POSTER: Using RPKI to Aggregate Autonomous Systems by their Managing Organization

Deepak Gouda
deepakgouda@gatech.edu
Georgia Institute of Technology
Atlanta, USA

Cecilia Testart
ctestart@gatech.edu
Georgia Institute of Technology
Atlanta, USA

Abstract

Accurate mapping of Autonomous Systems (ASes) to their owner organizations is fundamental for understanding the structure and dynamics of the Internet. However, as AS numbers have traditionally been delegated in an ad-hoc manner and organizational ownership has evolved over time, many organizations have registered resources under different names. Traditionally, researchers have relied on datasets like AS2Org, which map ASNs to organizations primarily using WHOIS records, but WHOIS inconsistencies often lead to missed and false relationships. We propose a new approach by leveraging the Resource Public Key Infrastructure (RPKI) to map ASNs to their managing organization. Our methodology combines multiple data sources: WHOIS records to extract organization names, RPKI certificates to identify potential siblings, and Large Language Models (LLMs) to find evidence not visible in WHOIS records currently. This integrated approach enables a more robust and accurate mapping of ASNs to organizations, notably improving inferences for 14% of multi-ASN clusters.

CCS Concepts

• **Networks** → **Network measurement**.

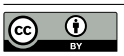
Keywords

Organization Mapping, WHOIS, RPKI, Sibling ASN, LLMs

ACM Reference Format:
Deepak Gouda and Cecilia Testart. 2025. POSTER: Using RPKI to Aggregate Autonomous Systems by their Managing Organization. In *ACM SIGCOMM 2025 Conference (SIGCOMM Posters and Demos '25), September 8–11, 2025, Coimbra, Portugal.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3744969.3748431

1 Introduction

The Internet is composed of thousands of independently administered networks, known as Autonomous Systems (ASes), each managed by a single administrative entity. This mapping enables a wide range of applications, including the analysis of Internet resource usage, investigation of traffic engineering, identification of misconfigurations and malicious activity on the Internet. Traditionally, researchers have relied on datasets like AS2Org [3] and more recently [1, 4], which map ASNs to organizations primarily using WHOIS records, and additionally PeeringDB data to fix some mappings. However, WHOIS inconsistencies and limited coverage of PeeringDB often lead to missed and false relationships between ASNs and organizations. To address these challenges, we leverage RPKI certificate infrastructure. Although RPKI certificates do not explicitly name the organization, they do contain the set of resources controlled by the same entity. This hints at the possibility of inferring sibling relationships.



SIGCOMM Posters and Demos '25, Coimbra, Portugal

ACM ISBN 979-8-4007-2026-0/2025/09
https://doi.org/10.1145/3744969.3748431

2 Related Work

The first methodology for mapping ASes to organizations was introduced by Cai *et al.* [2], and is implemented in the widely used CAIDA AS2Org [3] dataset. This approach relies on the Org-ID field and organization names from WHOIS records maintained by the Regional and National Internet Registries (RIRs and NIRs). While this method provides broad coverage, it is limited by the inconsistencies in how organizations register their resources and between WHOIS databases. Subsequent improvements to AS2Org by Chen *et al.* [4] and Arturi *et al.* [1], incorporated PeeringDB data to address issues with WHOIS data. However, these methods are limited by PeeringDB's partial coverage of ASNs and the availability and accuracy of voluntary data provided by operators. In contrast, our methodology relies on a source of data designed to provide verifiable attestations of Internet resources that has broad adoption. Indeed, we find more than 80% of routed ASNs in RPKI certificates.

3 RPKI Background

RPKI is a specialized PKI designed to secure Internet routing by enabling an entity to "verifiably assert that it is the legitimate holder of a set of IP addresses or a set of Autonomous System (AS) numbers" [7]. In this mechanism, a resource holder is issued a Resource Certificate (RC), which serves as a cryptographically verifiable attestation of the entity's right to use the Internet resources listed in the certificate. The certificate grants the holder a cryptographic key to issue further RPKI certificates, such as a Route Origin Authorizations (ROAs)-the most deployed use case of RPKI.

The structure of the RPKI certificate trees mirrors the hierarchical allocation of resources: from IANA to RIRs, then to Local Internet Registries (LIRs) and organizations. The RIRs issue direct-delegation RCs to entities such as ISPs, LIRs, or individual organizations to whom they have directly delegated sets of Internet resources. In this work, we use these direct-delegation RCs to infer the set of ASNs held by a single entity. Importantly, resources are grouped together in RPKI based on registry allocations, not on shared organization names or Org-IDs from WHOIS records. However, unlike WebPKI, RPKI certificates do not contain explicit organizational information. However, we can analyze the WHOIS records of listed Internet resources to identify the potential certificate holder.

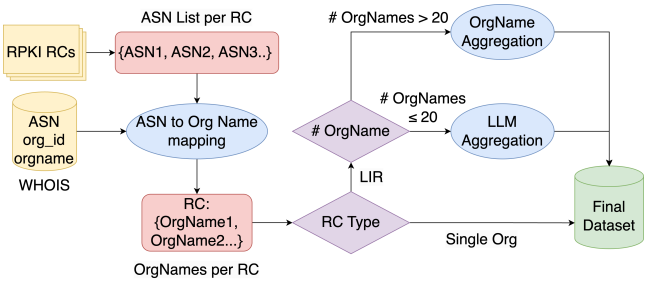


Figure 1: Methodology

4 Dataset and Methodology

Datasets: For this study, we utilize the RPKI certificate repository archive from the RPKIviews [9], which consists of all validated RPKI certificates. We use the first snapshot available for April 1, 2025. From this snapshot, we extract RCs and analyze their "resources" fields to identify the sets of ASNs associated with each certificate. We use WHOIS dataset to extract organization information.

Methodology: Our methodology for grouping ASNs by organization is outlined in Figure 1. We extract the set of ASNs from each RC[1]. Using WHOIS records, we identify the organization names associated with each ASN. If all ASNs map to a single organization name, we attribute the certificate to that organization.

If a single certificate lists multiple organization names, the certificate holder may be registering its resources under various names. Alternatively, the certificate could belong to an NIR or LIR, in which case it might include hundreds of organization names. This occurs because NIRs and LIRs receive large ASN blocks from RIRs and subsequently allocate them to their customers, resulting in individual RCs, that consist of resources allocated to distinct organizations.

LLM Aggregation: To focus on certificates likely owned by individual organizations rather than NIRs or LIRs, we analyze only RCs with fewer than 20 unique organization names. We use a Large Language Model (Sonar model by Perplexity) to assess whether different names refer to the same entity, considering alternative names, subsidiaries, and mergers. We prompt the model to provide us with a similarity score between each pair of organization names and supporting links to external sources, which help us verify the LLM outputs. The similarity scores are either very low or very high. We set a threshold of 70% similarity score to assert that two organization names belong to the same entity.

Very large RCs: We manually evaluated 23 RCs with more 20 unique organization names, identifying 9 belonging to individual organizations. For the remaining certificates, we group the ASNs by the organization name.

5 Results and Discussion

Across all RCs, we identified 107,999 ASNs, covering 82.8% of all routed ASNs on the Internet. Of the remaining 14,403 ASNs, 13,979 ASNs are ARIN legacy resources delegated before the RIR system, while the rest are absent from delegation files. We found 34,272 RCs with at least one ASN. 99.20% of these (covering 46,248 ASNs), mapped all ASNs to a single organization name, while 270 RCs (0.8%) included ASNs registered under multiple organization name.

To address organizational ambiguity, we focused on 247 RCs (encompassing 3,281 ASNs) with fewer than 20 unique organization names. We applied our LLM-based search methodology on these RCs, with an 80% similarity threshold. We manually verified most sources and links for these results. For example, our approach correctly grouped the ASNs of *Servers Australia Pty Ltd* and *Oz Servers Pty Ltd* into a single cluster of 12 ASNs (versus two clusters of 9 and 3 ASNs in traditional datasets) since they are related by a merger [6]. Our approach also combined the ASNs of *Vocus Communications* and its subsidiaries—*Vocus Pty Ltd* and *Nextgen Networks* [5, 8], into one cluster of 57 ASNs. Traditional datasets split them into three separate clusters of 53, 3 and 1 ASN.

Comparison with existing datasets: To evaluate our methodology, we compared our results against AS2Org with sibling inference datasets (referred to as AS2Org++). Our approach identified 6572 ASN clusters in AS2Org++ (covering 7884 ASNs), that consist of incorrect or missing links. Breakdown of required modifications: straightforward RPKI-based grouping modifies 8 clusters, LLM-based inference modifies 1992, and manual analysis of large RCs results aggregates 4544 clusters, while processing the 14 large RCs lead to modification of 28 clusters. The final dataset consists of 84,004 clusters, covering 108K ASNs. The total number of clusters (organizations) and ASNs is summarized in Table 1.

Method	Our Clusters	# modified AS2Org++ clusters
RPKI Grouping	33,226	8
LLM Inference	695	1992
Manual (9 RCs, Group)	9	4544
Manual (14 RCs)	50,074	28

Table 1: Comparison of ASN-to-Organization clusters by our methodology against existing ASN-to-Organization datasets

6 Limitations and Future Work

Local Internet Registries: LIRs and NIRs certificates often aggregate large numbers of ASNs belonging to many organizations. This makes it impractical to run LLM-based similarity checks for all possible organization name pairs. Our approach still reduces the search space and future research inspired by our manual investigation can tackle this.

Organizations with Multiple RCs: Some organizations hold multiple RCs, sometimes under different names or across different RIRs. When ASNs are registered under varying names or certificates, additional datasets such as the operator-maintained PeeringDB are needed to accurately map ASNs to organizations. Integrating such datasets is part of our planned future work.

Acknowledgements

This work is supported by the National Science Foundation grant CIEC-2419735. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

[1]We note that ROAs are issued from prefix attestations and there is no verification that the origin ASN in the ROA must relate to the prefix owner.

A Ethics

This work raises no ethical concerns.

References

[1] Augusto Arturi, Esteban Carisimo, and Fabián E. Bustamante. 2023. as2org+: Enriching AS-to-Organization Mappings with PeeringDB. In *Passive and Active Measurement: 24th International Conference, PAM 2023, Virtual Event, March 21–23, 2023, Proceedings*. Springer-Verlag, Berlin, Heidelberg, 400–428. doi:10.1007/978-3-031-28486-1_17

[2] Xue Cai, John Heidemann, Balachander Krishnamurthy, and Walter Willinger. 2010. Towards an AS-to-organization map. In *Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement* (Melbourne, Australia) *(IMC '10)*. Association for Computing Machinery, New York, NY, USA, 199–205. doi:10.1145/1879141.1879166

[3] CAIDA. 2020. *Mapping Autonomous Systems to Organizations: CAIDA's Inference Methodology*. https://www.caida.org/archive/as2org/

[4] Zhiyi Chen, Zachary S. Bischof, Cecilia Testart, and Alberto Dainotti. 2023. Improving the Inference of Sibling Autonomous Systems. In *Passive and Active Measurement: 24th International Conference, PAM 2023, Virtual Event, March 21–23, 2023, Proceedings*. Springer-Verlag, Berlin, Heidelberg, 345–372. doi:10.1007/978-3-031-28486-1_15

[5] Vocus Communications. 2016. VOCUS ANNOUNCES ACQUISITION OF NEXTGEN NETWORKS AND NWCS DEVELOPMENT PROJECT SUPPORTED BY A $652 MILLION CAPITAL RAISING. https://vocuscommunications.gcs-web.com/static-files/5aa34c88-e260-47dd-85c5-0b70de354b17

[6] Michael Jenkin. 2017. Servers Australia acquires Brisbane hosting provider Oz Servers. https://www.techpartner.news/news/servers-australia-acquires-brisbane-hosting-provider-oz-servers-473488?eid=4&edate=20170918

[7] M. Lepinski and S. Kent. 2012. An Infrastructure to Support Secure Internet Routing. RFC 6480 (Informational), 24 pages. https://www.rfc-editor.org/rfc/rfc6480.txt

[8] Corinne Reichert. 2016. Vocus completes AU$861m acquisition of Nextgen Networks. https://www.zdnet.com/home-and-office/networking/vocus-completes-au861m-acquisition-of-nextgen-networks/

[9] Job Snijders. 2023. RPKIViews. https://www.rpkiviews.org