 3.6K ROV-enforcing ASes by using the same methodology with more probes. In 2022, a total of 2.4K and 3.4K ROV-enforcing ASes were identified in [34] and [21], respectively. Such ROV measurements exemplify the sustained expansion of ROV coverage. Meanwhile, the expansion of coverage also attracts researchers to focus on the side-effects of ROV, such as misconfigurations of ROA [22, 25], technical bugs, and limitations of ROV [48], which may produce false positives. The issue of false positives, in particular, has a direct impact on routing security and is a significant concern for network operators when considering the deployment of ROV.

Improvement for ROV. Chung et al. [22] suggested that not all invalid BGP announcements with wrong ASes are hijacking attempts and listed 3 causes of representative misconfigurations: two different ASes managed by the same operator, provider–customer relationship, and DDoS protection, which are representative ROV false positives. To reduce false positives of ROV, Hlavacek et al. [35] proposed SROV as an extension of the Routinator [45] implementation. The SROV combines the time-based solution and Aggregated Resolution Algorithms (ARA) to identify benign conflicts between the BGP announcement and the ROA. However, there are some limitations to this methodology, which distinguishes benign conflicts and malicious hijacks based on duration threshold and assigned ARA scores. On the one hand, the reliability of ARA scoring is questionable because its criteria mostly depend on AS registration information, but not all AS scoring criteria are fully recorded in Internet number registries (IRRs) Whois datasets (i.e., AFRINIC [11], APNIC [12], ARIN [13], RIPE [54], LACNIC [52]). For instance, criteria like 'responsible', 'phone', and 'postalcode'. This means it is not always possible to obtain the full range of information that the ARA requires with ease, resulting in an imprecise score. Moreover, determining an appropriate duration threshold can be equally challenging. On the other hand, the SROV, as an extension tool of RP, provides a method to reduce false positives for a single operator, which focuses on a single conflicting pair of BGP announcements and ROA. In contrast, our methodology offers the legitimacy assessment of invalid routes in the global routing table, covering a broader network scope and enabling the use of more comprehensive routing state information, such as AS-path and traffic competitions between BGP routes.

Table 1: Impact of rejecting invalid routes.

Impact	Proportion	Count
Maintain Connectivity	40.5%	29.9K
Reduce Alternative Propagation Paths	19.9%	14.8K
Different Destination AS	7.9%	5.8K
Unreachable Prefix	31.7%	23.4K

4 Characteristics Derivation

In this section, we first categorize the types of CVRP based on the results of matching BGP routes with CVRP (§4.1). Subsequently, we introduce the derivation and limitations of five characteristics through systematic empirical analysis: same-organization (§4.2), path-neighbor (§4.3), traffic shunt point (§4.4), path-overlap (§4.5), and competitor (§4.6). These five characteristics form the basis of our methodology for assessing invalid routes.

4.1 CVRP Category

In previous work [22], ROV false positives are analyzed according to two categories of invalid routes: wrong AS and too-specific announcements. Similarly, in Routinator [45], invalid routes are categorized into two reasons: unmatched-AS and unmatched-length.

Our work assesses invalid routes based on their CVRPs (§2.1). In fact, many invalid routes have multiple CVRPs simultaneously, which implies that an invalid route may belong to multiple unmatched types at the same time. Therefore, we categorize CVRPs according to different types of mismatches between CVRPs and routes for distinct analysis processes, rather than categorizing the invalid routes themselves. Specifically, for a given route, we categorize its CVRPs into 4 types based on the criteria of CVRP AS, prefix, and Maxlength. The route (AS65536, 192.0.2.0/26) serves as an illustrative example, guided by [14] and [37]:

- **Unmatched-AS CVRP**: the CVRP AS is different from the route origin AS, while the route prefix length is less than or equal to Maxlength: CVRP (AS65551, 192.0.2.0/24, 26).
- **Unmatched-length CVRP**: the route origin AS is the same as the CVRP AS while the route prefix length exceeds the Maxlength: CVRP (AS65536, 192.0.2.0/24, 24).
- **Unmatched-A&L CVRP**: both CVRP AS and Maxlength are unmatched with the route: CVRP (AS65551, 192.0.2.0/24, 24).
- **Matched CVRP**: both CVRP AS and Maxlength are matched with the route, validating the route as valid: CVRP (AS65536, 192.0.2.0/24, 26).

Accordingly, we design different sets of characteristics tailored to various CVRP types to assess invalid routes. The first three characteristics(§4.2, §4.3, §4.4), are employed in scenarios where the route origin AS is inconsistent with the CVRP AS, addressing cases of unmatched-AS and unmatched-A&L CVRPs. The fourth characteristic, path-overlap(§4.5), is utilized for unmatched-length and condition-specific unmatched-A&L CVRPs. The final characteristic, competitor(§4.6), is essential for comprehensively considering all types of CVRPs. The combination of multiple characteristics can offset the limitation of each characteristic, improving accuracy and effectiveness in the overall methodology.

4.2 Same-organization

The invalid routes where the origin AS belongs to the same organization as its CVRP AS should be considered harmless. An organization can manage multiple ASes, and the address resources could be legitimately allocated within these ASes. When operators announce routes in different ASes for the same address resource but do not issue the corresponding ROAs on time, h-invalid routes are created with unmatched-AS or A&L CVRPs.

For example, on October 1, 2023, AS263210 announced a route for 179.51.117.64/27, which looks like a sub-prefix hijack targeted at the route for 179.51.117.0/24 originated by AS262220. AS262220 is authorized to be the legitimate origin of the prefix 179.51.117.0/24 with Maxlength 29 in the ROA. Due to the inconsistent AS, the route originated by AS263210 is invalid, while AS262220 and AS263210 belong to the same organization `HV TELEVISION S.A.S`, according to the AS organization information provided by CAIDA [19].

The same-organization is also generally agreed upon and widely acknowledged [22, 25, 35]: the invalid routes are more likely human error than malicious hijacks if their origin AS belongs to the same organization as the authorized AS, and CAIDA's AS organization dataset [19] is commonly used as a reference. Although it's common for different entities within the same organization to operate networks independently, the overestimation from same-organization cases is minimal: historical hijacks reported by BGPMon [3] (July 1 to December 30, 2023) that occur within the same organization are very rare (0.7%), even when no filters are applied to reduce BGPMon's false positive rate; and most same-organization h-invalid routes are stable (over 95% last >40 days), suggesting they stem from resource allocation or ROA errors, not misconfigurations.

Specifically, we utilize the attribute "org-id" of CAIDA's AS organization dataset [19] to distinguish whether ASes belong to the same organization. The "org-id" is subject to specific naming conventions and must be registered as a unique entity in the databases of IRR (e.g., ARIN, RIPE NCC, APNIC, etc.). However, only "org-id" is insufficient for the identification of all ASes that belong to the same organization in the real world. This is particularly the case in instances where the organization in question comprises multiple entities, such as subsidiaries of multinational groups in different regions. Therefore, besides this characteristic, we also consider 4 other characteristics to provide a more comprehensive assessment.

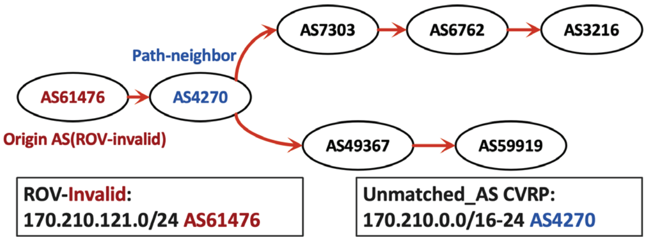


Figure 2: An example of path-neighbor. AS4270 is the path-neighbor of AS61476. The red vector is the propagation path (AS-path) of the invalid route (AS61476, 170.210.121.0/24), and AS4270 matches the ROA (AS4270, 170.210.0.0/16, 24).

4.3 Path-neighbor

The invalid routes where the CVRP AS appears as the next hop of the origin AS in the AS-path field should be considered harmless. Path-neighbor means that the origin AS forwards its BGP announcement to the CVRP AS to propagate to the rest of the Internet, which also implies that the CVRP AS forwards the origin AS's BGP announcement to its neighbors for further forwarding.

The forwarding behavior of path-neighbors is similar to that of ASes with provider-customer (p2c) or peer-peer (p2p) relationships, but requires a stricter condition: the path-neighbor AS is the legitimate owner authorized by ROAs of the prefix in the invalid route, and participated in the propagation of this invalid route.

We advocate employing the path-neighbor as a characteristic, rather than relying on the AS business relationship [24]. Since the AS business relationship is inherently derived through inference, which may be subject to inaccuracies and incompleteness [46]. In contrast, the path-neighbor is a conclusion drawn directly from empirical observations, thereby providing a more factual basis. The concept of path-neighbor differs from the traditional provider-customer (p2c) or peer-to-peer (p2p) relationships, as it is specifically constrained to the particular prefix of the invalid route being assessed. This constraint, rather than being a limitation, actually enhances the reliability of the path-neighbor by focusing specifically on a single route prefix, as opposed to broader AS relationships. Just as Fig. 2 shows, AS4270 (`Red de Interconexion Universitaria`) is the path-neighbor AS of AS61476 (`Universidad Nacional de La Pampa`), while they do not have same-organization or p2c/p2p relationship in CAIDA's dataset [19] issued on October 1, 2023.

To better illustrate the security and effectiveness of path-neighbor, we provide a more detailed rationale for this in the following:

Security. We consider whether using path-neighbor as a characteristic to determine h-invalid routes is secure, specifically regarding whether an invalid route with the next hop being the CVRP AS

Table 2: Comparison of path-neighbor and p2c relationship in the assessment of invalid routes on October 1, 2023.

Method	Total Count	Common Count	Unique Count
Path-neighbor	14498	8896	5602
Provider-Customer	9601	8896	705

is potentially a prefix hijack. Hijacks between upstream and downstream relationships are very rare. In particular, the topology filter proposed by Schlamp et al. [56] uses the upstream-downstream relationships to assess hijacking attacks. They posit that the behavior of the upstream AS propagating a BGP announcement from the downstream AS for the upstream AS's address resources indicates that the downstream AS's BGP announcement is approved by the upstream AS. This, in turn, suggests that BGP announcements from downstream ASes are not instances of hijacking attacks. In terms of the upstream-downstream relationship of propagation without considering AS-path forgery (§8.2), the existence of a physical connection between the origin AS and the CVRP AS also indicates a potential business relationship in reality, including p2c, p2p, sibling-to-sibling (s2s), and so on [24, 28].

To further illustrate the security, we compare upstream-downstream pairs with BGPMon [3]. In daily global routing tables acquired from RIB (RouteViews [63], RIPE RIS [55]) and snapshots (PCH [36]) files generated at 0:00 (UTC), considering only the origin AS and its next hop in alignment with path-neighbor, over 630K upstream-downstream pairs are identified. Comparing these pairs with BGPMon reported historical hijacking records (July 1 to December 30, 2023) shows that only 0.8% (7/813) hijacks happen within upstream-downstream pairs.

Due to the known false positives in BGPMons [56], we conduct a comprehensive analysis of these 7 hijacking records. Specifically, four of the records are deemed invalid because they lack a path-neighbor relationship: the CVRP (victim) AS is not the next hop for the origin (hijacker) AS in the AS-path of the invalid route. Instead, these records indicate upstream and downstream relationships in routes for other prefixes. Among the remaining three records, one is ROA-valid and is considered a false positive; one is unknown, unprotected by RPKI, and is beyond the scope of our study; and the last one is that the origin (*hijacker*) AS and CVRP (*victim*) AS belong to same organization based on CAIDA's AS relationship dataset [20]. Given these findings, it can be reasonably concluded that path-neighbor is a reliable characteristic of h-invalid.

Effectiveness. In the comparative experiment of invalid route assessment, the number of path-neighbor AS pairs is 1.5 times greater than the number of p2c AS pairs, as shown in Tab. 2. Moreover, 92% of invalid routes extracted by the p2c relationship also exhibit the path-neighbor characteristic. The result indicates that path-neighbor not only has a greater capacity for extracting h-invalid routes but also encompasses the majority of p2c extractions, demonstrating its superior efficiency compared to the p2c relationship. The BGP route data used in the comparative experiment is also acquired from RIB (RouteViews [63], RIPE RIS [55]) and snapshots (PCH [36]) files generated at 0:00 (UTC).

Furthermore, we hypothesize that the remaining 8% p2c extraction not encompassed by path-neighbor is due to the discrepancy between current snapshots of BGP routing data and CAIDA's AS business relationship inferences, which may use a much longer data span. The inferences are released once a month. Nevertheless, 174 of the p2c relationships are not encompassed by path-neighbor but would be identified by the same-organization. This illustrates that the combination of multiple characteristics could compensate for each other's limitations, thereby achieving accuracy and effectiveness in the overall methodology.

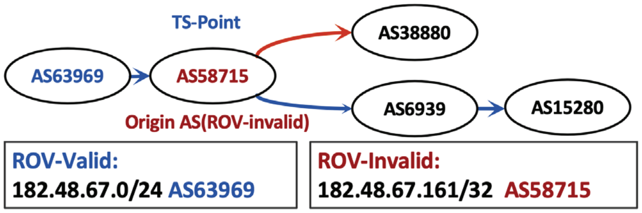


Figure 3: Example of TS-point. AS58715 is a TS-point of AS63969 under the route for 182.48.67.0/24. The red vector is the propagation path (AS-path) of the invalid route, and the blue vector represents a valid route propagation.

4.4 Traffic Shunt Point

Invalid routes resulting from the configuration of routing policies, such as route aggregation and traffic engineering, by other ASes in the propagation path should be considered harmless. In such cases, traffic can be forwarded to a new propagation path distinct from the valid route via a re-originated route. The characteristic used to identify this phenomenon is known as the Traffic Shunt Point (TS-point).

As shown in Fig. 3, the invalid routes originated from AS58715 meet the criteria of a TS-point: (1) the invalid routes carried a competitive prefix with valid routes, including parent-prefix, exact-prefix, and sub-prefix, and (2) AS58715 is involved in the propagation of the corresponding valid route, i.e., it appears on the valid route AS-path. Specifically, the standard of judgment for the TS-point is that the AS is present after the CVRP AS in the AS-path field of the valid route announced by the CVRP AS for a competitive prefix. TS-point may permit all ASes on a valid route path to hijack traffic aimed at the CVRP AS. However, they have no incentive to do so, as even without re-originating the invalid routes, traffic passes through them to reach the CVRP ASes, under the impression of the original valid routes.

TS-point and path-neighbor are proposed to address two distinct scenarios, and they represent separate concepts without intersection. For instance, TS-point deals with route aggregation on the propagation path, while path-neighbor pertains to the scenario where a provider AS assigns address resources to a customer AS without registering a ROA. Intuitively, the relationship between the origin AS of an invalid route and the CVRP AS on the AS-path is opposite in these scenarios: for the TS-point, the origin AS is upstream of the CVRP AS, whereas in the path-neighbor, the origin AS is downstream of the CVRP AS.

Furthermore, we use the historical hijacks reported by BGPMon [3] to validate the reliability of the TS-point. To reduce the impact of false positives in BGPMon [56], we apply the same filters as those used by Themis [49] to eliminate false positive hijacks.

Among the filtered historical hijacks, none are categorized as TS-point, confirming the accuracy of the TS-point.

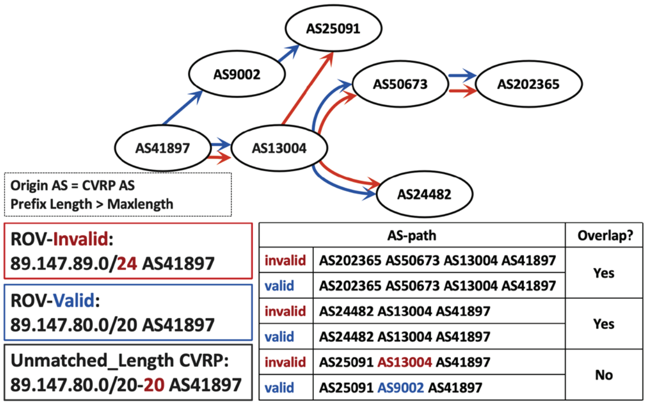
4.5 Path-overlap

Previous studies [22, 35] are concerned with invalid routes that result from too-specific prefixes and tend to classify them as misconfigurations rather than hijacks. The SROV [35] explicitly cites the use of `too long prefixes` (ARA score = 1) as a strong indicator of a benign conflict. However, this perspective overlooks the protective role of Maxlength in preventing forged-origin sub-prefix hijack [27]. Consequently, we believe that a further evaluation of such invalid routes is necessary.

In our study, we frequently notice that the AS-path field in the invalid routes with unmatched-length CVRPs is identical to the AS-path of the valid parent prefix originated by the same AS, which are not multiple origin AS (MOAS) events. **If the propagation paths of invalid sub-prefixes are identical to those of valid parent prefixes, the invalid route is deemed harmless, as it does not indicate intent to hijack traffic**. From this consideration, we propose the path-overlap, specifically for invalid routes that have unmatched-length CVRPs, to describe the invalid sub-prefix propagates along with the identical forwarding paths as the valid parent prefix, observed by the same vantage point (VP). For instance, as shown in Fig. 4, the valid route announced by AS41897 for 89.147.80.0/20 and the invalid route for its sub-prefix 89.147.89.0/24 with the same origin AS are path-overlap.

Hijack defense systems [2, 58, 65] all involve a well-established concept: routes without path changes are not typically considered "real" hijacks. ARTEMIS [58] stores locally the following lists of directed AS-links: previously verified AS-links list, AS-links list from monitors and local BGP routers, to detect the fake link against Type-N (N ≥ 2) hijacks. The N indicates the position of the right-most fake link in the forged announcement, which determines the type. For example, forged AS-path (hijacker AS, AS2, AS1, victim AS) is classified as a Type-3 hijacking, and forged AS-path (hijacker AS, AS1, victim AS) is classified as a Type-2 hijacking. The most commonly observed hijack is Type-0 hijacking, which is also called origin hijacking. And the forged-origin hijacking introduced by Gilad et al. [27] is Type-1 hijacking. In contrast to the comparison with locally pre-stored AS-links, path-overlap compares the AS-paths of the valid parent-prefix and the invalid sub-prefix from the same origin AS at the same moment, which is more reliable and time-sensitive. We posit that an invalid sub-prefix announced by the CVRP AS will typically share the same propagation path as the valid parent-prefix, assuming no specific routing policies are in effect. Furthermore, the identical propagation path precludes the possibility of forged-origin hijacks (as in [27]) and path manipulation, in the context that we default the valid route is legitimate.

4.5.1 ***Directionally Identical.*** Additionally, we explicitly regard "Directionally Identical" as identical paths, too. "Directionally Identical" refers to the propagation of invalid/valid routes aimed at the same destination (the same VP), and the AS-path of the valid route fully covers the AS-path of the invalid route. In essence, concerning a given destination AS, the sequence of ASes comprising a valid parent route AS-path must exclusively include those present in the invalid prefix route of the same origin AS. This implies that the set of ASes in an invalid route AS-path is a subset of the set of ASes in a valid parent route AS-path. For instance, guided by documentation ASNs (RFC5398 [37]), if the AS-path of the invalid route is (AS64496 AS64497 AS64498 AS64500), and the AS-path of the valid route is (AS64496 AS64497 AS64498 AS64499 AS64500), we can consider them to have a directional identity. Conversely, if the AS-path of the invalid route is (AS64496 AS64497 AS64511 AS64500), it is not “Directionally Identical”.



Origin AS = CVRP AS
Prefix Length > Maxlength

ROV-Invalid:
89.147.89.0/24 AS41897

ROV-Valid:
89.147.80.0/20 AS41897

Unmatched_Length CVRP:
89.147.80.0/20-20 AS41897

AS-path		Overlap?
invalid	AS202365 AS50673 AS13004 AS41897	Yes
valid	AS202365 AS50673 AS13004 AS41897	
invalid	AS24482 AS13004 AS41897	Yes
valid	AS24482 AS13004 AS41897	
invalid	AS25091 AS13004 AS41897	No
valid	AS25091 AS9002 AS41897	

Figure 4: An example of path-overlap (partial path-overlap: two of three paths overlap). Red arrows represent the propagation (AS-path) of invalid routes, and blue arrows represent the propagation of valid routes. AS41897 announces one valid parent prefix (89.147.80.0/20) which matches the ROA (89.147.80.0/20, AS41897, 20), and one invalid sub-prefix (89.147.89.0/24). Two AS-paths of the invalid route are identical to that of the valid parent prefix, and one is different.

4.5.2 ***Full and partial path-overlap.*** Following this definition of identical paths, full path-overlap is defined as instances where all AS-paths of the invalid route are identical to those of the valid parent prefix. Partial path-overlap, on the other hand, encompasses situations where both identical paths and different paths exist. As illustrated in Fig. 4, for an invalid route with n AS-paths, if m (0<m<n) AS-paths can be overlapped, it is a partial path-overlap.

Both full path-overlap and partial path-overlap are h-invalid. In the case of full path-overlap, the invalid sub-prefix propagates exactly along the propagation path of the valid parent prefix, so it is not a malicious hijack. For partial path-overlap routes, some traffic is forwarded along the same path as the valid parent prefix, while other traffic is forwarded along different paths, which is similar to load balancing. Of the more than 4000 partially path-overlap routes identified daily from July 2023 to July 2024, 80% to 95% of these have next hops of the origin AS that also appear in the AS-paths of other valid routes originated by its same AS. This further substantiates the assertion that the different paths are deliberate choices made by the operators to engineer traffic between different neighbors. To circumvent the consequences of forged-origin hijacking, the remaining 5% to 20% of the different paths are disregarded, as the next hop of their origin only appears on the invalid route's AS-path. In addition, among the historical hijacks reported by BGPMon [3], there are no instances (0/813) of path-overlap, including both full path-overlap and partial path-overlap. This demonstrates the effectiveness of path-overlap.

For unmatched-A&L CVRPs meeting same-organization and path-neighbor criteria, we transform them into unmatched-length CVRPs. Specifically, we assess whether they meet the path-overlap criteria by replacing the origin AS as the CVRP AS in the AS-path under the transformed unmatched-length conditions.

4.6 Competitor

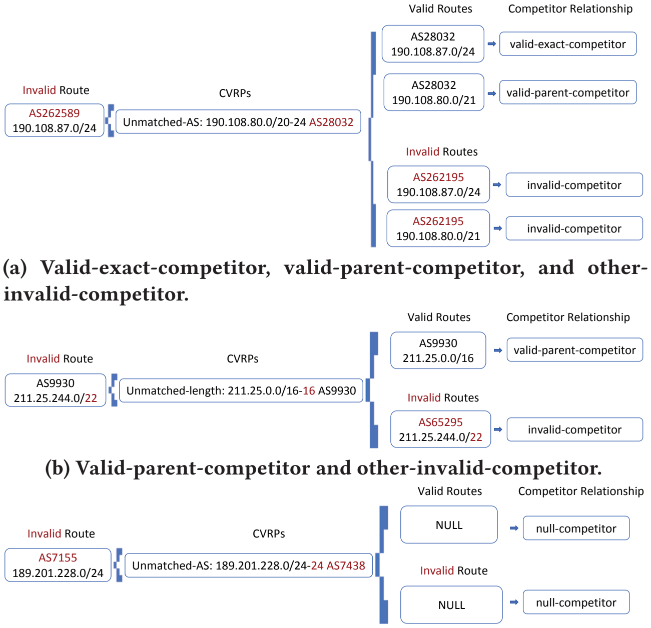
The characteristic competitor focuses on the traffic competition within routes to exclude the hijacks. We use the competitor to exclude the possibility of malicious hijacks, as traffic competition is the most likely scenario for such hijacks. It is evident that routes without competitors cannot constitute malicious hijacks, as there are no potential victims involved.

Distinct from De Facto Ownership. DISCO [30] is a new mode that aims to avoid the security threats posed by the authoritative trust anchors of RPKI. Instead of relying on RPKI for authoritative authorization, DISCO determines the **de facto ownership** of address resources based on the duration. However, DISCO will put the legitimate routes that appear briefly at a disadvantage. Another key limitation of DISCO arises when multiple ASes originate the same prefix (routes with multiple competitors). In such cases, it becomes challenging to determine a single de facto owner.

In the context of the competitor, we address not only the no-competitor scenario (single de facto ownership) but also the multi-competitor scenario. To handle the latter, we utilize two characteristics introduced before: same-organization and path-neighbor. Furthermore, since DISCO considers non-overlapping IP address blocks rather than prefixes as the objects in its certificates, it can identify the address space of a sub-prefix as a de facto owner, while assigning another de facto owner to the address space outside the sub-prefix for parent-prefix. This approach creates potential opportunities for sub-prefix hijacking. Therefore, in our study, we consider prefixes as objects of competitive categorization, where sub-prefixes compete for the traffic of the parent-prefix, i.e., for de facto ownership of the parent prefix's address space.

Criteria for Detecting Competitor. Specifically, we detect competitors through CVRPs. For one invalid route, any routes that share the same CVRP are considered its competitors. They may compete for the same address resources because their prefixes are covered by the same CVRP prefix. We further compare their prefixes to determine whether competition indeed exists and the type of competitor. Consequently, we classify competitors into four categories, including valid-exact-competitor, valid-parent-competitor, other-invalid-competitor, and null-competitor:

- Valid-exact-competitor: The CVRP AS announces a valid route for the same prefix as the invalid route, as shown in Fig. 5a. Valid-exact-competitors can be classified into different types based on their CVRP sources, and each type will be treated differently.
- Valid-parent-competitor: the CVRP AS announces a valid route with the parent prefix of this invalid route, as shown in Fig. 5a and 5b. Like valid-exact-competitor, valid-parent-competitor could also be classified into different types.
- Other-invalid-competitor: other invalid routes also compete for the address resource, as shown in Fig. 5a and 5b.
- Null-competitor: the origin AS is the sole announcer for the address resource, as shown in Fig. 5c.



(a) Valid-exact-competitor, valid-parent-competitor, and other-invalid-competitor.

(b) Valid-parent-competitor and other-invalid-competitor.

(c) Null-competitor.

Figure 5: Representative examples of competitors. NULL in the box indicates the absence of such competitors. Only invalid routes for which both the valid competitor and the invalid competitor are NULL have null-competitor.

We ignore the valid-child-competitor since the longest-prefix-match routing ensures that the sub-prefix is always preferred. As for valid-child-competitor, the invalid route can not affect the valid routes they compete with.

Benign competition. There is a clear distinction between benign competition and aggressive competition for an invalid route's competitors. An invalid route is considered with benign competition if it only includes the following types of competitors: (1) null-competitor, (2) valid-parent-competitor from unmatched-length (both common and transformed), (3) other-invalid-competitor that all ASes are same-organization with origin AS. The null-competitor means no victim. Regarding the latter two cases of benign competition, we demonstrate the equivalence of these ASes by using the same-organization and path-neighbor.

In the above three corresponding scenarios, we classify the traffic competition between the routes as benign, meaning that invalid routes do not encroach upon the address resources of valid routes. However, filtering out the majority of these routes can lead to the disappearance of prefixes from the Internet, particularly when they have only a null-competitor. An invalid route that has competitors not in the above three types should be considered with aggressive competition. **In light of the potential impact on routing security, it is advisable to categorize invalid routes facing benign competition as h-invalid**.

It is important to note that benign competition is distinct from legitimate ownership. Therefore, we do not directly designate h-invalid routes as legitimately invalid. Moreover, we utilized the characteristics mentioned earlier in our competition-type judgment, but this was not a simple repetition. By combining characteristics, we compensate for the limitations of a single characteristic, thereby enhancing the accuracy of our overall methodology. For instance, an invalid route that meets the criteria of path-neighbor or same-organization for an unmatched-A&L CVRP, i.e., has a transformed unmatched-length CVRP, is not considered h-invalid if it does not also satisfy the path-overlap criterion after replacing the origin AS in the AS-path. However, if the route only has valid-parent-competitors from the transformed unmatched-length CVRP, it is recognized as h-invalid because it fulfills the conditions for benign competition. Thus, the combination is justified.

5 Datasets and Methodology

To assess the legitimacy of invalid routes, we design a characteristics-based methodology that distinguishes the h-invalid and s-invalid routes, based on the above five characteristics. We introduce the input datasets (§5.1) and the methodology workflow (§5.2).

5.1 Datasets

The datasets for our methodology consist of three parts: (1) **The global routing tables**, (2) **ROAs**, and (3) **AS organization datasets**.

The global routing tables. The global routing tables, retrieved from PCH [36], RIPE RIS [55], and RouteViews [63], including both IPv6 and IPv4 prefixes, are collected from more than 200 collectors worldwide. We select the RIB (RIPE RIS [55] and RouteViews [63]) and snapshots (PCH [36]) files generated at 0:00(UTC) to represent the daily routing table. The data from PCH is in a different format compared to that from Routeviews and RIPE RIS. This disparity poses a challenge when attempting to integrate them. To obtain sufficient information on BGP announcements and the global state of the inter-domain routing system, we design an auto-processing program to standardize data formats from three public platforms and construct global routing tables. The total number of route pairs acquired per day exceeds 1.4 million.

ROAs. We use Routinator [45] as RP software to retrieve ROAs from the RPKI repository at the same time as BGP data retrieved in the default configuration. The total number of ROAs acquired at one time exceeds 460K and continues to increase over time.

AS organization datasets. The AS organization datasets [19] are retrieved from CAIDA.

5.2 Methodology Workflow

The workflow of our methodology can be described as three parts: (1) **Data Preprocessing**, (2) **Characteristics Analysis**, and (3) **Classification Scheme**.

5.2.1 ***Data Preprocessing.*** In this procedure, we clean up the collected BGP data obtained from public probe services (§5.1), discarding routes involving reserved or private ASes and prefixes. Likewise, routes that include a loop in the AS-path are also eliminated. The corresponding time ROA is also an essential input, as we described in §5.1. Firstly, we need to extract route pairs to denote each route and store the corresponding AS-path set for subsequent characteristics analysis. ROV is the second step. We validate route pairs to determine the ROV status and categorize the CVRPs of each route pair into unmatched-AS, unmatched-length, and unmatched-A&L, respectively, for different characteristics analysis processes. After data preprocessing, BGP data and ROAs become the collection of route entities (§2.1).

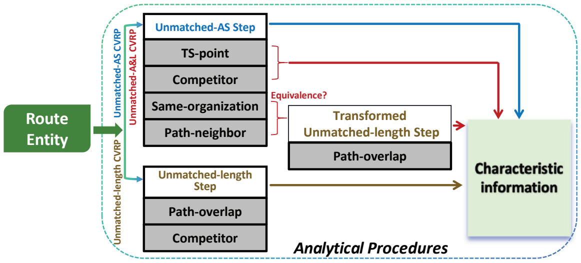


Figure 6: Overview of Characteristics Analytical Procedures.

5.2.2 ***Characteristics Analysis.*** Each route entity has a set of AS-paths and a set of CVRPs, and we assess the invalid part. We outline the analytical procedures for each type of CVRP in Fig. 6. Specifically, for unmatched-AS CVRPs, we analyze the same-organization, path-neighbor, TS-point, and competitor as the unmatched-AS step; for unmatched-length CVRPs, we analyze path-overlap and competitor as the unmatched-length step. For the unmatched-A&L CVRP, it is necessary to perform an unmatched-AS step first due to the mismatch of both the origin AS and prefix length. If the equivalence between the CVRP AS and the origin AS is established, the CVRP is transformed into a transformed unmatched-length CVRP. Subsequently, the transformed unmatched-length step is executed, with the origin AS replaced by the CVRP AS in the AS-path. Equivalence here means they satisfy the same-organization or path-neighbor. After this procedure, we obtain the characteristic information of each route entity, and the info is used to classify the invalid routes into different categories in the classification scheme.

5.2.3 ***Classification Scheme.*** The information generated by the characteristics analysis is employed to extract h-invalid routes. We integrate the reliability of individual characteristics and their combinations to develop a more nuanced classification scheme.

First, we categorize each invalid route's competition as either benign or aggressive. Next, for invalid routes with unmatched-length CVRPs, we identify those that satisfy the path-overlap criterion. Given concerns about CVRP ASes that continue to announce address blocks after leasing or allocating sub-prefixes to other ASes, we assign higher confidence (strong evidence) to invalid routes exhibiting both path-overlap and benign competition (C1), while routes with path-overlap and aggressive competition (C2) are assigned lower confidence (weak evidence). For invalid routes with unmatched-AS CVRPs, we directly assign high confidence (strong evidence) to those with same-organization (C3) and path-neighbor (C4). For invalid routes with unmatched-A&L CVRPs, we term them "double fault", considering the use of two-step analysis and path-overlap, and classify them into two categories based on the type of competition (C5 and C6), similar to the case of path-overlap alone. For invalid routes deemed h-invalid exclusively due to benign competition, we categorize them as "benign competition only" (C7) and assign them lower credibility (weak evidence). Invalid routes identified by the TS-point are categorized as C8 with strong evidence. Invalid routes that do not fit into the above eight categories are not considered h-invalid and are recorded as C9 (s-invalid). Based on this scheme, the following classification rules are formed:

- C1 (strong): Path-overlap and benign competition. Invalid routes with unmatched-length CVRPs that fulfill the requirements of path-overlap and engage in benign competition.

- C2 (weak): Path-overlap and aggressive competition. Invalid routes with unmatched-length CVRPs that fulfill the requirements of path-overlap but face aggressive competition.
- C3 (strong): Same-organization. Invalid routes with unmatched-AS CVRPs, where the origin AS belongs to the same organization as at least one CVRP AS.
- C4 (strong): Path-neighbor. Invalid routes with unmatched-AS CVRPs, where the origin AS is a path-neighbor to at least one CVRP AS.
- C5 (strong): Double fault h-invalid with benign competition. Invalid routes with unmatched-A&L CVRPs that satisfy the path-overlap criterion in the transformed unmatched-length step. Based on the type of competition, double fault h-invalid routes are classified into two categories: C5 and C6.
- C6 (weak): Double fault h-invalid with aggressive competition.
- C7 (weak): Benign competition only. This type of invalid route is deemed h-invalid exclusively due to benign competition.
- C8 (strong): TS-point. Invalid routes with unmatched-AS or unmatched-A&L CVRPs, where the origin AS is the TS-point of a CVRP AS.
- C9 (no): S-invalid. Invalid routes not fitting into any of the above categories are deemed suspicious.

Weak and strong evidence. In the aforementioned classification scheme, we identify additional concerns regarding specific h-invalid routes and assign them weak evidence. Conversely, other h-invalid routes are assigned strong evidence to indicate their higher confidence level. By specifying these two confidence levels, we aim to offer the community more realistic guidance for classification rules. This guidance can assist operators in selecting appropriate filtering rules if they intend to apply our approach in a production environment. We will elaborate on the additional concerns for assigning weak evidence in the Discussion (§8).

Multi-category classification. As we described in §4.1, each invalid route could have at most three types of CVRPs simultaneously, and the classification scheme can produce results based on any type of CVRPs. Therefore, an invalid route may belong to more than one category. If any classification result supports the assessment of h-invalid routes, it is considered h-invalid, no matter which type of h-invalid it is. After the above three procedures, all the invalid routes are classified into three categories: h-invalid routes (strong and weak evidence) and s-invalid.

In general, our methodology integrates dynamic routing information, e.g., traffic competition and AS-path, with static AS information, to propose five characteristics. Some of them are novel and customized insights, such as TS-point, path-overlap, and competitors. Others represent advancements in the state-of-the-art: from p2c to path-neighbor. We combine them to propose eight rules and identify a multitude of h-invalid routes. We anticipate that these findings will garner attention from the community and contribute to the robust advancement of RPKI.

6 Measurement on the Global Routing Tables

In this section, we introduce the measurement and longitudinal analysis of h-invalid routes based on the global routing tables from July 2023 to November 2024.

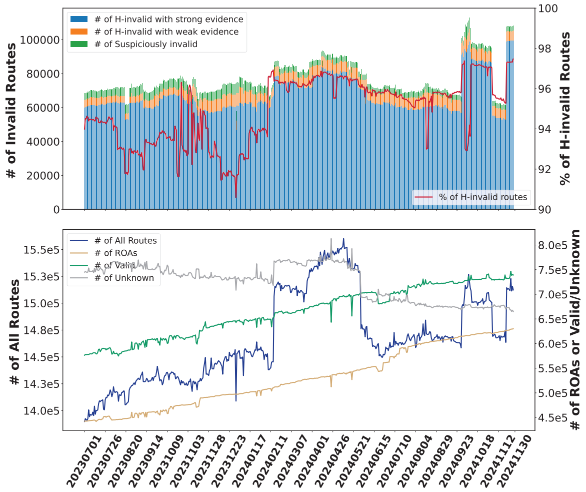


Figure 7: Measurement results of h-invalid routes. The top figure shows the number of invalid routes and the proportion of h-invalid routes. The bottom figure shows the number of BGP routes, ROAs, valid routes, and unknown routes.

First, we present the detailed measurement results, showing that over 91% of invalid routes are h-invalid (§6.1). Subsequently, we validate the accuracy of our measurements through three distinct approaches (§6.2). Following that, we analyze the reasons behind the emergence of h-invalid routes (§6.3). Finally, we give some recommendations based on our measurement results (§6.4).

6.1 Measurement Results of H-invalid Routes

To understand the quantity and trends of h-invalid routes in the global routing table, we conducted measurements of h-invalid routes from July 2023 to November 2024, providing a long-term perspective and comparing them with the changes in the overall routes, ROAs, valid and unknown routes, as shown in Fig. 7. Our measurement reveals a massive proportion (over 91%) of h-invalid routes existing in invalid routes and typically increases when the total number of new routes surges, indicating network operators are prone to negligence when announcing new routes. We notice two **key takeaways** as following:

1) Massive false positives of ROV: Fig. 7 (top) plots the percentages and the number of h-invalid routes and shows that over 91% of invalid routes are h-invalid, up to a maximum of 97.47% on November 30, 2024. Rejecting or deprioritizing will damage their announcers' interests, which hinders the deployment of ROV, and further reduces the desire and incentive to issue ROAs for their address resources [48]. It is like being caught in a vicious cycle that needs to be broken: improving the accuracy on the side of ROV by adding an extra function like SROV [35], or improving coverage and accuracy on the side of ROA issuance.

2) Gap between routes and ROAs: Fig. 7 (bottom) plots the number trends of routes, ROAs, unknown, and valid routes. It is observed that after February 15, 2024, as the number of entire routes rises considerably, the proportion and number of invalid routes also increase significantly. Meanwhile, there was a slight increase in unknown routes, while the number of valid routes continued to grow at a slow pace, similar to the growth trend of ROAs.

The significant surge is primarily attributed to a specific VP AS51185 (MainStreaming SpA), observed over 65K new routes on February 15. Specifically, AS3269 (TIM S.p.A.) announced over 50K unknown routes (divided its /16 prefix into multiple /20-24 sub-prefixes), while AS12357 (VODAFONE ESPANA S.A.U.) and AS30722 (Vodafone Italia S.p.A) contributed most of the invalid routes. The decline after May 29 followed VP (AS51185) ceasing to report AS3269's unknown routes. The invalid cases exhibit a similar situation. The event was only captured by the collector RIPE-RIS-RRC25 [54]. Similarly, after September 29, AS17557 (Pakistan Telecommunication Company Limited) announced over 33.8K invalid routes, AS52863 (UPX TECNOLOGIA LTDA) announced over 8K invalid and 6K unknown routes, and AS22773 (Cox Communications Inc.) announced over 6K invalid. These ASes divided their valid prefix into multiple too-specific sub-prefixes, resulting in a notable increase in invalid routes, the majority of which are classified as h-invalid. A similar event occurred on November 22, 2024, when AS11845 announced over 45K invalid routes, the majority of which were /32 sub-prefixes derived from its /23-24 valid prefix. We also observed the notable event where the number of routes dropped sharply on December 31, 2023, and recovered the next day. Analysis revealed that the global routing table on this day was missing 12K unknown routes from AS12969 (Vodafone Iceland) and a total of approximately 13K invalid routes from AS16010 (Magticom Ltd.), AS15964 (CAMTEL), AS210664 (Virtua-Networks SARL), and AS15435 (DELTA Fiber Nederland B.V.), compared to the two days before and after. This led to a notable decrease in the corresponding numbers in Fig. 7. We speculate that this was due to occasional data loss from the publicly available services, which may also account for other single-day outliers.

The notable increases of routes in Fig. 7 can be clearly attributed to the routing policies implemented by network operators, which involve dividing otherwise valid prefixes into numerous too-specific sub-prefix announcements. This practice generates a large number of h-invalid routes.

Although Fig. 7 indicates a continuous increase of ROAs, a gap still exists between the ROA increase and network operators' dynamic announcement behavior, leading to a notable number of false positives in ROV. On one hand, the process of ROA issuance, which involves multiple roles from CAs to routers and the data plane, is not instantaneous. On the other hand, network operators are also prone to negligence when configuring new announcements.

Significance of h-invalid routes problem. Despite the growing number of ROAs and the expanding coverage of ASes and address space over time, the number of h-invalid routes remains high and shows no clear downward trend, as shown in Fig. 7. This indicates that the problem persists and has not been significantly mitigated, highlighting the ongoing severity, which remains a critical concern for RPKI deployment. Additionally, occasional surges in h-invalid routes reveal that changes in routing policies can easily lead to h-invalid routes, underscoring the urgency of addressing this problem. The large number of h-invalid routes burdens route hijacking detection systems. If operators need to conduct further legitimacy validation on invalid routes.

On the other hand, our comparison of the propagation scope of invalid, valid, and unknown routes reveals that, even with the current low ROV adoption rate, the number of propagation nodes for invalid routes (i.e., the size of the AS set across all AS-paths for the same route) is significantly lower than that for valid and unknown routes. The median number of propagation nodes for invalid routes is approximately one percent of that for valid and unknown routes. This finding aligns with the observation by Hlavacek et al. [34] that ROV-enforcing routers have fragmented the Internet into several regions, with invalid routes having a very limited scope of influence. As ROV adoption increases, the viability of h-invalid routes is expected to diminish further.

In summary, addressing the issue of h-invalid routes is crucial, whether from the perspective of promoting RPKI and ROV deployment to enhance the security of the Internet's inter-domain routing system or from the standpoint of meeting network operators' routing needs. H-invalid routes represent a pressing problem that requires immediate attention.

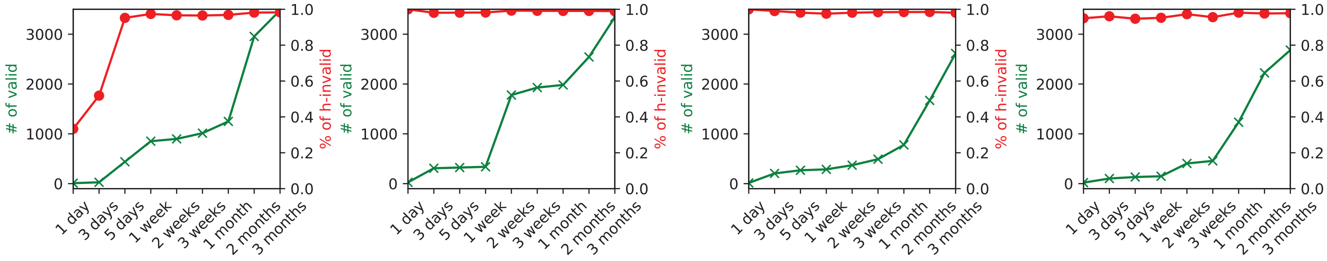
6.2 Validation

Despite the continued interest and previous discussions[22, 25, 35] regarding the ROV errors, no comprehensive database currently exists. Consequently, the validation remains an unresolved issue. To address this, we employ multiple approaches to validate our findings, focusing on whether harmless routes can be correctly identified as h-invalid, and whether hijacking routes are not mistakenly identified as h-invalid.

The routes that were invalid at the time but later became valid (valid-in-the-future routes) indicate that these routes were deemed invalid due to register delays rather than hijack attempts. Furthermore, we manually searched online sources to confirm that their reported hijacks do not align with h-invalid in our measurements. Therefore, we build two ground truth datasets, including an across-timeline validation dataset and a hijacks dataset, to focus on benign and malicious cases, respectively. Additionally, we compared our results with the Aggregated Resolution Algorithms (ARA) in SROV[35], which aimed at differentiating ROV errors from hijacks and was already introduced in §3.

6.2.1 ***Across-timeline Validation.*** Instead of malicious hijack attempts, a multitude of routes are validated as invalid due to the absence or inaccuracy of ROA. This idea is also confirmed by real-world data in our study. We find that some routes currently validated as invalid will become valid after a few days. Intuitively, it seems that the operators notice the ROA issue over time and subsequently implement a solution. Therefore, we construct the across-timeline validation dataset consisting of invalid routes that will be valid in the future as ground truth.

As an efficient and accurate assessment methodology, routes that will become valid in the future should be accurately identified as h-invalid. To prove this assertion, we set a series of monitoring windows to observe the switching of the ROV validation status of routes and build a dataset. Specifically, we set 9 monitoring windows of varying sizes, from 1 day to 3 months, based on specific reference points. We monitor routes that were validated as invalid in the routing table on those specific reference point dates and subsequently re-validated as valid-in-the-future, following the duration of each monitoring window. For instance, we designate 4 reference points: October 1, November 1, December 1, 2023, and January 1, 2024, and summarize the validation results in Fig. 8.



(a) No.1: October 1, 2023 **(b) No.2: November 1, 2023** **(c) No.3: December 1, 2023** **(d) No.4: January 1, 2024**

Figure 8: Across-timeline validation. Our methodology identifies over 96% of routes that switch from invalid to valid as h-invalid. Green lines present the number of invalid routes switching to valid in the monitoring window. Red lines present the h-invalid proportion of these routes.

Our methodology effectively identifies the overwhelming majority (exceeding 96%) of routes that switch from invalid to valid as h-invalid on the four reference points. As shown in Fig. 8, we summarize the results for all monitoring windows at each observation point and find that almost all valid-in-the-future routes can be recognized as h-invalid, highlighting the effectiveness and accuracy of our methodology in assessing the legitimacy of invalid routes. The proportion of routes identified as h-invalid in short monitoring windows is slightly lower in Fig. 8a, which is due to the fact that the number of valid-in-the-future routes is very low, below 30, making them susceptible to occasional fluctuations.

To provide additional context, we designate four additional reference points. The outcomes of across-timeline validation II, utilizing February 1, March 1, April 1, and May 1, 2024, as reference points, are illustrated in Fig. 9. Our methodology identifies over 96% of routes that switch to valid as h-invalid.

Furthermore, we set the monitoring window to 30 days and analyzed all valid-in-the-future routes throughout the measurement. We found that our method is capable of classifying 95.79% of such routes as h-invalid routes. This high classification rate underscores the effectiveness of our methodology in accurately identifying routes that become invalid due to reasons other than hijacking attempts. Overall, the switching from invalid to valid over time indicates a delay that exists between BGP announcements and ROAs, which is caused by human errors and should be accountable in such cases.

6.2.2 ***Historical Hijacks.*** We utilized BGPMon's data as a reference for characteristics derivation (§4). To ensure the independence of our validation process, we selected an alternative source of hijacked data to construct a historical hijack ground truth: Qrator.Radar, developed by Qrator Labs [4], is a pioneering tool of network security and routing analytics, providing incidents such as BGP route leaks and BGP hijacks.

Since Qrator.Radar does not publish all hijacking-related prefixes and AS-path information, we manually collected all BGP announcements of the hijack ASes from update files (PCH [36], RIPE RIS [55], and RouteViews [63]) at the corresponding time. This allowed us to extract the complete set of hijacking events. We then filtered out the ROV-valid or IRR-matched routes (false positives in Qrator.Radar) and ROV-unknown routes (out of the h-invalid scope), using the remaining data as our hijack ground truth dataset. Eventually, we obtained over 55K hijack routes from Qrator.Radar[5–10] and incorporated them into our measurement to validate our methodology in assessing invalid routes. Among these potential hijack incidents, our methodology identifies none as h-invalid and classifies 100% as s-invalid. This result indicates that our methodology does not introduce hijacks, thereby validating the accuracy of our measurements.

6.2.3 ***Comparison to Aggregated Resolution Algorithms (ARA).***

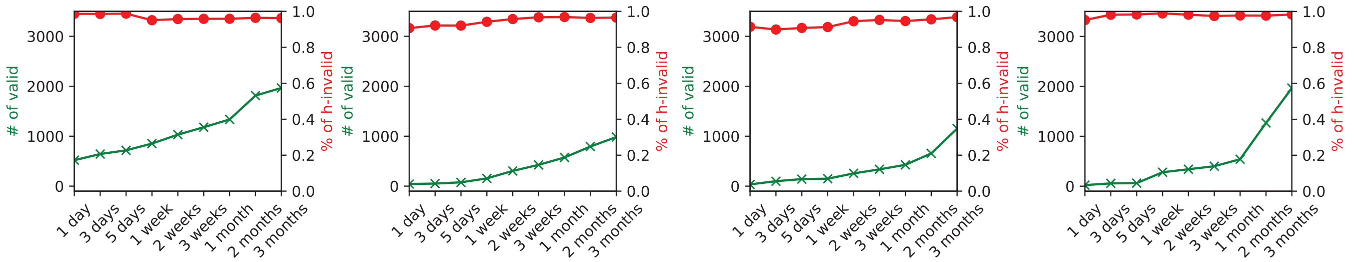
Comparison with existing research enables the verification of our methodology's accuracy. Given the limited availability of suitable methods, we select the ARA [35] for this purpose. As described in §3, the ARA calculates the score of invalid routes based on scoring items. Invalid routes with scores exceeding the threshold are classified as benign conflicts, while those below the threshold are considered malicious hijacks.

We match IRR *route object* with route prefixes using exact matching, calculate the ARA scores for invalid routes in the daily global routing table based on assigned ARA scores provided by SROV [35], and apply a threshold of 0.4, consistent with SROV, to differentiate between benign conflicts (ARA-benign) and malicious conflicts (ARA-malicious). The exact match principle is the common IRR matching rule [50, 51].

Comparative Results. Upon comparing the results obtained from the ARA with those from our methodology, we observe an approximately 63.5% overlap between ARA-benign and h-invalid routes throughout our measurements, i.e., 63% invalid routes are considered not malicious hijacks in both methodologies. While differences in results do exist due to the inherent differences between our methodology and that of ARA, the overall consistency is notable. For approximately 1.2% of the invalid routes, the results are inconsistent: specifically, 0.5% are classified as h-invalid and ARA-malicious, while about 0.7% are identified as s-invalid and ARA-benign. Additionally, ARA is unable to determine around 35.3% of invalid routes (ARA-no-data), since no matching IRR *route object* for the invalid route's origin AS or prefix.

As a supplement, under the longest prefix matching rule, which significantly reduces the number of ARA-no-data, achieving a consistency rate of approximately 92%, an inconsistency rate of about 7%, and an ARA-no-data rate around 1%. Both matching principles show high consistency, except when there is no matching IRR route object, highlighting the accuracy of our method.

We conducted a further analysis of the inconsistencies. For invalid routes categorized as h-invalid and ARA-malicious, the vast majority fall into the C1, C3, and C4 h-invalid routes, which are assigned with strong evidence, yet their ARA scores are 0. This discrepancy may arise from the outdated information present in the IRR; Conversely, for invalid routes categorized as s-invalid and



(a) No.5: February 1, 2024 **(b) No.6: March 1, 2024** **(c) No.7: April 1, 2024** **(d) No.8: May 1, 2024**

Figure 9: Across-timeline validation II.

benign, we find them fall under the subsidiary issues and other limitations of our methodology discussed in §7.

Advantages. Under the exact match principle, our methodology and ARA demonstrate high consistency in the comparative experiments, with only 1.2% of the results showing discrepancies, which underscores the accuracy of our methodology. Compared to ARA, the most prominent advantage of our method is the reduced reliance on third-party databases.

ARA heavily depends on third-party IRR data, and without IRR *route object*, 35.3% of the routes cannot be determined. In contrast, our method primarily leverages comprehensive routing state information, such as AS-path and traffic competition, thereby minimizing dependence on third-party data sources.

As shown in Fig. 10, the influence of the same-organization, which relies on the CAIDA AS Organization dataset [19], is limited in the overall measurement results. On the other hand, the expiration and absence of IRR *route object* [38] can affect the results of the ARA method, while the CAIDA AS Organization dataset remains relatively stable. Only 5.24% of ASes changed "org-id" between the datasets released in July 2023 and November 2024.

6.3 Analysis of H-invalid Routes

To gain a deeper understanding of how the assessment is executed and the main reasons for h-invalid routes, we analyze the number of routes categorized as C1-C8 throughout the measurement period to infer the primary reasons for h-invalid routes and assess the contribution of each characteristic. To better interpret the measurements, we present a detailed analysis of CVRPs and characteristics on the global routing table of Oct. 1, 2023, as an example.

The cumulative number of h-invalid routes categorized as C1-C8 is illustrated in Fig. 10. The distribution of CVRP types for the invalid routes on October 1, 2023, is presented in Fig. 11 and results of the h-invalid routes are detailed in §A. Combining the long-term and specific measurement results, we conclude that the main reasons for h-invalid routes on the Internet are the following:

1) Mismatch between Maxlength and BGP configuration. The number of invalid routes with unmatched-length CVRPs is the highest (Fig. 11), highlighting the severity of prefix length mismatches. Also, the path overlap (C1 and C2) is observed to be particularly extensive in scale (Fig. 10). Path-overlap denotes that the legitimate owner AS announces too-specific sub-prefixes that exceed the Maxlength, implying that one of the main reasons for h-invalid routes is the mismatch between Maxlength and BGP configuration. We notice that operators pay more attention to this problem than invalid routes with wrong AS: Based on our across-timeline validation introduced in §6.2.1, over two-thirds of h-invalid routes switching to valid in the future are C1 or C2.

2) Address transfer within ASes with business relationships. Double-fault (C5, C6), same-organization (C3) and path-neighbor (C4) are also important reasons for identifying h-invalid routes (Fig. 10). Double-fault (C5, C6) means both wrong AS and too-specific prefixes. In this situation, the wrong AS means address transfer, and the too-specific prefix also proves the allocation of sub-prefixes. Although same-organization (C3) and path-neighbor (C4) do not account for a large proportion among h-invalid routes, since unmatched-AS CVRPs constitute just 9.74% of the overall CVRPs (Fig. 11), these two characteristics have already taken up a significant share. They reveal address transfer in the legitimate range of address resources between ASes with business relationships.

3) AS-path re-origination. TS-point (C8) is another important reason for h-invalid routes (Fig. 10). It is defined to detect re-originated routes for traffic engineering or route aggregation.

As for the competitor, routes belonging to the benign competitor only (C7) are categorized as h-invalid with weak evidence to reduce the threat of network disconnection as a conservative choice. Besides, the benign competition improves the evidence confidence of h-invalid in the situation of path-overlap and double-fault.

6.4 Recommendations

A collaborative effort between address resource owners and network operators, who are often distinct entities, is necessary to resolve the issue of h-invalid routes and avoid disconnecting networks from legitimate destinations. We are focusing on 2 key areas:

1) Matching Maxlength and sub-prefix configuration. As shown in Fig. 11, over 53% of invalid routes have unmatched-length CVRPs, and nearly all are h-invalid. This indicates that too-specific sub-prefixes in BGP configurations are a significant and pervasive issue that must be addressed seriously to avoid h-invalid results.

2) Timely ROA registration for legitimate announcers and less aggregation/re-origination. It is recommended that the address resource owner issue ROAs for legitimate announcers that have been assigned addresses and actually originated in BGP [26] timely manner, with particular attention to networks belonging to the same organization and customer AS. While this approach may lead to an increase in the number of ROAs, it represents an effective method for reducing ROV false positives at the source, as opposed to introducing additional judgments that could interfere with the ultimately deterministic nature of RPKI/ROV. Additionally, it is

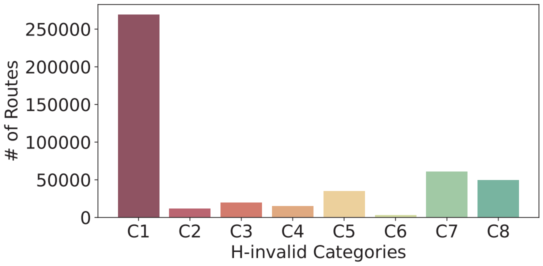


Figure 10: The cumulative distribution of h-invalid routes over the complete measurements from July 2023 to November 2024.

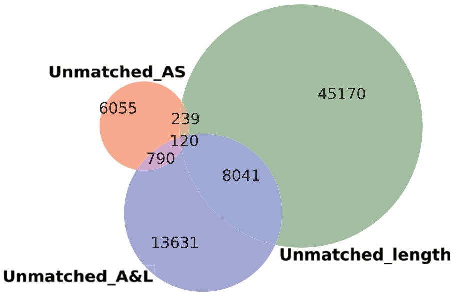


Figure 11: The distribution of CVRPs on October 1, 2023.

advised that some core network nodes avoid aggregation and reorigination of routes. These networks are not the address resource owners, and their announcement behavior can lead to false positives, extracted by TS-point, and make resolution under the RPKI mechanism challenging.

7 Suspiciously Invalid (C9)

Through analyzing the s-invalid routes, we consider that the failure to classify invalid routes as h-invalid can be attributed to three potential scenarios. These scenarios also represent the limitations of our methodology, which may influence the effectiveness:

Address transfer between ASes with undefined relationships. In across-timeline validation (§6.2.1), less than 5% of routes switch to valid are s-invalid. Most of their CVRP ASes are cloud service providers or IP leasing providers. In contrast to AS relationships, such as the same-organization or path-neighbor, which explicit rules can define, the relationships between ASes of these kinds of address service providers and commercial customers are challenging to describe and analyze automatically. Consequently, address transfers occurring between them are hard to identify through the application of explicit rules. For example, AS59432 is managed by an ISP of Spain, named '`GINERNET S.L`', and originated a route for 185.214.100.0/24, which was invalid on October 1, 2023. However, the route is valid before September 30 and after October 3, 2023. During the period of 'invalid', from 2023-09-30 20:00 UTC to 2023-10-02 16:00 UTC, the ROA (AS834, 185.214.100.0/24, 24) produces the 'invalid' status. It does not mean the precise time because we retrieved ROAs every 2 hours. Before and after the period of 'invalid', the ROA (AS59432, 185.214.100.0/24, 24) makes the route valid. AS834 is owned by a provider of IP leasing, named '`IPXO`'. It would appear that this invalid route is the result of a brief ROA adjustment, rather than the consequence of a malicious hijacking. Our methodology is unable to identify this invalid route as h-invalid since the two ASes lack a clear relationship. Moreover, we find another greater frequency of address transfer in the context of IP leasing, as illustrated in Fig. 12. The prefix 163.5.223.0/24 persists in transferring from AS834 to other ASes, which gives rise to a small number of invalid but non-hijacking routes. We also confirm the occurrence of address leasing between AS834 and other ASes in a related study [23].

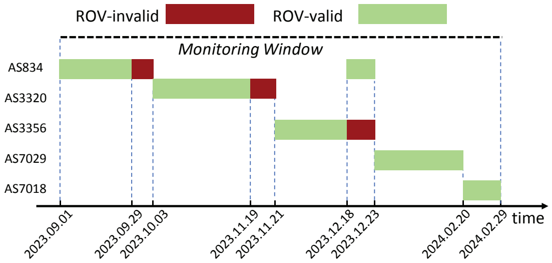


Figure 12: The origin AS and ROV status of 163.5.223.0/24.

Route visibility. The distribution bias of VPs has an impact on the global routing table, reducing route visibility. Despite our efforts to collect data from over 200 collectors of three publicly available data sources, it is possible that this bias still affects the data. Invalid routes that fail to be identified as h-invalid may be the result of incomplete route visibility, especially in the analysis of AS-path-related characteristics like path-neighbor and path-overlap. Route visibility also impacts the judgment of competitors, that is, one of the main reasons we define benign competition only (C7) as h-invalid with weak evidence and consider the combination of all characteristic analysis results as the final categories.

Subsidiaries with different 'org-id'. Our investigation reveals that some invalid routes are observed between the parent company and its subsidiaries, which had different 'org-id' in CAIDA's dataset [19]. For example, AS19527 (Google LLC) and AS139190 (Google Asia Pacific Pte. Ltd.), have different 'org-id' in CAIDA's AS organization but exhibit a notable degree of resemblance in 'org-name', and we infer they belong to the same organization. Furthermore, we infer over 100 groups (involving over 1000 ASes) of possible parent companies and subsidiaries by comparing the 'org-name and other ways, like information on their official website, by which over 900 suspiciously invalid routes would be h-invalid on the measurement of October 1, 2023. To defend the accuracy of our methodology and to avoid the potential pitfalls of inaccurate subsidiary inference, we do not include the subsidiary in the same-organization criteria. Consequently, a number of invalid routes originated by genuine subsidiary ASes are classified as s-invalid.

8 Discussion

8.1 IP Address Squatting

In the categories of invalid routes, C7 (benign competition only) is defined as h-invalid routes with weak evidence. It is a conservative choice to maintain network accessibility as we discussed in §4.6 and §5.2.3. The h-invalid routes are defined as those that preclude the possibility of malicious hijacks. Therefore, even if there might be a small number of routes that involve IP address squatting, we still categorize them as h-invalid, albeit with weak evidence.

For network operators, disregarding h-invalid routes with weak evidence would be imprudent. The observation that a quarter of the valid-in-the-future routes in the across-timeline validation (§6.2.1) are designated as weak evidence highlights the significance of these routes and suggests that they should not be overlooked.

Intuitively, we propose assigning higher weights to h-invalid routes with strong evidence, or even treating qualified invalid routes as valid. For h-invalid routes with weak evidence, more aggressive filtering rules could ignore these identification results and classify them as s-invalid routes, effectively treating them as invalid. This approach enhances security but comes with some trade-offs, potentially leading to certain prefixes becoming unreachable, altering traffic forwarding paths, or changing the destination AS, as described in the second part of the Analysis of Invalid Route Status (§2.2). A more conservative approach should assign them lower weights than those given to h-invalid routes with strong evidence, allowing operators to set their own thresholds to determine whether to retain them as valid, reject them as invalid, or deprioritize them.

8.2 AS-path Forgery

The determination of path-neighbor and path-overlap is contingent upon the AS-path. It should be noted that, at the technical level, operators can insert additional ASes in the AS-path to implement specific routing policies or to facilitate hijacking attacks, such as forged-origin hijacks [27]. However, our objective is to extract h-invalid routes on historical data rather than implement a filtering system. Consequently, there is no incentive for attackers to forge AS-paths against our path-dependent characteristics. And even if they do, it doesn't prevent their attacks from being validated as invalid in real-time. Therefore, the AS-paths we collect by default truly reflect the propagation paths of route announcements and the forwarding paths of the traffic.

On the other hand, as noted in RFC6811 [41], ROV does not protect against "AS-in-the-middle attacks" nor does it provide any path validation. It merely attempts to verify the origin. In general, if our method is directly used to formulate filtering rules, it similarly faces the threat of AS-path forgery, which is the design purpose of ASPA (Autonomous System Provider Authorization) [15].

9 Conclusion

In this work, we propose a characteristic-based methodology to assess invalid routes and find that over 91% of the daily global routing table is harmless. The characteristics also demonstrate the main reasons for h-invalid routes in the long-term measurement, which could assist operators in avoiding human error, thereby reducing false positives at the source. By our measurement, we urge ASes to accelerate the deployment of RPKI/ROV and avoid mismatching configurations between ROA and BGP. This work does not raise any ethical issues, and the contents in our open-sourced link are anonymized.

Acknowledgments

We are grateful to all the anonymous reviewers for their insightful comments. The research is supported in part by the National Natural Science Foundation of China (NSFC) under Grant No. 62472246 and Grant No.62221003. Yangyang Wang, Jia Zhang and Mingwei Xu are the corresponding authors of the paper.

References

[1] 2008. YouTube Hijacking: A RIPE NCC RIS case study. https://www.ripe.net/publications/news/industrydevelopments/youtube-hijacking-a-ripe-ncc-ris-case-study.

[2] 2019. ARTEMIS. https://labs.ripe.net/author/vasileios_kotronis/artemis-an-open-source-tool-for-detecting-bgp-prefix-hijacking-in-real-time/.

[3] 2023. BGPMon. https://www.bgpmon.net.

[4] 2023. Qrator.radar. https://qrator.net/services/radar.

[5] 2023. Qrator.radar hijack report 2023-08-29. https://x.com/Qrator_Radar/status/1749844226513232293.

[6] 2024. Qrator.radar hijack report 2024-01-23. https://x.com/Qrator_Radar/status/1749844226513232293.

[7] 2024. Qrator.radar hijack report 2024-05-25. https://x.com/Qrator_Radar/status/1794383153727160499.

[8] 2024. Qrator.radar hijack report 2024-07-09. https://x.com/Qrator_Radar/status/1810561605526245815.

[9] 2024. Qrator.radar hijack report 2024-08-22. https://x.com/Qrator_Radar/status/1826600702996873489.

[10] 2024. Qrator.radar hijack report 2024-10-20. https://x.com/Qrator_Radar/status/1848111622520447148.

[11] AFRINIC Whois Database 2024. AFRINIC Whois Database. https://www.afrinic.net/services/whois-query.

[12] APNIC Whois Database 2024. APNIC Whois Database. https://wq.apnic.net/apnic-bin/whois.pl.

[13] ARIN Whois Database 2024. ARIN Whois Database. https://whois.arin.net/ui.

[14] J. Arkko and et al. 2010. IPv4 Address Blocks Reserved for Documentation, RFC 5737. https://www.rfc-editor.org/rfc/rfc5737.

[15] ASPA 2023. BGP AS PATH Verification Based on Autonomous System Provider Authorization (ASPA) Objects. https://datatracker.ietf.org/doc/html/draft-ietf-sidrops-aspa-verification.

[16] Henry Birge-Lee, Yixin Sun, Anne Edmundson, Jennifer Rexford, and Prateek Mittal. 2018. Bamboozling certificate authorities with {BGP}. In *27th USENIX Security Symposium (USENIX Security 18)*. 833–849.

[17] R. Brandom. 2018. Hackers emptied Ethereum wallets by breaking the basic infrastructure of the internet. https://www.theverge.com/2018/4/24/17275982/myetherwallet\-hack-bgp-dns-hijacking-stolen-ethereum.

[18] R. Bush. 2018. Clarifications to BGP Origin Validation Based on Resource Public Key Infrastructure. RFC 8481. http://www.ietf.org/rfc/rfc8481.txt.

[19] CAIDA AS Organizations Dataset 2024. CAIDA AS Organizations Dataset. http://www.caida.org/data/as-organizations/.

[20] CAIDA AS Relationships Dataset 2024. CAIDA AS Relationships Dataset. https://publicdata.caida.org/datasets/as-relationships/.

[21] Wenqi Chen, Zhiliang Wang, Dongqi Han, Chenxin Duan, Xia Yin, Jiahai Yang, and Xingang Shi. 2022. ROV-MI: Large-Scale, Accurate and Efficient Measurement of ROV Deployment.. In *NDSS*.

[22] Taejoong Chung, Emile Aben, Tim Bruijnzeels, Balakrishnan Chandrasekaran, David Choffnes, Dave Levin, Bruce M Maggs, Alan Mislove, Roland van Rijswijk-Deij, John Rula, et al. 2019. RPKI is coming of age: A longitudinal study of RPKI deployment and invalid route origins. In *Proceedings of the Internet Measurement Conference*. 406–419.

[23] Ben Du, Romain Fontugne, Cecilia Testart, Alex C Snoeren, and kc claffy. 2024. Sublet Your Subnet: Inferring IP Leasing in the Wild. In *Proceedings of the 2024 ACM on Internet Measurement Conference*. 328–336.

[24] Lixin Gao. 2001. On inferring autonomous system relationships in the Internet. *IEEE/ACM Transactions on networking* 9, 6 (2001), 733–745.

[25] Yossi Gilad, Avichai Cohen, Amir Herzberg, Michael Schapira, and Haya Shulman. 2016. Are we there yet? On RPKI's deployment and security. *Cryptology ePrint Archive* (2016).

[26] Y. Gilad and et al. 2022. The Use of maxLength in the Resource Public Key Infrastructure (RPKI), RFC 9319. https://datatracker.ietf.org/doc/rfc9319/.

[27] Yossi Gilad, Omar Sagga, and Sharon Goldberg. 2017. Maxlength considered harmful to the RPKI. In *Proceedings of the 13th International Conference on emerging Networking EXperiments and Technologies*. 101–107.

[28] Vasileios Giotsas, Matthew Luckie, Bradley Huffaker, and KC Claffy. 2014. Inferring complex AS relationships. In *Proceedings of the 2014 Conference on Internet Measurement Conference*. 23–30.

[29] DAN GOODIN. 2019. BGP event sends European mobile traffic through China Telecom for 2 hours. https://arstechnica.com/information-technology/2019/06/bgp-mishap-sends-european-mobile-traffic-\through-china-telecom-for-2-hours/.

[30] Tomas Hlavacek, Italo Cunha, Yossi Gilad, Amir Herzberg, Ethan Katz-Bassett, Michael Schapira, and Haya Shulman. 2020. DISCO: Sidestepping RPKI's deployment barriers. In *Network and Distributed System Security Symposium (NDSS)*.

[31] Tomas Hlavacek, Amir Herzberg, Haya Shulman, and Michael Waidner. 2018. Practical experience: Methodologies for measuring route origin validation. In *2018 48th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*. IEEE, 634–641.

[32] Tomas Hlavacek, Philipp Jeitner, Donika Mirdita, Haya Shulman, and Michael Waidner. 2022. Behind the scenes of RPKI. In *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security*. 1413–1426.

[33] Tomas Hlavacek, Philipp Jeitner, Donika Mirdita, Haya Shulman, and Michael Waidner. 2022. Stalloris:{RPKI} downgrade attack. In *31st USENIX Security Symposium (USENIX Security 22)*. 4455–4471.

[34] Tomas Hlavacek, Haya Shulman, Niklas Vogel, and Michael Waidner. 2023. Keep Your Friends Close, but Your Routeservers Closer: Insights into RPKI Validation in the Internet. In *Proceedings of the 32nd USENIX Conference on Security Symposium*. 1–18.

[35] Tomas Hlavacek, Haya Shulman, and Michael Waidner. 2022. Smart RPKI validation: Avoiding errors and preventing hijacks. In *European Symposium on Research in Computer Security*. Springer, 509–530.

[36] Packet Clearing House. [n. d.]. PCH. https://www.pch.net/resources/Routing"-"Data.

[37] G. Huston. 2008. Autonomous System (AS) Number Reservation for Documentation Use, RFC 5398. https://www.rfc-editor.org/rfc/rfc5398.

[38] Minhyeok Kang, Weitong Li, Roland van Rijswijk-Deij, Taejoong Chung, et al. 2024. IRRedicator: Pruning IRR with RPKI-Valid BGP Insights. In *Network and Distributed System Security Symposium, NDSS 2024*.

[39] Weitong Li, Zhexiao Lin, Md Ishtiaq Ashiq, Emile Aben, Romain Fontugne, Amreesh Phokeer, and Taejoong Chung. 2023. RoVista: Measuring and analyzing the route origin validation (ROV) in RPKI. In *Proceedings of the 2023 ACM on Internet Measurement Conference*. 73–88.

[40] Donika Mirdita, Haya Schulmann, Niklas Vogel, and Michael Waidner. 2023. The CURE to vulnerabilities in RPKI validation. *arXiv preprint arXiv:2312.01872* (2023).

[41] P. Mohapatra and et al. 2013. BGP Prefix Origin Validation, RFC 6811. https://tools.ietf.org/rfc/rfc6811.txt.

[42] O. Moll. 2020. Border Gateway Protocol Hijacking - Examples and Solutions. https://www.anapaya.net/blog/bordergateway-protocol-hijacking-examples-and\-solutions.

[43] Reynaldo Morillo, Justin Furuness, Cameron Morris, James Breslin, Amir Herzberg, and Bing Wang. 2021. ROV++: Improved Deployable Defense against BGP Hijacking.. In *NDSS*.

[44] Mutually Agreed Norms for Routing Security 2024. Mutually Agreed Norms for Routing Security. https://manrs.org/netops/network-operator-actions/.

[45] NLnet Labs 2024. Routinator. https://nlnetlabs.nl/projects/\rpki/routinator.

[46] Lars Prehn and Anja Feldmann. 2021. How biased is our validation (data) for AS relationships?. In *Proceedings of the 21st ACM Internet Measurement Conference*. 612–620.

[47] Lancheng Qin, Li Chen, Dan Li, Honglin Ye, and Yutian Wang. [n. d.]. Understanding Route Origin Validation (ROV) Deployment in the Real World and Why MANRS Action 1 Is Not Followed. ([n. d.]).

[48] Lancheng Qin, Li Chen, Dan Li, Honglin Ye, and Yutian Wang. 2024. Understanding Route Origin Validation (ROV) Deployment in the Real World and Why MANRS Action 1 Is Not Followed. In *NDSS*.

[49] Lancheng Qin, Dan Li, Ruifeng Li, and Kang Wang. 2022. Themis: Accelerating the Detection of Route Origin Hijacking by Distinguishing Legitimate and Illegitimate {MOAS}. In *31st USENIX Security Symposium (USENIX Security 22)*. 4509–4524.

[50] RFC 2622 1999. Routing Policy Specification Language (RPSL). https://www.rfc-editor.org/rfc/rfc2622.

[51] RFC 2725 1999. Routing Policy System Security. https://datatracker.ietf.org/doc/html/rfc2725.

[52] RLACNIC Whois Database 2024. RIS. http://lacnic.net/cgi-bin/lacnic/whois.

[53] Nils Rodday, Ítalo Cunha, Randy Bush, Ethan Katz-Bassett, Gabi D Rodosek, Thomas C Schmidt, and Matthias Wählisch. 2021. Revisiting rpki route origin validation on the data plane. In *Proc. of Network Traffic Measurement and Analysis Conference (TMA), IFIP*.

[54] Routing Information Service 2024. RIPE NCC Whois Database. https://apps.db.ripe.net/search/query.html.

[55] Routing Information Service 2024. RIS. https://www.ripe.net/analyse/internet\measurements/routing-information-service-ris.

[56] Johann Schlamp, Ralph Holz, Quentin Jacquemart, Georg Carle, and Ernst W Biersack. 2016. HEAP: reliable assessment of BGP hijacking attacks. *IEEE Journal on Selected Areas in Communications* 34, 6 (2016), 1849–1861.

[57] Pavlos Sermpezis, Vasileios Kotronis, Alberto Dainotti, and Xenofontas Dimitropoulos. 2018. A survey among network operators on BGP prefix hijacking. *ACM SIGCOMM Computer Communication Review* 48, 1 (2018), 64–69.

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

[59] A. Siddiqui. 2020. A Major BGP Hijack by AS55410-Vodafone Idea Ltd. https://www.manrs.org/2021/04/a-major-bgphijack-by-as55410-vodafone-idea-ltd/.

[60] Yixin Sun, Maria Apostolaki, Henry Birge-Lee, Laurent Vanbever, Jennifer Rexford, Mung Chiang, and Prateek Mittal. 2021. Securing internet applications from routing attacks. *Commun. ACM* 64, 6 (2021), 86–96.

[61] Cecilia Testart, Philipp Richter, Alistair King, Alberto Dainotti, and David Clark. 2019. Profiling BGP serial hijackers: capturing persistent misbehavior in the global routing table. In *Proceedings of the Internet Measurement Conference*. 420–434.

[62] Andree Toonk. 2010. Chinese ISP hijacks the Internet. https://www.bgpmon.net/chinese-isp-hijacked-10-of-the-internet/.

[63] University of Oregon Route Views Project 2024. RouteViews. http://www.routeviews.org.

[64] Pierre-Antoine Vervier, Olivier Thonnard, and Marc Dacier. 2015. Mind Your Blocks: On the Stealthiness of Malicious BGP Hijacks.. In *NDSS*.

[65] Yang Xiang, Zhiliang Wang, Xia Yin, and Jianping Wu. 2011. Argus: An accurate and agile system to detecting IP prefix hijacking. In *2011 19th IEEE International Conference on Network Protocols*. IEEE, 43–48.

A Measurement Results on October 1, 2023

We present the measurement result and characteristic analysis on the global routing table of October 1, 2023.

Invalid routes and CVRPs. On October 1, 2023, we collected over 1.43 million routes. Using the ROA obtained at 0:00(UTC) on October 1, we validated that there are 74K invalid routes, accounting for 5.18% of the total number of routes. Among the remaining routes, 611K are valid (42.8%) and 743K are unknown (52.02%). We focus on the invalid routes and use our methodology to assess whether they are h-invalid. We record the CVRPs of each invalid route, noting that each invalid route may possess one or more types of CVRPs. Fig. 11 presents the distribution of CVRP types for the routes. First, the number of invalid routes with unmatched-length CVRPs is the highest. Second, 12.41% of invalid routes have at least two types of CVRPs, with 0.16% having three types. The number of invalid routes with multiple CVRP types is not negligible and provides more characteristic information.

Characteristic of routes. To better interpret the measurements, we present the results of characteristic analysis.

Relying on the unmatched-AS CVRPs (7204), we find that 38.89% (2802/7204) of them belong to the same-organization, and 19.72% (1421/7204) are path-neighbor. Both characteristics have approximately one-fifth hit rates with considerable independence, which could efficiently extract h-invalid routes.

According to the unmatched-length CVRPs (53570), we find that 94.81% (50791/53570) of the invalid routes are identified as path-overlap, including full and partial ones in the situation of "Directionally Identical".

Based on the unmatched-A&L CVRPs (22582), in the unmatched-AS step, we find that 40.55% (9156/22582) of them belong to the same-organization, and 37.02% (8360/22582) are path-neighbor. Next, we replace the origin AS of these cases with CVRP AS in their AS-paths and consider them as transformed unmatched-length CVRPs to execute the unmatched-length step. Consequently, 66.97% (15123/22582) invalid route entities with unmatched A&L CVRPs could be transformed. Then we find that 59.34% (8974/15123) of these route entities with transformed unmatched-length CVRPs belong to path-overlap. As a result, we consider the above 8,974 invalid route entities as double fault h-invalid (C5 and C6) routes.

As to the competitor, we consider about 83.58% of the total invalid routes to be benign competition(C1, C5, and C7). We also find that 9276 invalid route entities meet the definitions of TS-point.

In general, the results show that at least 93.76% of the invalid routes on October 1, 2023, are h-invalid, including about 65.4K h-invalid with strong evidence and about 3.9K h-invalid with weak evidence. Besides, there are 4.6K s-invalid routes.