IRR Hygiene in the RPKI Era

Ben Du[1(✉)], Gautam Akiwate[1], Thomas Krenc[1], Cecilia Testart[2], Alexander Marder[1], Bradley Huffaker[1], Alex C. Snoeren[1], and KC Claffy[1]

[1] CAIDA/UC San Diego, San Diego, USA
bendu@ucsd.edu, {gakiwate,snoeren}@cs.ucsd.edu,
{tkrenc,amarder,bhuffake,kc}@caida.org
[2] MIT, Cambridge, USA
ctestart@csail.mit.edu

Abstract. The Internet Route Registry (IRR) and Resource Public Key Infrastructure (RPKI) both emerged as different solutions to improve routing security in the Border Gateway Protocol (BGP) by allowing networks to register information and develop route filters based on information other networks have registered. RPKI is a crypto system, with associated complexity and policy challenges; it has seen substantial but slowing adoption. IRR databases often contain inaccurate records due to lack of validation standards. Given the widespread use of IRR for routing security purposes, this inaccuracy merits further study. We study IRR accuracy by quantifying the consistency between IRR and RPKI records, analyze the causes of inconsistency, and examine which ASes are contributing correct IRR information. In October 2021, we found ROAs for around 20% of RADB IRR records, and a consistency of 38% and 60% in v4 and v6. For RIPE IRR, we found ROAs for 47% records and a consistency of 73% and 82% in v4 and v6. For APNIC IRR, we found ROAs for 76% records and a high consistency of 98% and 99% in v4 and v6. For AFRINIC IRR, we found ROAs for only 4% records and a consistency of 93% and 97% in v4 and v6.

1 Introduction

The Border Gateway Protocol (BGP) is the protocol that networks use to exchange (announce) routing information across the Internet. Unfortunately, the original BGP protocol lacked mechanisms for route authentication, allowing for unauthorized announcement of network addresses, also known as *prefix hijacking* [25]. Prior to the adoption of security mechanisms like Resource Public Key Infrastructure (RPKI), the primary means of protecting against unauthorized origin announcements was to register their routing information in public databases and use these databases to verify route advertisements (for those networks with resources and incentive to do so). These databases, first deployed by various organizations in 1990s, now are collectively known as the Internet Routing Registry (IRR) [7]. However, the IRR depends on voluntary (although sometimes contractually required) contribution of routing information. Moreover, many Internet Service Providers (ISPs) are reluctant to share their routing

O. Hohlfeld et al. (Eds.): PAM 2022, LNCS 13210, pp. 321–337, 2022.
https://doi.org/10.1007/978-3-030-98785-5_14

policies to avoid leaking sensitive business information [18,26]. Perhaps more critically, the IRR information is not strictly—and sometimefirs not at all—validated. As such, the accuracy of IRR information is not guaranteed [44].

While attempts to create variations of the IRR by adding validation mechanisms have been proposed none have ever gained traction [22,27]. After years of debate, a significant set of stakeholders including many ISPs agreed on an alternative to improving routing security known as the Resource Public Key Infrastructure (RPKI) [31]. RPKI tackles the data integrity problem by allowing networks to register their prefixes with their origin AS and using cryptography to authenticate these records, with each Regional Internet Registry (RIR) operating as a "root" of trust.

Similar to the IRR, operators can use RPKI to discard routing messages that do not pass origin validation checks. Although the RPKI deployment process is standardized and network equipment supports Route Origin Validation (ROV), additional challenges arise because of the configuration and operation of relying parties [28]. Furthermore, there are concerns that RPKI gives too much power to the RIRs [17,23,43], which combined with associated legal risks [43,48] and business concerns [46] hindered its adoption.

As of now, the IRR and RPKI operate in parallel. The routing security initiative known as Mutually Agreed Norms for Routing Security (MANRS) requires its participants to use either IRR or RPKI to facilitate routing security globally but does not enforce the requirement [3]. For example, Telia Carrier (recently rebranded to Arelion), a participant of MANRS, helps its customers keep their IRR records current, and drops all RPKI invalid routes [5]. Google also requires their peers and customers to register in an IRR database [4]. In addition, various large cloud service providers, Internet Service Providers (ISPs), and transit providers such as Cloudflare, Comcast, and Cogent have registered in RPKI and deployed RPKI-based filtering [15].

To reduce cost and complexity, networks may choose not to deploy RPKI filtering and continue using only existing IRR-based route filtering. However, IRR information may be inaccurate due to *improper hygiene*, since there is no penalty to the address space owner for not updating the origin information after changes in routing policy or prefix ownership [30]. Such inaccurate information limits the ability of networks to construct correct BGP route filters and, thus, compromises routing security.

Networks who decide to move to RPKI may not (continue to) keep their IRR records accurate, which means an increase in RPKI adoption can further increase inconsistency between IRR and RPKI, and therefore widen the gap between routing decisions based on IRR and RPKI. In this paper, we study the inconsistency between the data registered in the IRR and RPKI and the underlying causes for those inconsistencies.

2 Background and Related Work

IRR and RPKI studies have focused on deployments, limitations, and impact on routing security. In contrast, we conduct a joint analysis of the IRR and RPKI to shed light on the data consistency across these two infrastructures. In this section, we provide background on both the IRR and RPKI.

2.1 Internet Routing Registry

The IRR, first introduced in 1995 as a combination of the internal routing policy storage system from RIPE and functionality extensions from Merit [8], was designed to facilitate sharing of routing policies across networks to improve routing security. Currently, the IRR consists of 25 distributed databases maintained by RIRs, commercial corporations, and non-profit organizations [33]. Networks can use the Routing Policy Specification Language (RPSL) to register and retrieve routing policy information in a set of distributed IRR databases maintained by different organizations. RPSL formats objects as lists of attribute-value pairs. The following objects are of particular relevance: (1) `mntner` objects contain authentication information required to create, modify, and delete other IRR objects; (2) `aut-num` objects contain the name and routing policies of an Autonomous System (AS). (3) `route` and `route6` objects contain IPv4 and IPv6 prefixes and their origin AS information.

The `route` and `route6` objects are particular significance to routing security. Every route objects has two mandatory attributes: `route` and `origin`. The example `route` object below shows `AS7377` intends to announce `137.110.0.0/16`.

```
route: 137.110.0.0/16  origin: AS7377
```

This data allows researchers to better understand the Internet topology and identify anomalies in BGP. Di Battista *et al.* [9] extracted BGP peering information from the IRR and Wang *et al.* [47] infer AS relationships from IRR routing policies. Shi *et al.* [42] and Schlamp *et al.* [41] used IRR information to detect and filter potential BGP hijacking events. This use of IRR data critically depends on its accuracy.

To find out the accuracy of the IRR in practice, in 2013 Khan *et al.* [26] conducted a comparative analysis of prefix origin information in IRR and BGP. They found that 87% of prefix origin pairs in 14 IRR databases matched with those in BGP. 55% of the mismatching prefix origin pairs were outdated in the IRR. They also found that the quality of IRR data depended on the routing registries, RIRs, and ASes. They found that stub ASes were more likely to register in IRR than small transit providers, followed by tier-1 transit ASes. Routing registries maintained by the RIPE NCC, APNIC, and AFRINIC had more consistency with BGP than that of ARIN and LACNIC.

2.2 Resource Public Key Infrastructure

RPKI was introduced in 2012 to help prevent BGP prefix origin hijacking. In contrast to the lack of information validation in the IRR, RPKI binds IP addresses and AS numbers to public keys using certificates. Each of the five Regional Routing Registries (RIRs) operates as the root of trust, a.k.a. *trust anchor*, for its corresponding service region. There are currently two RPKI deployment models: *hosted RPKI* and *delegated RPKI*. In hosted RPKI, The RIRs host Certificate Authority (CA) certificates and sign Route Origin Authorizations (ROAs) for the IP address space and AS numbers of their registered members. In delegated RPKI, the RIR members can host their own CA certificates to sign ROAs for their own or their customers' address space.

The most important RPKI object for BGP origin information validation is the `ROA` object. Inside a `ROA` object, `IP Prefix` specifies the IPv4 or IPv6 address resource owned by the network. `ASN` specifies the AS number used to announce the `IP Prefix` in BGP. `Max Length` specifies the length of the most specific subprefix of the `IP Prefix` allowed in BGP. The example ROA below allows `AS7377` to announce in BGP `137.110.0.0/16` and any subprefix whose length does not exceed `20`.

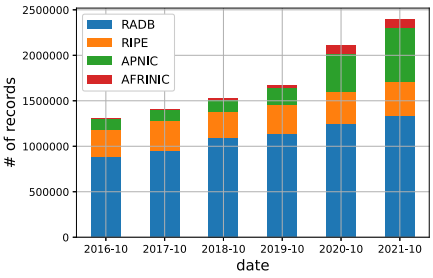
(`IP Prefix, ASN, Max Length`) (`137.110.0.0/16, AS7377, 20`)

Deploying RPKI can bring significant security benefits even with limited deployment [16], but was slow to take off due to early instances of collateral damage from insufficient/erroneous RPKI deployment [19]. Chung *et al.* [14] found that when RPKI was first deployed in 2012, 27.47% of invalid announcements were caused by misconfigurations, and by 2019, the fraction of misconfigurations dropped to 5.39%. Apparently RPKI promotion efforts have had positive effects in RPKI deployment. In 2020, Testart *et al.* [45] found that more transit and content providers had started to enforce RPKI-based filtering and thus fewer illicit BGP announcements were propagating across networks. In 2020, Kristoff *et al.* [28] found that the caching servers of up to 20% of deployed RPKI relying parties did not fetch complete or timely copies of RPKI data.

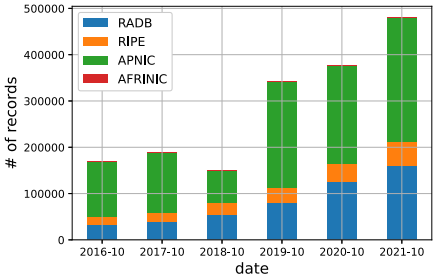
3 Datasets

The IRR and the RPKI datasets are the main focus of this paper. Additionally, we used CAIDA's Inferred AS to Organization Mappings (as2org) [12], Routeviews Prefix to AS mappings for IPv4 and IPv6 (pfx2as) [13], AS Relationships [11], and AS Rank [10] datasets to facilitate our analysis of the causes of inconsistency between the IRR and RPKI.

IRR Dataset. We collected historical IRR database dumps from the four IRR databases: the Routing Assets Database (RADB) [32], the RIPE IRR [36], the APNIC IRR [6], and the AFRINIC IRR [1]. RADB publicly hosts IRR archives starting in 2016 and we downloaded monthly snapshots from August 2016 to



(a) RADB has the most v4 records.



(b) APNIC has the most v6 records.

Fig. 1. Number of v4 and v6 IRR records in RADB, RIPE, APNIC, and AFRINIC IRR databases.

Table 1. RPKI coverage of IPv4 and IPv6 address space expanded almost 10×.

Date	IPv4			IPv6		
	ROAs	Prefixes	ASNs	ROAs	Prefixes	ASNs
2016–10	23k	22k	3,874	3.5k	3.3k	1,911
2017–10	38k	35k	5,067	6.1k	5.3k	2,519
2018–10	50k	46k	6,465	9.0k	8.1k	3,370
2019–10	92k	84k	10,232	15k	14k	5,274
2020–10	160k	143k	16,276	26k	24k	8,651
2021–10	205k	185k	21,265	40k	37k	10,878

October 2021. We downloaded CAIDA's quarterly snapshots from October 2016 to October 2021 of the RIPE, APNIC, and AFRINIC databases and obtained their **authoritative** IRR information (which only include IP address space administrated by the respective RIRs). We extracted the `route` and `route6` objects from the databases above and referred to them as the *RADB IRR dataset*, the *RIPE IRR dataset*, the *APNIC IRR dataset*, and the *AFRINIC IRR dataset* respectively. Figure 1 summarizes these datasets including their growth over time. As of October 2021, RADB had the most v4 records and the APNIC IRR had the most v6 records.

RPKI Dataset. RIPE NCC publishes daily validated ROA objects from all five RPKI trust anchors (APNIC, ARIN, RIPE NCC, AFRINIC, LACNIC) from 2011 to now, even after the retirement of their RPKI Validator [38,40]. These snapshots include both IPv4 and IPv6 RPKI information. We downloaded the monthly validated ROA archive starting August 2016 to October 2021 and refer to it as the *RPKI dataset*.By definition, each ROA can contain a list of prefixes and the MaxLength values for each prefix. But in this paper, we consider each unique `(IP Prefix, ASN, Maxlength)` a unique ROA, adhering to the data

format published by RIPE NCC [39]. Table 1 summarizes the number of IPv4 and IPv6 ROA objects.

MANRS Participants. The Mutually Agreed Norms for Routing Security (MANRS) project [3] publishes its list of participants on its website. We downloaded 787 AS numbers of network operators and CDN-and-cloud providers as of October 25, 2021 and refer to it as the *MANRS dataset*. We compared the IRR hygiene of MANRS ASes and that of other ASes in Sect. 6.

4 Methodology

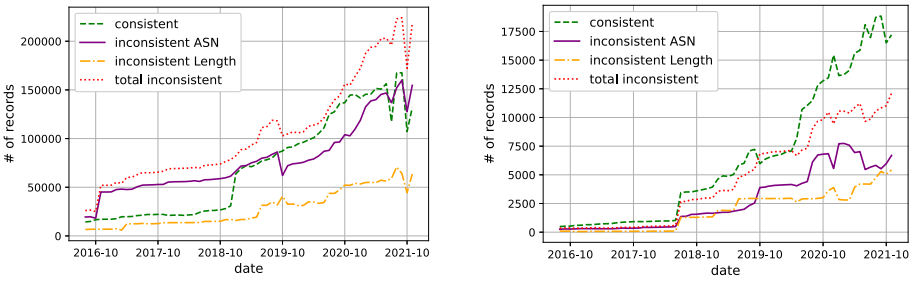
We present the steps we used to investigate the hygiene of IRR records. We use prefix origin pairs found in the *IRR datasets* and classify them according to their consistency with information found in the *RPKI dataset*. Thereby, we use the cryptographically signed ROAs as baseline for our analyses. Then, based on on the classified IRR records, we classify ASes by their maintenance practices.

4.1 Classification of IRR Records

To classify IRR records by their consistency with ROAs, we define the following four classes: Records that show full consistency in prefix and origin AS fall in **consistent**. When the ASN in an IRR record does not equal that of the ROA, the IRR record is in **inconsistent ASN**. When the ASNs are the same, but the prefix length differs, the IRR record is in **inconsistent length**. Otherwise, if there is no corresponding prefix in the ROA, the record falls into **not in RPKI**. We apply the following algorithm, a modified version of Route Origin Validation [24] and similar to the IRR Explorer [35], for the four IRR datasets *RADB IRR*, *RIPE IRR*, *APNIC IRR*, and *AFRINIC IRR*:

1 We choose snapshots of the same date from the *IRR dataset* and the *RPKI dataset*.
2 We denote each `route` or `route6` object as R_x. We denote the list of ROAs in the RPKI dataset as $ROALIST$.
3 For each record R_x, we denote the prefix as P_x and origin AS as AS_x.
4 We look for exact matching prefixes or covering prefixes of P_x in $ROALIST$. The resulting list of candidate ROAs are denoted L_{ROA}
5 If L_{ROA} is empty, then we put R_x in **not in RPKI**.
6 For each candidate ROA, C_{ROA}, in L_{ROA}, we put C_{ROA} in a list, M_{ROA}, if the origin AS in C_{ROA} equals AS_x.
7 If M_{ROA} is empty, then we classify R_x as **inconsistent ASN**.
8 For each C_{ROA} in M_{ROA}, we put C_{ROA} in a final list, V_{ROA}, if the prefix length of P_x does not exceed maxLength field in C_{ROA}.
9 If V_{ROA} is empty, we classify R_x as **inconsistent length**, otherwise as **consistent**.

Using the example from Sect. 2.2, if the *RPKI dataset* contains ROA `(137.110.0.0/16, AS7377, 20)`, a record `(137.110.0.0/24, AS195)` from the *RADB IRR dataset* falls into **inconsistent ASN** while `(137.110.0.0/24, AS7377)` falls into **inconsistent length**.



(a) RADB v4 (b) RADB v6

Fig. 2. RADB IPv6 records were more consistent with RPKI than IPv4 records.

4.2 Classification of ASes Registered in IRR

To further study the IRR record maintenance practices of different ASes, we discard all IRR records in the **Not in RPKI** category and group the ASes into three categories based on the classification of the remaining IRR records:

1) An AS is **Entirely consistent (EC)**, if all its IRR records are classified **consistent**.
2) An AS is **Entirely inconsistent (EI)**, if all associated records are classified either **inconsistent ASN** or **inconsistent length**.
3) An AS is **Mixed**, if it is associated with both **consistent** and **inconsistent** IRR records.

5 Prefix Origin Pair Consistency

Thus far, we have introduced our datasets and our methodology to determine inconsistencies in IRR records and characterize the maintaining ASes. In this section, we provide the results of our longitudinal analysis for the two *IRR datasets*, in IPv4 and IPv6, respectively. Specifically, we look at the consistency of all IRR records with a corresponding ROA in the *RPKI dataset*.

5.1 IPv4 vs. IPv6

RADB IRR. We begin by investigating the records in the *RADB IRR dataset*. From a total of 1.33M IRR records in October 2021, we found a corresponding ROA for 279,402 (21%) of the v4 prefix origin pairs. We found that 107,882 (or 38%) pairs are **consistent** with the ROA, while 127,099 (46%) showed an **inconsistent ASN** and the remaining 44,421 pairs (16%) exhibited an **inconsistent length**.

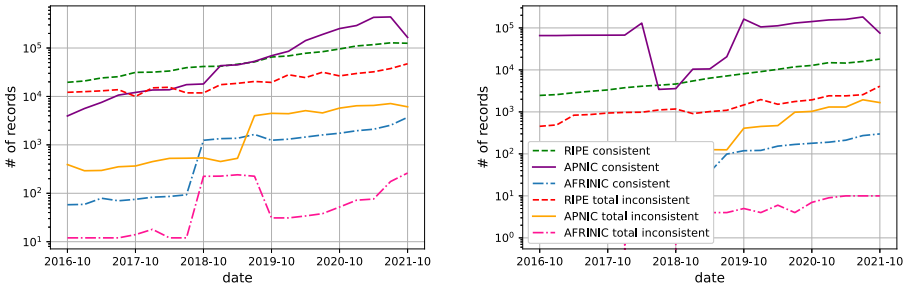
However, the fractions varied across our 5-year observation window (Fig. 2a) We observed an increase in the fraction of prefixes with a corresponding ROA,

starting with 882,220 (5%) in 2016, peaking at 30% in August 2021 and dropping to 1,335,602 (20%) in October 2021. We attribute this increase to accelerated adoption of RPKI during that time period (Table 1). The total number of **consistent** records increased by around 1000%, from 14,359 in October 2016 to its peak of 167,370 in August 2021. Also, the number of **inconsistent** records increased by around 850%, from 26,057 in October 2016 to 222,982 in August 2021.

We noticed some outstanding events in Fig. 2a: In October 2016, there was a sudden increase of 26,647 **inconsistent ASN** records. Those records were registered under `AS26415` (Verisign), with a description of `verisign customer route`. Customers of Verisign registered their prefixes under their own AS numbers (27 total) in RPKI, which caused this inconsistency. Later, in September 2019, Verisign deleted 26,682 records from RADB in an effort to clean up their records in RADB. Figure 2a also shows an increase of 34,430 **consistent** records in January 2019. Those records were registered by 10 TWNIC ASes. We speculate that this event was the outcome of TWNIC obtaining an delegated RPKI CA certificate from APNIC in late 2018 [29]. Later, from July 2021 to October 2021, fluctuations in the green line were caused by TWNIC RPKI records disappearing and reappearing from our *RPKI dataset*, possibly due to instability from the retirement of the RIPE NCC RPKI Validator [38].

Next, we look at the consistency of RADB v6 records. Figure 2b shows that for around two years until May 2018, there were few records in any category, constituting fewer than 5% of all prefix origin pairs in the *IRR RADB data set*. After that, the fraction of prefixes with a corresponding ROA increased to more than 10% and peaks 20%, which indicates a steady adoption of RPKI for v6 prefixes. Of 27,540 prefixes with a matching ROA in October 2021, around 16,506 (60%) were **consistent**, 5,977 (22%) **inconsistent ASN**, and 5,057 (18%) **inconsistent length**. Interestingly, the number of inconsistent records stabilized after October 2020, while consistent records continued to increase. This contrasts with our IPv4 observations, where the inconsistency was high. Also, on November 25, 2019 the RIPE NCC announced the complete exhaustion of IPv4 address space [37], after which the growth rate of **consistent** records increased as a potential outcome of the greater incentive to deploy IPv6 operationally.

The sudden increase of records in all three categories in June 2018 can be attributed to 2,411 **consistent** records of 7 APNIC ASes, 827 **inconsistent ASN** records of 2 (a subset of the 7) ASes, and 1,180 **inconsistent length** records of 2 ASes operated by Advanced Wireless Network Company Limited (AWN), which is a large Thai ISP. Later in September 2019, `AS45430` registered 2 prefixes in RPKI, causing 1,313 IRR records that belonged to `133481` to become **Inconsistent**. Using CAIDA's AS Rank Dataset [10], we found that both ASes belonged to AWN, and `AS45430` was the provider of `AS133481`. This is an example of the cause of some **Inconsistent ASN** cases in Sect. 5.3, where the customer AS failed to remove their IRR record after the provider AS reclaimed its address space.



(a) RIR IRR v4 (legend same as Fig. 3b) (b) RIR IRR v6

Fig. 3. In IRR databases maintained by the RIRs, the number of consistent records was at least 10 times the number of inconsistent records within the same RIR.

RIPE IRR. In October 2021 in the *RIPE IRR dataset*, there were 125,424 **consistent**, 30,764 **inconsistent ASN**, and 16,543 **inconsistent length** v4 records (Fig. 3a). For IPv6, the three categories had 18,154, 3,374, and 721 records, respectively (Fig. 3b). In the RIPE IRR, the prefix origin consistency for v4 and v6 records were similar. The consistent records have grown steadily while there have been minimal increase in the number of records in the inconsistent categories, showing good IRR hygiene over time.

Comparing the RADB and the RIPE IRR, we confirm prevailing knowledge that RIPE IRR has better-maintained records. As of October 2021, 8.3% of total records in *RADB IRR dataset* were **consistent** compared to 34.4% of total records in the *RIPE IRR dataset*. RIPE IRR's better hygiene may be due to RIPE NCC's authorization: To create a `route` object, a validation process first checks if the maintainer (`mntner`) has the authority to announce the IP prefix by either looking for parent maintainer information or referencing IP address space ownership information [36]. Our *RIPE IRR dataset* only included route objects in the authoritative RIPE IRR, which contained only prefixes managed by the RIPE NCC.

In the *RADB IRR dataset*, 28.3% of records had matching ROAs in the *RPKI dataset*, and the corresponding fraction for the *RIPE IRR dataset* was 45.2%. This shows that a larger fractions of networks registered in the RIPE IRR have also registered in the RPKI.

APNIC IRR. In July 2021 in the *APNIC IRR dataset*, there were 438,143 **consistent**, 3,426 **inconsistent ASN**, and 3,702 **inconsistent length** v4 records. For IPv6, the three categories had 182,563, 1,014, and 928 records, respectively. We found that the APNIC IRR had the highest consistency with RPKI compared to all other IRRs. The sudden drop in the solid purple line in Fig. 3b shows the number of **consistent** records decreased by 128,056 in July 2017. This decrease was caused by AS10091 (Starhub, Singapore) and AS45224 (Lanka Bell Limited, Sri Lanka) removing the entirety of their IRR records (65,491 and 62,555

records) from the APNIC IRR. Figure 3b also shows that there were significantly more v6 records in the APNIC IRR than the other RIR IRRs. Dhamdhere et al. [37] found that APNIC made a big push towards IPv6 deployment because it was the first geographical region to experience IPv4 exhaustion. We speculate that the IPv6 push caused such a high presence of IPv6 records in the APNIC IRR.

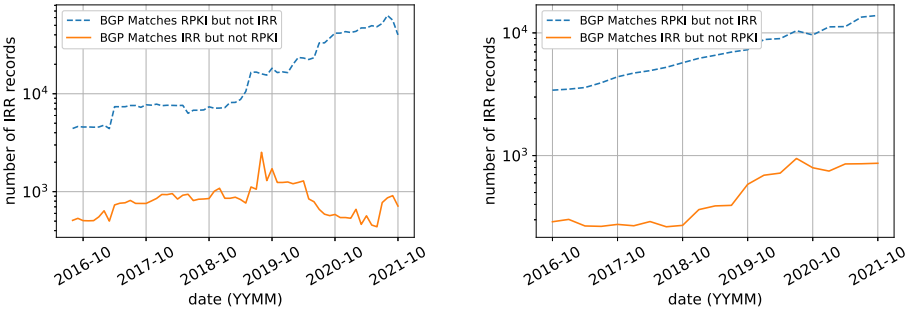
AFRINIC IRR. In October 2021 in the *AFRINIC IRR dataset*, there were 3,702 **consistent**, 180 **inconsistent ASN**, and 82 **inconsistent length** v4 records. For IPv6, the three categories had 299, 5, and 5 records, respectively. Although the number of consistent records exceeded that of inconsistent records, the AFRINIC IRR overall contained few records compared to other IRRs. Figure 3a shows a increase of both **consistent** and **inconsistent** records in October 2018 (dash-dotted blue and pink lines). We found this event is caused by AS30844 (Liquid Telecom, UK) registering its prefixes in RPKI and caused 1,082 **consistent** and 212 **inconsistent ASN** IRR records. We speculate that AFRINIC has the lowest number of IRR records compared to other RIRs because AFRINIC only launched its IRR in 2013, and had used the RIPE IRR before then [1].

Overall, the IRR databases operated by RIRs showed higher consistency with RPKI compared to RADB, as a result of the RIR's ability to regulate the creation process of `route` objects with address ownership information. In the following sections, we conduct inconsistency analysis only on RADB and RIPE IRR records due to the low inconsistency of APNIC IRR records and scarcity of AFRINIC IRR records.

5.2 Causes of Prefix Length Inconsistency

We further analyze the **inconsistent length** category. Networks that registered such records were one step away from good hygiene. Those networks successfully kept the origin ASes of their prefixes consistent in IRR and RPKI, but registered too-specific prefixes in the IRR. We speculate that this phenomenon could be caused by RPKI misconfiguration instead of bad IRR hygiene. To find out whether the networks registered inaccurate IRR records or incorrectly used the RPKI `Max Length`, we compared the **inconsistent length** records to their BGP announcement and corresponding RPKI ROAs. For each **inconsistent length** IRR record, we looked for its exact or covering prefix in BGP. If the BGP prefix origin pair was the same as that in the IRR record, we labeled it *BGP matches IRR but not RPKI*, and this indicates that the IRR record was correct and the network likely misconfigured RPKI. If the BGP prefix was less specific than the IRR record prefix length and the RPKI ROA `Max Length`, we label it *BGP matches RPKI but not IRR*, and this indicates that the IRR record was indeed inaccurate.

Figure 4a shows that as of October 2021, out of 44,421 **inconsistent length** v4 records in RADB, 713 (1.6%) agreed with BGP (*BGP matches IRR but not*



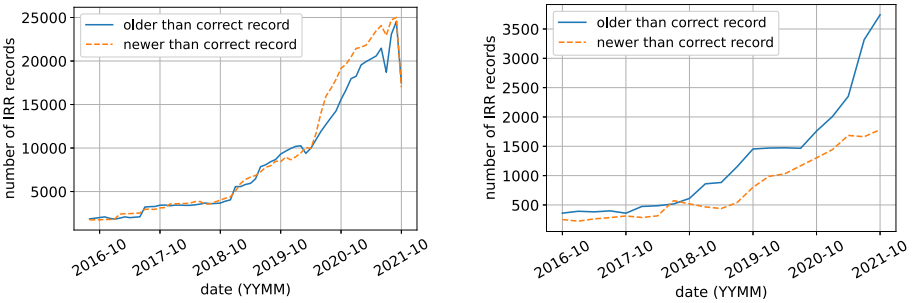
(a) RADB v4 (b) RIPE IRR v4

Fig. 4. BGP sometimes agrees with IRR records but not RPKI, showing that some IRR records in the Inconsistent Length category may actually be correct.

RPKI) and 39,968 (90.0%) disagreed with BGP (*BGP matches RPKI but not IRR*). Most prefix-length inconsistency in the IRR was caused by networks registering too-specific prefixes in the IRR, while in BGP and RPKI, they aggregated the prefixes. In much fewer cases, networks used the correct prefixes in IRR and announced them in BGP, but registered less-specific prefixes in RPKI and failed to set the proper *Max Length* attribute. Figure 4b shows a similar situation for v4 records in the RIPE IRR. Of 16,543 **inconsistent length** records, 866 (5.2%) were *BGP matches IRR but not RPKI* and 13,872 (83.9%) were *BGP matches RPKI but not IRR*. We found similar distributions for v6 records in both RADB and the RIPE IRR (not shown due to space constraints).

To summarize, prefix length inconsistency came from two types of mistakes with different operational impacts. The first type of mistake is having incorrect IRR entries that do not correctly reflect their prefix owners' routing intentions in BGP due to mismatching prefix lengths. If the upstream provider of the prefix owner requires it to register its exact BGP announcements in IRR, the prefix owner's current BGP announcements may be dropped (e.g. the upstream provider sees (`137.110.0.0/24, AS7377`) in IRR but sees (`137.110.0.0/16, AS7377`) in BGP).

The second type of mistake is misconfiguring the RPKI Max Length field. A too-small Max Length value in RPKI will almost immediately cause disruption to the prefix owner if its upstream provider uses RPKI filtering. The prefix owner's BGP announcement will be marked *RPKI invalid* and the upstream provider will drop the prefix owner's BGP announcement. Some operators reported that the Max Length feature of RPKI can cause confusion especially for RPKI newcomers because no similar feature exist in the IRR [34]. Gilad et al. [21] stated that the use of Max Length brings more harm than good to routing security, and operators have drafted proposals to discourage the use of the RPKI Max Length attribute [20].



(a) RADB v4 (b) RIPE IRR v4

Fig. 5. Sometimes inconsistent ASN records are registered more recently than their corresponding consistent records.

5.3 Analysis of ASN Inconsistency

We further analyze the causes of **inconsistent ASN** records. Networks may stop maintaining their IRR records after initial registration, inducing outdated information [26]. To study such registration practices, we compared records within the same IRR dataset. First, we defined *conflicting* records to be any two IRR records with the same IP prefix but different origin ASes. Then, we took each **inconsistent ASN** record and looked for *conflicting* records in the same IRR dataset. We called the corresponding *conflicting* record a **correct** record if it was categorized as **consistent** (Sect. 5.1). We compared the registration dates of the **inconsistent ASN** record and the **correct** record.

We found 34,174 (26.9%) [5,524 (18.0%)] **correct** records for 127,099 [30,764] **inconsistent ASN** RADB [RIPE] v4 records in the October 2021 snapshot. Figure 5a shows that in RADB, 17,319 (13.6%) records were older than their **correct** records and 16,855(13.2%) were newer. Figure 5b shows that in RIPE IRR, 3,741 (12.1%) records were older than their **correct** records and 1783 (5.8%) were newer.

This result contradicts the intuition that an **inconsistent ASN** record should be older than its **correct** counterpart, because the inaccurate IRR records are likely stale [26]. To explore this surprising phenomenon, we used the October 2021 snapshots of the *IRR datasets* and use the CAIDA AS Relationship dataset [11] from October 2021 to examine the relationship between the ASes that registered those records. We retrieved the AS relationships for 10,382 (30.4%) out of 34,174 **correct** records for RADB and 3,203 (58.0%) out of 5,524 for RIPE IRR. In RADB, we found that out of 5,468 **inconsistent ASN** records with older **correct** records, 4,464 (81.6%) **correct** records were registered by providers of the networks that registered the **inconsistent ASN** records. Similarly in RIPE IRR, out of 839 **inconsistent ASN** records, 563 (67.1%) fell into the same situation. Such a high percentage of provider-customer relationships suggests that the providers first registered their prefixes in an IRR database, and then assigned address space to their customers. Those customers also registered their assigned

prefixes in the same IRR database but failed to delete the IRR records after their providers revoked the address space. We found anecdotal evidence that some large ISPs such as Advanced Wireless Network Company Limited (AWN) in Thailand (Sect. 5.1) had customers who failed to delete outdated IRR records.

To prevent this inconsistency, providers could proactively require their customers to remove their IRR entries upon reclamation of address space to promote good IRR maintenance. Alternatively, if 0providers still intend to allow their customers to use such address space, the providers could register additional ROAs in RPKI under their customers' ASes.

6 ASes Behind IRR Inconsistency

We took the October 2021 snapshots for the *RADB IRR dataset* and the *RIPE IRR dataset* and classified the ASes according to Sect. 4.2. We used the CAIDA Inferred AS to Organization dataset [12] and the MANRS dataset [3] to classify the ASes by their RIR and whether they were MANRS participants. We labeled the ASes in RIPE, ARIN, APNIC, LACNIC, and AFRINIC regions as EU, NA, AP, SA, and SA respectively. Table 2 shows that LACNIC (SA) ASes had the best IRR hygiene among all RIRs as of October 2021. Although there were more **entirely consistent (EC)** ASes than **entirely inconsistent (EI)** ASes, the number of **consistent** records was lower than **inconsistent ASN** and **inconsistent Length** records combined. This discrepancy is because the **entirely consistent (EI)** ASes registered fewer v4 records in RADB. We also used the CAIDA AS Rank [10] dataset to look at the AS *customer cone* size distribution in each category, but found no correlation between AS size and AS registration practice in RADB. Table 3 shows that the authoritative RIPE IRR had few users outside of the RIPE service region, and the ASes had good IRR hygiene as a result of the validation requirement of the RIPE Database.

As of October 2021, of 787 MANRS ASes, 326 appeared in the *RADB IRR dataset* and had corresponding ROAs in the *RPKI dataset*. The MANRS ASes had better IRR hygiene than ASes in RADB, because the fraction of **entirely consistent (EC)** ASes were higher (53.1% vs. 45.2%). However, fewer MANRS ASes registered in the RIPE IRR compared to RADB so the fraction of **entirely consistent (EC)** ASes dropped below that of RIPE ASes (63.6% vs. 71.2%). Note that MANRS **only** requires their participants to register in either IRR or RPKI. MANRS ASes are not required to keep their records consistent between IRR and RPKI, and any MANRS AS that appears in Table 2 and Table 3 registered in both IRR and RPKI, which is more than required by MANRS.

7 Limitations

Incomplete IRR Data Coverage. We do not have historical IRR data for all IRR database providers. Although RADB mirrors all IRR databases, its IRR archives only include information directly registered in RADB itself. The number

Table 2. Classification of ASes that registered v4 records in RADB. MANRS ASes were more consistent than other ASes.

AS Class	Record	General AS count (9,675)						MANRS ASes (326)
		Total	EU	NA	AP	SA	AF	
EC	54,488	4,375 (45.2%)	562	1,084	1,184	1,452	67	173 (53.1%)
EI	367,795	3,513 (36.3%)	415	1,366	1,075	271	50	43 (13.2%)
Mixed	489,172	1,787 (18.5%)	271	428	739	324	16	110 (33.7%)

Table 3. Classification of ASes that registered v4 records in RIPE IRR. Most ASes were in the RIPE region and were highly consistent.

AS class	Record	General AS count (13,109)						MANRS ASes (220)
		Total	EU	NA	AP	SA	AF	
EC	75,589	9,339 (71.2%)	9,039	175	85	31	6	140 (63.6%)
EI	18,613	1,478 (11.3%)	1,309	86	63	7	4	7 (3.2%)
Mixed	144,284	2,292 (17.5)	2,193	70	26	0	3	73 (33.2%)

of consistent and inconsistent IRR records in our analysis is treated as a lower bound of the actual situation.

Sparse IRR Data Granularity. Changes to IRR and RPKI databases happen daily or even hourly as networks change configurations to adapt to routing needs. Our dataset does not have the granularity to monitor such frequent events and provides only a longitudinal analysis.

Aggregated RPKI Data. The RPKI infrastructure has a hosted model and delegated model, which may have different consistency with IRR databases. We cannot distinguish which model our RPKI data is collected from.

8 Summary

The recent growth of RPKI usage gives us the opportunity to study the accuracy of the IRR. In this paper we explored IRR hygiene by comparing the consistency between IRR and RPKI records and analyzing the IRR maintenance practices of ASes. Although RPKI has gained popularity, it still has far fewer participating ASes or database records. Comparing RADB and RPKI, we found 61.4% of v4 and 40.1% of v6 records (that appeared in both databases) were inconsistent. In contrast, the RIPE IRR had only 27.4% v4 and 18.4% v6 inconsistent records. We discovered some causes of inconsistency: complicated customer-provider relationships among ASes in the IRR, and possible misconfiguration in the RPKI. Finally, we found that ASes participating in the routing security intiative MANRS were more likely to keep IRR records consistent with RPKI than ASes in general.

Future Work. Our work helps to broadly identify inaccurate and suspicious IRR records and can serve as the foundation for IRR false registration detection.

On the operational side, the future of IRR can be promising, as new tools such as IRRd Version 4 [2] have been developed to help operators automatically validate IRR information against RPKI. This could further improve the accuracy of the IRR and contribute to better routing security.

Acknowledgment. This material is based on research sponsored by the National Science Foundation (NSF) grants CNS-1901517, OAC-2131987, CNS-2120399, and OAC-1724853. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of NSF. We appreciate network operators who gave us valuable insight. We also thank our shepherd for the helpful feedback.

References

1. AFRINIC's Internet Routing Registry (2021). https://afrinic.net/internet-routing-registry
2. IRRd Version 4.2.0 (2021). https://irrd.readthedocs.io/en/stable/
3. Mutually Agreed Norms for Routing Security (2021). https://www.manrs.org/
4. Peering with Google (2021). https://peering.google.com/#/options/peering
5. Routing Security (2021). https://www.teliacarrier.com/our-network/bgp-routing/routing-security.html
6. APNIC Internet Routing Registry (2022). https://www.apnic.net/about-apnic/whois_search/about/what-is-in-whois/irr/
7. Alaettinoglu, C., et al.: Routing policy specification language (RPSL). RFC 2622, RFC Editor, June 1999
8. Bates, T., et al.: Representation of IP routing policies in a routing registry (ripe-81++). RFC 1786, RFC Editor, March 1995
9. Battista, G.D., Refice, T., Rimondini, M.: How to extract BGP peering information from the internet routing registry. In: Proceedings of the 2006 SIGCOMM Workshop on Mining Network Data, MineNet 2006, pp. 317–322. Association for Computing Machinery, New York (2006). https://doi.org/10.1145/1162678.1162685
10. CAIDA: AS Rank (2021). https://asrank.caida.org/
11. CAIDA: AS Relationships (2021). https://www.caida.org/catalog/datasets/as-relationships/
12. CAIDA: Inferred AS to Organization Mapping Dataset (2021). https://www.caida.org/catalog/datasets/as-organizations/
13. CAIDA: Routeviews Prefix to AS mappings Dataset (pfx2as) for IPv4 and IPv6 (2021). https://www.caida.org/catalog/datasets/routeviews-prefix2as/
14. Chung, T., et al.: RPKI is coming of age: a longitudinal study of RPKI deployment and invalid route origins. In: Proceedings of the Internet Measurement Conference, IMC 2019, pp. 406–419. Association for Computing Machinery, New York (2019). https://doi.org/10.1145/3355369.3355596
15. Cloudflare: Is BGP Safe Yet? (2022). https://isbgpsafeyet.com/
16. Cohen, A., Gilad, Y., Herzberg, A., Schapira, M.: One hop for RPKI, one giant leap for BGP security. In: Workshop on Hot Topics in Networks, 7 p., November 2015
17. Cooper, D., Heilman, E., Brogle, K., Reyzin, L., Goldberg, S.: On the risk of misbehaving RPKI authorities. In: Proceedings of the Twelfth ACM Workshop on Hot Topics in Networks. HotNets-XII, Association for Computing Machinery, New York (2013). https://doi.org/10.1145/2535771.2535787

18. Durand, J., Pepelnjak, I., Doering, G.: BGP operations and security. BCP 194, RFC Editor, February 2015
19. Gilad, Y., Cohen, A., Herzberg, A., Schapira, M., Shulman, H.: Are we there yet? On RPKI's deployment and security. In: Network and Distributed System Security Symposium (2017)
20. Gilad, Y., Goldberg, S., Sriram, K., Snijders, J., Maddison, B.: The use of maxLength in the RPKI. Internet-Draft draft-ietf-sidrops-rpkimaxlen-09, IETF Secretariat, November 2021. https://www.ietf.org/archive/id/draft-ietf-sidrops-rpkimaxlen-09.txt
21. Gilad, Y., Sagga, O., Goldberg, S.: MaxLength considered harmful to the RPKI. In: Proceedings of the 13th International Conference on Emerging Networking EXperiments and Technologies, CoNEXT 2017, pp. 101–107. Association for Computing Machinery, New York (2017). https://doi.org/10.1145/3143361.3143363
22. Goodell, G., Aiello, W., Griffin, T., Ioannidis, J., McDaniel, P.D., Rubin, A.D.: Working around BGP: an incremental approach to improving security and accuracy in interdomain routing. In: NDSS, vol. 23, p. 156. Citeseer (2003)
23. Heilman, E., Cooper, D., Reyzin, L., Goldberg, S.: From the consent of the routed: improving the transparency of the RPKI, SIGCOMM 2014, pp. 51–62. Association for Computing Machinery, New York (2014). https://doi.org/10.1145/2619239.2626293
24. Huston, G., Michaelson, G.: Validation of route origination using the resource certificate public key infrastructure (PKI) and route origin authorizations (ROAs). RFC 6483, RFC Editor, February 2012
25. Huston, G., Rossi, M., Armitage, G.: Securing BGP - a literature survey. IEEE Commun. Surv. Tutor. **13**(2), 199–222 (2011). https://doi.org/10.1109/SURV.2011.041010.00041
26. Khan, A., Kim, H.C., Kwon, T., Choi, Y.: A comparative study on IP prefixes and their origin ASes in BGP and the IRR. SIGCOMM Comput. Commun. Rev. **43**(3), 16–24 (2013). https://doi.org/10.1145/2500098.2500101
27. Kim, E.Y., Xiao, L., Nahrstedt, K., Park, K.: Secure interdomain routing registry. IEEE Trans. Inf. Forensics Secur. **3**(2), 304–316 (2008). https://doi.org/10.1109/TIFS.2008.922050
28. Kristoff, J., et al.: On measuring RPKI relying parties. In: Proceedings of the ACM Internet Measurement Conference, IMC 2020, pp. 484–491. Association for Computing Machinery, New York (2020). https://doi.org/10.1145/3419394.3423622
29. Ku, C.H.: 98% of Taiwan's IP address holders have signed RPKI ROAs (2020). https://blog.apnic.net/2020/10/16/98-of-taiwans-ip-address-holders-have-signed-rpki-roas/
30. Kuerbis, B., Mueller, M.: Internet routing registries, data governance, and security. J. Cyber Policy **2**(1), 64–81 (2017)
31. Lepinski, M., Kent, S.: An infrastructure to support secure internet routing. RFC 6480, RFC Editor, February 2012. http://www.rfc-editor.org/rfc/rfc6480.txt
32. Merit Network: The Internet Routing Registry - RADb (2021). https://www.radb.net/
33. Merit Network, Inc.: Internet Routing Registry (2018). http://www.irr.net
34. Michaelson, G.: IRR and RPKI: a problem statement (2017). https://conference.apnic.net/44/assets/files/APCS549/Global-IRR-and-RPKI-a-problem-statement.pdf
35. NLNOG: IRR explorer (2021). https://irrexplorer.nlnog.net/
36. RIPE NCC: Managing Route Objects in the IRR (2019). https://www.ripe.net/manage-ips-and-asns/db/support/managing-route-objects-in-the-irr

37. RIPE NCC: The RIPE NCC has run out of IPv4 Addresses (2019). https://www.ripe.net/publications/news/about-ripe-ncc-and-ripe/the-ripe-ncc-has-run-out-of-ipv4-addresses
38. RIPE NCC: Ending Support for the RIPE NCC RPKI Validator (2021). https://www.ripe.net/publications/news/announcements/ending-support-for-the-ripe-ncc-rpki-validator
39. RIPE NCC: RPKI Dataset (2021). https://ftp.ripe.net/ripe/rpki/
40. RIPE NCC: RPKI Validator (2021). https://rpki-validator.ripe.net/
41. Schlamp, J., Holz, R., Jacquemart, Q., Carle, G., Biersack, E.W.: HEAP: reliable assessment of BGP hijacking attacks. IEEE J. Sel. Areas Commun. **34**(6), 1849–1861 (2016). https://doi.org/10.1109/JSAC.2016.2558978
42. Shi, X., Xiang, Y., Wang, Z., Yin, X., Wu, J.: Detecting prefix hijackings in the internet with argus. In: Proceedings of the 2012 Internet Measurement Conference, IMC 2012, pp. 15–28. Association for Computing Machinery, New York (2012). https://doi.org/10.1145/2398776.2398779
43. Shrishak, K., Shulman, H.: Limiting the power of RPKI authorities. In: Proceedings of the Applied Networking Research Workshop, ANRW 2020, pp. 12–18. Association for Computing Machinery, New York (2020). https://doi.org/10.1145/3404868.3406674
44. Testart, C.: Reviewing a historical internet vulnerability: why isn't BGP more secure and what can we do about it? TPRC (2018)
45. Testart, C., Richter, P., King, A., Dainotti, A., Clark, D.: To filter or not to filter: measuring the benefits of registering in the RPKI today. In: Sperotto, A., Dainotti, A., Stiller, B. (eds.) PAM 2020. LNCS, vol. 12048, pp. 71–87. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-44081-7_5
46. Wählisch, M., Schmidt, R., Schmidt, T.C., Maennel, O., Uhlig, S., Tyson, G.: RiPKI: the tragic story of RPKI deployment in the web ecosystem. In: Proceedings of Fourteenth ACM Workshop on Hot Topics in Networks (HotNets). ACM, New York (2015)
47. Wang, F., Gao, L.: On inferring and characterizing internet routing policies. In: IMC 2003, pp. 15–26. Association for Computing Machinery, New York (2003). https://doi.org/10.1145/948205.948208
48. Yoo, C.S., Wishnick, D.A.: Lowering legal barriers to RPKI adoption. U of Penn Law School, Public Law Research Paper (19-02) (2019)