Poster: IRR Hygiene in the RPKI Era

Ben Du
UC San Diego
bendu@ucsd.edu

Alex C. Snoeren
UC San Diego
snoeren@cs.ucsd.edu

kc claffy
UC San Diego
kc@caida.org

Problem and motivation

The Border Gateway Protocol (BGP) is the protocol that networks use to exchange (announce) routing information across the Internet. Unfortunately, BGP has no mechanism to prevent unauthorized announcement of network addresses, also known as *prefix hijacks*. Since the 1990s, the primary means of protecting against unauthorized origin announcements has been the use of routing information databases, so that networks can verify prefix origin information they receive from their neighbors in BGP messages. In the 1990s, operators deployed databases now collectively known as the Internet Routing Registry (IRR), which depend on voluntary (although sometimes contractually required) contribution of routing information without strict (or sometimes any) validation. Coverage, accuracy, and use of these databases remains inconsistent across ISPs and over time.

In 2012, after years of debate over approaches to improving routing security, the operator community deployed an alternative known as the Resource Public Key Infrastructure (RPKI). The RPKI includes cryptographic attestation of records, including expiration dates, with each Regional Internet Registry (RIR) operating as a "root" of trust. Similar to the IRR, operators can use the RPKI to discard routing messages that do not pass origin validation checks. But the additional integrity comes with complexity and cost. Furthermore, operational and legal implications of potential malfunctions have limited registration in and use of the RPKI. In response, some networks have redoubled their efforts to improve the accuracy of IRR registration data. These two technologies are now operating in parallel, along with the option of doing nothing at all to validate routes.

Although RPKI use is growing, its limited coverage means that security-conscious operators may query both IRR and RPKI databases to maximize routing security. However, IRR information may be inaccurate due to improper hygiene, such as not updating the origin information after changes in routing policy or prefix ownership. Since RPKI uses a stricter registration and validation process, we use it as a baseline against which to compare the trends in accuracy and coverage of IRR data.

Related work

Researchers have compared prefix-origin pairs in both IRR and RPKI to those observed in BGP announcements and routing tables, finding 80% (2013) [2] and 90.4% (2019) [1] of records are consistent with BGP announcements, respectively. However, we know of no detailed comparisons between IRR and RPKI databases since the launch of RPKI.

Dataset and methodology

Our dataset consists of the *IRR dataset*, a longitudinal dataset of an IRR database, and *RPKI dataset*, a longitudinal dataset of RPKI databases. To build the *IRR dataset*, we collected the IRR database archives from the Routing Assets Database (RADB). RADB provides the largest IRR database mirror on the Internet, including IRR databases hosted by 34 other organizations such as NTT and RIPE NCC. We downloaded monthly snapshots of RADB from August 2016 to September 2021 and extracted *route objects* for their IP prefixes and origin AS information.

To build the *RPKI dataset*, we collected validated ROA objects (verified prefix origin information) from RIPE NCC's RPKI validator. RIPE NCC publishes validated ROA objects from all five RPKI trust anchors (APNIC, ARIN, RIPE NCC, AFRINIC, LACNIC). We download the monthly ROA archive starting August 2016 to September 2021 to match the interval of our *IRR dataset*.

Our methodology explores differences between the information registered in the *IRR dataset (IRR)* and the *RPKI dataset (RPKI)* from the following two aspects: database completeness and record consistency. To study database completeness, we look in each dataset at the number of IP prefixes and ASes and their coverage of the allocated IPv4 address space. To study record consistency, we take all records in our *IRR dataset* and perform RPKI validation on those prefixes, similar to the mechanism of BGP route origin validation. We put the IRR records into 4 categories – *consistent*, *inconsistent ASN*, *inconsistent maxLength*, and *not in RPKI* by applying the following validation logic:

1. For each record R_x in the *IRR dataset*, we denote the prefix as P_x and origin AS as AS_x.
2. We look for an exact matching prefix or covering prefixes of P_x in the ROA objects in the RPKI dataset. The resulting list of candidate ROAs are denoted L_{ROA}
3. If L_{ROA} is empty, then we put R_x in *not in RPKI*.
4. 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.
5. If M_{ROA} is empty, then we classify R_x as *inconsistent ASN*.
6. 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}.

7 If V_{ROA} is empty, we classify R_x as *inconsistent maxLength*, otherwise as *consistent*.

Results and contribution

Our results summarize the completeness and consistency between the *IRR dataset* and *RPKI dataset*. Table 1 shows that *IRR dataset* contained 38 times more prefixes and covers almost 7 times more allocated IPv4 address space than RPKI in 2016. Table 1 also shows that RPKI almost doubled its number of prefixes over 6 years, more rapid growth than IRR.

	IRR			RPKI		
Year	Prefix	ASN	IP Space	Prefix	ASN	IP Space
2016	769k	24,112	70.52%	20k	3,741	11.62%
2017	813k	27,151	73.39%	34k	4,918	14.25%
2018	900k	30,531	74.23%	44k	6,185	15.08%
2019	958k	33,608	74.73%	75k	9,394	23.55%
2020	1M	37,427	82.59%	128k	15,039	35.06%
2021	1.06M	40,574	92.73%	209k	23,472	49.26%

Table 1: RPKI is growing faster than IRR, but the *IRR dataset* is still more complete than the *RPKI dataset*.

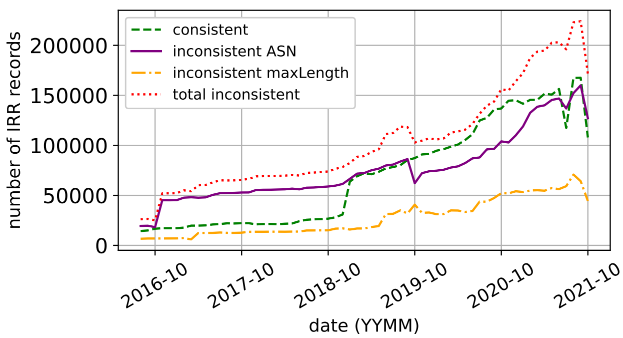

Figure 1: There are more conflicting records than agreeing records between IRR and RPKI.

Record consistency We observed substantial inconsistency between IRR records and corresponding ROAs in RPKI. The number of IRR records in each category has grown as RPKI gained popularity, but notably there are more inconsistent records between IRR and RPKI than consistent ones (Fig. 1). The purple line reflects a significant uptick of *inconsistent ASN* records in October 2016, when 26 thousand more IRR records were inconsistent with their corresponding RPKI records. Those 26 thousand IRR records were registered under Verisign's AS number, with a description of `verisign customer route`. In October 2016, those customer ASes of Verisign registered their prefixes in RPKI under their own AS numbers, causing significant inconsistency between IRR and RPKI. In September 2019, Verisign deleted those conflicting records from the IRR, causing the purple line downtick in Figure 1. In January 2019, the green line shows an increase of *consistent* records, caused by TWNIC ASes bulk registering their prefixes in RPKI.

Table 2 classifies ASes into three categories based on their record consistency. We define *record age* as the number of years from a record's last update to September 2021. Figure 2 shows that the *record age* of records maintained by **entirely inconsistent ASes** are older than records by ASes in other categories. This disparity suggests that some RPKI users may have given up on IRR hygiene long before they began using the RPKI.

AS Class	AS Count	Records	
		Consistent	Inconsistent
Entirely Consistent	4326	31,897	0
Entirely Inconsistent	3600	0	47,395
Mixed	2040	123,609	156,552

Table 2: More ASes keep their entire IRR records consistent with RPKI.

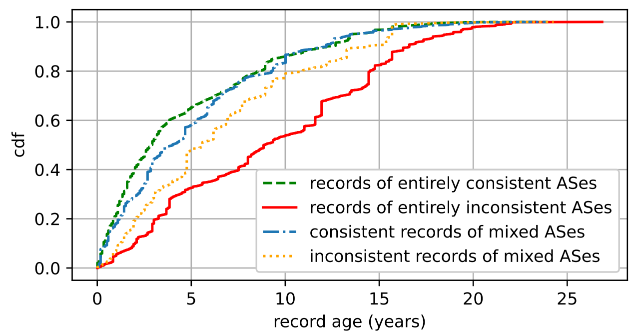

Figure 2: Entirely inconsistent ASes are less likely to update their IRR records.

The main contribution of this work is to explore IRR hygiene by comparing completeness and consistency between IRR and RPKI. We find the rapid growth of RPKI adoption helpful for measuring IRR correctness and better routing security. However, because lack of consistency suggests stale IRR data, tools that identify such inconsistencies can help those wanting (or willing) to maximize the utility of both platforms. We will expand our analysis to correlate usage and consistency with other network properties, e.g., network size, location, type, that reveal insight into how the ecosystem is evolving.

REFERENCES

[1] Taejoong et al. Chung. 2019. RPKI is Coming of Age: A Longitudinal Study of RPKI Deployment and Invalid Route Origins. In *Proceedings of the Internet Measurement Conference (IMC '19)*.

[2] Akmal et al. Khan. 2013. A Comparative Study on IP Prefixes and Their Origin Ases in BGP and the IRR. *SIGCOMM Comput. Commun. Rev.* (July 2013), 16–24.