Prefix2Org : Mapping BGP Prefixes to Organizations

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

Alberto Dainotti
dainotti@gatech.edu
Georgia Institute of Technology
Atlanta, USA

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

Abstract

Accurately mapping Internet address space to organizations is critical to understanding the Internet's organizational ecosystem. Traditional approaches, which rely on individual WHOIS queries often suffer from unclear ownership structure of IP addresses and inconsistent organization names, resulting in ambiguous inferences. Alternative methods that map BGP prefixes to Autonomous Systems Numbers (ASNs) and ASNs to organizations are also inaccurate since ASes often originate prefixes on behalf of their customers. This paper introduces Prefix2Org, a comprehensive prefix-to-organization mapping framework. We introduce a taxonomy for the holders of IP addresses and a methodology to map IP addresses to organizations, based on the operational rights over them. We develop string processing heuristics and leverage RPKI Certificates and routing data to address inconsistencies in organizational names and aggregate prefixes under unified management. Our public dataset covers 99.96% (99.99%) of IPv4 (IPv6) prefixes. We validate 9.3% of routed IPv4 addresses with a 99% recall, and 5.6% of IPv6 prefixes with a 99.34% recall. For the two large organizations where we obtained complete ground truth, Prefix2Org produced no false positives. Finally, in two case studies, *(i)* we characterize organizations that hold address space without an ASN and *(ii)* demonstrate how RPKI adoption measured through Prefix2Org differs from the previously used AS-centric view.

CCS Concepts

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

Keywords

Prefix-to-Organization Mapping, IP Ownership, BGP, RPKI

ACM Reference Format:
Deepak Gouda, Alberto Dainotti, and Cecilia Testart. 2025. Prefix2Org : Mapping BGP Prefixes to Organizations. In *Proceedings of the 2025 ACM Internet Measurement Conference (IMC '25), October 28–31, 2025, Madison, WI, USA.* ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3730567.3764485

1 Introduction

Numerous organizations operate resources and infrastructure on the Internet for their business purposes. Mapping *number resources*, *i.e.,* IP address blocks and Autonomous System Numbers (ASNs), to the organizations that are actually using them for their operation, is crucial to understanding various Internet phenomena. This mapping enables better decision-making to support availability, performance, security, and more generally, the successful operation of services running on the Internet. In addition, it facilitates transparency and accountability, relevant to coordination between organizations, industry groups, and policy-makers.

Efforts to map Internet resources to organizations have mostly focused on ASNs. Indeed, the mapping between ASNs and organizations (AS2Org) has received increased attention [16, 62], and recent efforts have provided the community with improved datasets [13, 18]. In contrast, identifying the organization behind a given IP prefix is a largely unexplored problem. Researchers and operators sometimes use the organization operating the ASN that originates in BGP, the IP prefix they are interested in. But this practice is not necessarily accurate, since many ASNs originate prefixes on behalf of their customers.

More appropriately, we can query the WHOIS databases maintained by the Regional Internet Registries (RIRs), where each address block (sub-)delegation is declared. However, two challenges arise in this process, which we address for the first time in this paper. First, the concept of "ownership" is undefined and rests on a complex reality: *(i)* an address block can be associated with multiple levels of delegation and sub-delegations, each to a different organization; and *(ii)* delegations and sub-delegations come in different "flavors" (22 types across 5 RIRs!), each with a different meaning in terms of rights a delegated organization has on an address block, and taxonomized differently by each RIR. Second, similarly to the parallel AS2Org research problem, in order to be useful, a mapping approach must coalesce the various names under which an organization appears in WHOIS. This necessity arises due to *(i)* the complex ownership structure of resources (*i.e.,* subsidiaries, mergers, and acquisitions over time) and *(ii)* the well-known inconsistencies in naming and in record types across WHOIS databases.

We tackle these two challenges by first defining and proposing a concept of *Direct Ownership* (Section 2.2). Based on our in-depth study of the taxonomies and nomenclatures used by the five RIRs, we define the *Direct Owner* of a prefix the organization with the most authoritative control over the management of the prefix address block. While we focus on this definition of ownership, our approach and the resulting dataset also captures the complete delegation structure for each prefix. This information enables us to identify both intermediate and end-users of an address block, providing a more nuanced understanding of address space utilization.

Building on our definition of Direct Ownership, we then develop a methodology that addresses the second challenge, wherein the same Direct Owner may appear under inconsistent names across different records. Our approach combines data sources clearly associated with this subject, *i.e.,* WHOIS and BGP, with a data source previously unexplored in this context—Resource Public Key Infrastructure (RPKI). Specifically, we leverage the Resource Certificates

ACM ISBN 979-8-4007-1860-1/2025/10
https://doi.org/10.1145/3730567.3764485

in RPKI to identify Internet resources under the same management authority. Through our method, we construct the first dataset that maps prefixes to the organizations holding ownership over them (direct or delegated) while also mapping "sibling" prefixes under the same ownership. Our method supports periodic updates and the generation of historic snapshots, subject to the availability of WHOIS, BGP, and RPKI data.

Our main contributions are:

- We introduce two macro-levels of ownership of prefixes: "Direct Owner" and "Delegated Customer(s)", which we define through their granted rights of address space usage. These two ownership levels allow us to consistently classify prefix ownership across 22 allocation types from all RIRs into two meaningful categories.
- We aggregate prefixes registered by the same Direct Owner under inconsistent names across WHOIS records, using a combination of string processing methods, BGP data, and RPKI resource certificates.
- Utilizing data from September 2024, we build the Prefix2Org dataset, providing a mapping for 99.96% of prefixes in BGP. We validate 9.3% of the routed address space (3.1% of the routed prefixes), and obtain an overall recall score exceeding 99%.
- We conduct two case studies: first, we provide an initial characterization of IP address space held by organizations that do not operate an ASN; second, we demonstrate that *Prefix2Org* enables a more accurate assessment of organizational RPKI adoption compared to previous approaches.

Prefix2Org enables the research, industry, and policy communities to consider organizational ownership of IP space when studying the Internet and Internet services. To the best of our knowledge, this is the first work focusing on mapping prefixes to organizations and consolidating ownership of prefixes by organization at scale.

Roadmap: The rest of the paper is organized as follows: §2 explains what we consider an organization and the concept of prefix ownership, while §3 reviews previous work. §4 explains the data we use and §5 describes the different parts of our methodology. §6 presents the dataset and §7 the validation results. Finally, §8 reports new insights using the dataset, §9 presents limitations and future work, and §10 concludes the paper.

Our data is available at https://doi.org/10.5281/zenodo.17237945 and code to generate the dataset is available at https://github.com/ISS-GT/Prefix2Org-Code.

2 Organizations and IP Delegations

In this work, our focus is to map IP address space (specifically, prefixes announced in BGP) to "owner" organizations. The same block of address space may be linked to different organizations, each holding different rights on the address space. Additionally, organizations often have complex structures and naming conventions. In this section, we discuss what we consider an organization and define two levels of ownership.

2.1 Organizations

An *organization* is an entity incorporated in one or multiple countries to pursue specific purposes, activities, or business objectives.

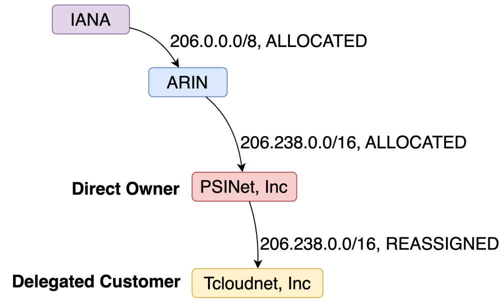

Figure 1: Prefix delegation hierarchy.

To support their operations and infrastructure, organizations frequently require dedicated IP address space. The process of acquiring such resources, known as *IP delegation*, involves the procurement of a contiguous block of IP addresses (a *prefix*) that can be independently routed globally. Depending on their needs and management preferences, organizations may acquire IP address blocks directly from one of the five Regional Internet Registries (RIRs), based on the geographic location of their operations, or indirectly through a National Internet Registry (NIR), a Local Internet Registry or an Internet Service Provider (ISP). In some cases, individuals also obtain Internet resources; for the purposes of this work, we treat such individuals as organizations in their own right.

Organizations evolve over time, and as a consequence, the IP addresses they have may be registered under different names. Multinational organizations may change their legal entity in a country or register in multiple countries under different legal names. An organization can also expand its services and create subsidiaries with similar names. In this work, we consider an organization as the group of all business units and subsidiaries that are related and can be identified by a distinct naming scheme. For instance, we consider *Amazon.com*, *Amazon UK*, and *Amazon Web Services* to be part of the same organization.

2.2 Who owns an address block?

The root of IP address delegation is the Internet Assigned Numbers Authority (IANA). The IANA allocates large pools of IP addresses to RIRs [22]. RIRs then delegate IP address blocks to organizations in their administrative regions. RIRs' policies explicitly state that IP addresses are not "owned as property" [1]. Instead, RIRs delegate the "right to use" IP addresses to organizations, which are said to *hold* IP address blocks. In some cases, holders are allowed to re-delegate their space, either partially or entirely, to other organizations. As a result, for a given IP address block, a chain of sub-delegations may occur after the initial delegation. We refer to both the delegating organization and the recipient(s) of sub-delegation(s) as *holders* of the address space. Importantly, some (sub-)delegations may grant restricted rights, *e.g.*, whether further sub-delegation is permitted or the independence to choose an upstream provider is granted. In each RIR administrative region, the possible terms and restrictions of delegations are taxonomized into different IP allocation types with inconsistent nomenclatures across RIRs—a challenge we address in this section.

Figure 1 shows an example of prefix delegation from the root. IANA allocates `206.0.0.0/8` to ARIN, the RIR for North America. ARIN delegates `206.238.0.0/16` to *PSINet, Inc* granting them several rights of use of the prefix. Then, *PSINet, Inc* re-delegates the full address block to *Tcloudnet, Inc* for their use as a *re-assignment*. In the ARIN zone, that type of address space re-delegation means that the re-delegating party (*PSINet, Inc* in this case) requires the customer party (*Tcloudnet, Inc*) to get Internet connectivity through them (*PSINet, Inc*), without allowing further re-delegation.

Traditionally, the various RIR allocation types primarily capture two rights. First, the right to further sub-delegate the address space: they distinguish whether the address space is given to an *end-user* organization for its own business and operations (an *Assignment*), or to an Internet Service Provider (ISP) that acts as a Local Internet Registry (LIR), subsequently delegating IP blocks and usually providing connectivity to end-user organizations (an *Allocation*)[1]. Second, the right to change upstream provider: they indicate whether there are connectivity requirements related to the upstream provider, *i.e.,* whether the holder may freely decide how to connect to the Internet and route the address space. RIRs use different keywords to signal this second aspect, including *Portable, Non-Portable, Provider Independent (PI), Provider Aggregatable (PA).*

A third key right that we identify and consider to define ownership is the ability to issue RPKI certificates for a prefix. With the emergence of routing security measures and more dynamic IP allocation and usage, the ability to issue cryptographic certificates under RPKI has become relevant. For instance, this ability allows the holder to issue cryptographic certificates that assert authority over the routing characteristics of the address space (*e.g.,* Route Origin Authorizations (ROAs)). Not all types of address space delegations enable the holder organization to issue those cryptographic certificates, even if the organization wants to use them.

In summary, we identify three fundamental rights associated with the use and management of address space allocated to an organization, as indicated by the various allocation type keywords used across RIRs. These are the rights to:

(R_1) Change upstream provider (provider independence);
(R_2) Further sub-delegate all or portions of the address;
(R_3) Issue RPKI certificates for the address space.

In §5.1 we describe in detail how we identify the presence/absence of these rights in the 22 allocation types differently used by the 5 RIRs (with differences also between IPv4 and IPv6 allocations) and the types used for legacy address space, which was delegated before the creation of the RIRs and thus does not fall under the RIR policies. We utilize these three rights (R_1, R_2, R_3) to define two macro levels of control over IP address space, which we refer to as *Direct Owner* and *Delegated Customer*. We define them as follows:

Direct Owner: These holders of address space have the right to choose any upstream provider for connectivity (R_1), *may* have the right to further delegate portions of the address block to downstream customers (R_2), but can always issue or request RPKI certificates for any part of the address space (R_3). The Direct Owner has administrative control over the address block, as they can formulate the routing policies and exercise security practices for their addresses. Consequently, Direct Owners play a key role in how address space is routed and are responsible for implementing and maintaining operational practices for the address space, including routing security practices.

Delegated Customer: These are holders of sub-delegated address space who receive it from another organization, either the *Direct Owner* of the prefix or an intermediate *Delegated Customer*. Delegated Customers do not have direct administrative autonomy to change upstream providers (R_1)—though Direct Owners may decide to provide it, usually cannot further fragment the address block and re-allocate it to other customers unless agreed by the Direct Owner (R_2), and cannot issue RPKI certificates (R_3) unless the Direct Owner provides the infrastructure or issues the certificate for them. Delegated Customers use their address blocks to receive traffic, which can be directly originated by the Direct Owner Autonomous System (AS) or their own AS but should have the Direct Owner as upstream provider unless authorized by the Direct Owner. Essentially, what Delegated Customers can do with the address space is limited by the terms of the contract with the Direct Owner, and they must return the address space upon contract termination.

In this work, to map prefix ownership, for every routed prefix we identify the *Direct Owner* and (the chain) of *Delegated Customer(s)* when there is one. Finally, with the goal of better identifying the organizations that hold the main control of address space, we aggregate Direct Owners when we infer they are the same organization—*i.e.,* in this case, we cluster under the same Direct Owner organization prefixes associated with different organization name strings in WHOIS records. In §5 we describe all the steps to build the Prefix2Org dataset. In particular, in §5.1 we describe how we infer address space rights and then map them to Direct Owner and Delegated Customer(s).

In the context of leased IP addresses—*i.e.,* usage agreements of IP address blocks obtained through transactions involving an IP leasing or brokerage entity independent of connectivity services—Prefix2Org does not explicitly distinguish whether the Direct Owner decided to offer parts of its address space for IP leasing, becoming a lessor, or whether the Delegated Customer is the lessee, getting the same right as noted by the IP allocation type through an IP leasing transaction and contract. Depending on the IP leasing contract, the lessee may appear as a Delegated Customer in Prefix2Org, as there is a WHOIS record for the leased prefix and the lessee holds the right of the IP allocation type. In addition, leasing entities may appear as Direct Owner when they have direct allocations from RIRs or as (intermediate) Delegated Customers when the contract with the lessors provides them the same rights as IP allocation types, some of which can then be further sub-delegated. Further details on how leased IPs are represented in Prefix2Org are provided in Appendix E.

3 Previous Work

Network operators use tools like WHOIS to obtain information about the organizational ownership of prefixes. Alternatively, the research community has also used the mapping of prefixes to ASes and of ASes to organizations, which are sometimes combined to link prefixes to organizations [55]. Finally, there are commercial

[1]Address space provided to a National Internet Registry (NIR) is generally not considered an allocation until the NIR delegates it. We discuss details about NIRs in Section 5.1.

tools that also provide a mapping between prefixes and organizations. This section reviews these approaches and discusses their relationship to our work.

The WHOIS protocol and datasets, maintained by RIRs and NIRs, are the *de facto* tools to find organizations associated with an IP address. However, using WHOIS data has two challenges. First, WHOIS responses are inconsistent across RIRs and NIRs [33, 51], with organization names embedded in various fields such as *org*, *owner*, or *descr*. For example, RIPE records often include an organization ID that requires additional queries to resolve the organization name. Second, WHOIS records across the five RIRs contain more than 20 different keywords indicating the type of allocation for an address block, and lack consistency in the rights granted to address block holders. As noted earlier, multiple organizations can be the "holder" of an address block, each with different rights. This inconsistency complicates the identification of organizational responsibility for a given prefix. In our work, we address these issues by introducing a mapping from the various allocation types to two categories: organizations with "direct" or "delegated" ownership of an address block.

An alternative to WHOIS-based approaches is the use of ASN-to-organization mapping datasets. Since 2005, CAIDA has published daily files mapping IP prefixes to their origin ASN in BGP visible from the Route Views collectors [15]. Additionally, since 2004, CAIDA has also published the Inferred AS to Organizations Mapping Dataset [16], largely based on direct inference from WHOIS data using the methodology from Cai et al. [14]. Recent studies by Chen et al. and Arturi et al. [13, 18] have improved the inferences of CAIDA's AS2Org dataset by incorporating operator-maintained datasets. Previous work has combined BGP data with AS2Org to map prefixes to organizations [55]. However, if used for mapping prefixes to organizations, this approach will misattribute prefixes and will return only organizations that own ASes. Many organizations instead hold address space but do not operate an ASN, relying on a provider ASN for routing—an issue also noted in prior studies [55]. In contrast, our work presents a methodology for directly mapping prefixes to organizations.

Commercial databases such as IPinfo [19, 24, 38] also provide mappings from IP addresses to organizations. However, their methodologies are undisclosed. These services aim for high granularity, often mapping individual IP addresses when possible. IPinfo, for example, requires organizations to have an active domain name; otherwise, it traverses the delegation hierarchy to find an organization with a resolvable domain. In our work, we focus on prefix-level granularity as advertised in BGP, targeting the organization responsible for routing and administrative management of the IP block, and do not require organizations to have an active DNS domain. Moreover, commercial datasets typically lack a concept of "ownership" and do not distinguish between allocation types.

4 Datasets

In this section, we describe the data sources we use to create the Prefix2Org dataset.

4.1 BGP

In this work, we focus on mapping routed prefixes to organizations. Thus, we extract all IPv4 and IPv6 prefixes and the origin ASNs visible in BGP data on September 1, 2024, from all RouteViews and RIPE RIS collectors [36, 39] using BGPStream [40]. We drop prefixes less specific than /8 for IPv4 and /16 for IPv6, since no such IP delegations have been made by RIRs.[2] Our prefix list consists of 1.16M IPv4 and 260k IPv6 prefixes originated from 84.3k ASes.

4.2 WHOIS

WHOIS Database records are the primary source of registration information on Internet resources. We collect the September 2024 snapshots of Bulk WHOIS data from the five RIRs—AFRINIC, APNIC, ARIN, LACNIC, RIPE—and three NIRs—JPNIC, KRNIC, TWNIC.

We parse several types of WHOIS records to extract information about IPv4 and IPv6 prefixes (*e.g.,* `inetnum`, `inet6num` and `NetRange` objects) and the member organizations (`org` objects) that hold these addresses. We create custom parsers for each WHOIS database as RIRs use different record types, field names, and values. For example, APNIC and AFRINIC WHOIS records contain organization information in the `descr` (description) field. RIPE records have an `OrgID` field that identifies an organization record which in turn provides the organization name. When we find several records for the same prefix and allocation type, we use the latest WHOIS records by checking the `last-updated` or an equivalent field.

Additionally, we also extract the prefix allocation type information found in the `status`/`NetType` field of the `inetnum`/`NetRange` records. The Bulk WHOIS data provided by JPNIC does not include the allocation type for prefixes. However, this information can be obtained through individual queries to the JPNIC WHOIS database. Therefore, we perform individual WHOIS queries for each JPNIC address block to retrieve the allocation type field. We can extract the prefix, its owner information (via organization name or ID) and the allocation type information for almost all prefixes in our list from BGP data (99.96% and 99.99% of routed IPv4 and IPv6 prefixes).

4.3 RPKI Certificates

The Resource Public Key Infrastructure (RPKI) is an off-band security framework designed to cryptographically verify the association between Internet resources (IP addresses and ASNs) and their legitimate holders. It enables the holders of Internet resources to issue cryptographic certificates, such as Route Origin Authorization (ROA) certificates, that can then be used by networks to validate routing information.

Each of the five RIRs serves as the trust anchor in RPKI for the resources it delegates, establishing the root of trust at the top of the RPKI certificate tree. When an RIR delegates address space to an organization, it issues a *Resource Certificate* (RC) that attests to the organization's right to use those resources [2, 7, 9, 30, 37]. These Resource Certificates enable the holders to issue further RPKI certificates including: *(i)* child Resource Certificates to provide another organization with the ability to issue certificates for sub-delegated resources, or *(ii)* ROAs, which authorize a specific ASN to originate in BGP the prefixes in the certificate. Prefixes listed in the same Resource Certificate require the same cryptographic key to issue further certificates. As such, the list of IP prefixes

[2]We check RIR delegation files—the daily file with the status of all resources managed by each RIR—and verify that there is no larger delegation than /8 and /16 for IPv4 and IPv6, respectively.

in a Resource Certificate provides a mechanism for identifying a common level of management within the RIR system.

In this work, we leverage Resource Certificates to check if group of prefixes share a common level of management. From the RPKIviews archive [54], we download all RPKI objects of one snapshot. We use the first snapshot for September 1, 2024, which contains all the fetched RPKI certificates from all five RIRs and linked repositories.

4.4 AS2Org and Sibling Inference Datasets

The CAIDA AS2Org dataset [16] maps ASNs to the organizations that own and operate the ASNs. Organizations may own multiple ASNs, which are called *sibling ASNs*. To group together ASes that are controlled by the same organization, we use sibling inferences from *as2org+* by Arturi et al. [13] and from IIL-AS2Org by Chen et al. [18]. We define an *ASN Cluster* as a set of ASNs we find are owned by the same organization. We also use CAIDA AS2Org to fetch the organization names of ASes for the case studies in section 8.

5 Methodology

In this section, we describe our methodology to map prefixes to organizations. Figure 2 shows the pipeline starting from data sources and all the steps to generate the final *Prefix2Org* mapping.

5.1 Mapping Allocation Types to Direct Owners and Delegated Customers

The many types of IP allocation used by RIRs indicate varying rights of usage of the address space that can be granted to a holder organization. While there are numerous nuanced differences among the various allocation types, even within each RIR, the finer distinctions are less relevant for our primary objective of mapping Internet resources to organizations with substantial management and usage rights. Instead, we focus on the three key operational rights described in §2.2 that meaningfully differentiate allocation types: provider independence (R_1), the ability to sub-delegate (R_2), and the authority to issue RPKI certificates (R_3). We identify the rights R_1, R_2, and R_3 for all allocation types and map them to the two categories of ownership—*Direct Owner* and *Delegated Customer*. In the next paragraphs, we explain each right and how it relates to the main allocation types. We also describe how this works for address space delegated by National Internet Registries (NIRs) and legacy address space. In Appendix B.1, we provide an extensive explanation of the mapping for these three rights, handling legacy address space, and the role of National Internet Registries (NIRs) along with the mappings for each allocation type in Tables 8 to 12.

Provider Independence (R_1): Not all IP address allocation types allow holders to change their Internet provider without returning their address block. Allocation types with keywords such as *Portable* or *Provider Independent (PI)* allow holders to change their upstream provider, ensuring provider independence. In contrast, *Provider Aggregatable (PA)* or *Non-Portable* allocations require the holder to use the delegating entity as their upstream provider, as these addresses are typically aggregated for routing purposes.

Sub-delegation of address space (R_2): Not all IP address allocations grant the holder the authority to further delegate, either partially or fully, the address space to other entities such as customer organizations. To determine which allocation types allow sub-delegation, we reviewed RIR IP allocation policy manuals and then complemented this with a data-driven approach: we constructed prefix trees from WHOIS records to examine which allocation types are associated with further re-delegations. Through this process, we found that allocations labeled *Assign* typically do not permit sub-delegation, except in rare cases with LACNIC, where directly assigned blocks can sometimes be reassigned.

RPKI certificate issuance (R_3): RPKI certificates enable the holder of an IP address space to manage its routing behavior in BGP, such as authorizing only specific ASNs to originate the (sub-)prefix. However, not all IP address allocations permit the holder to issue RPKI certificates. Through analysis of RIR policies [3, 8, 10, 31, 46], RPKI Certification Practice Statements [2, 7, 9, 30, 37], certificate chains (see §4.3), and direct communication with RIRs, we find that only organizations with direct IP delegations from RIRs have the authority to issue RPKI certificates themselves. Organizations holding sub-delegations can issue RPKI certificates only if the direct address holder facilitates the process.

National Internet Registries: We examine the role of National Internet Registries (NIRs) in APNIC and LACNIC regions regarding the delegation of IP address space, and how they grant the rights R_1, R_2, and R_3 to their member organizations. Our findings indicate that NIRs act as intermediaries that delegate Internet resources using the same allocation types and rights R_1 and R_2 as their parent RIRs. By analyzing NIR documentation, websites, and certificate chains, we discover that although the technical system may vary across NIRs, all organizations with direct delegations from the NIRs can issue RPKI certificates just as those with direct delegations from RIRs (R_3). In conclusion, direct delegations from NIRs have the same rights as those from RIRs.

Legacy Address Space: About 30% of the routed IPv4 address space, primarily in the ARIN and RIPE regions, falls outside the scope of current RIR policies, as these addresses were allocated before the establishment of the RIRs. This address space is called *legacy* address space [44]. Legacy address space records are maintained in WHOIS databases even if the address holders are not members of the relevant RIRs. All RIRs use the regular allocation and sub-allocation types for the legacy address space except RIPE, which explicitly labels these allocations as `Legacy`. In RIPE, holders of legacy address space can further sub-delegate their allocations, and the sub-delegations retain the `Legacy` designation. Even if there are WHOIS records for legacy address space, legacy address holders cannot directly issue RPKI certificates. Nonetheless, legacy holders may sign agreements with RIRs enabling them to issue RPKI certificates for their address space [6, 49].

In summary, we are able to identify the rights R_1, R_2 and R_3 of all allocation types and map them to two macrocategories of ownership, reflecting the level of control over IP address space: *Direct Owner*, the organization that has the maximum level of control of the address space, receiving a delegation granting them provider independence (R_1), usually the ability to further sub-delegate the address space (R_2), and the authority to issue RPKI certificates (R_3); and *Delegated Customer(s)*, the organizations that are usually the end-users of the address space and hold some of the rights of the address space. A Delegated Customer' rights depend on the terms

RIR	Direct Owner	Delegated Customer
ARIN	Allocation	Reallocation Reassignment
LACNIC	Allocated Assigned	Reallocated Reassigned
RIPE	Allocated PA Assigned PI Legacy‡ Allocated-BY-RIR† Assigned Anycast Allocated-Assigned PA	Assigned PA Assigned† Sub-Allocated PA Allocated-By-LIR† Aggregated-By-LIR†
AFRINIC	Allocated PA Assigned PI Allocated-BY-RIR† Assigned Anycast	Assigned PA Sub-Allocated PA
APNIC	Allocated Portable Assigned Portable	Allocated Non-Portable Assigned Non-Portable

‡ : IPv4 allocations only † : IPv6 allocations only

Table 1: Allocation type values used across five RIRs.

of the contractual relationship with the Direct Owner and, if any, intermediate Delegated Customers.[3]

Mapping a prefix to its Direct Owner and Delegated Customers allows us to treat inconsistent allocation types that grant the same set of operational rights equivalently, thereby simplifying the mapping process while focusing on the key distinction of levels of control between delegations and further sub-delegations. Table 1 provides the mapping of 22 allocation types used across all RIRs into the categories of Direct Owner and Delegated Customer.

5.2 Finding Direct Owners and Delegated Customers of Prefixes

Our objective is to identify the Direct Owner and Delegated Customer for routed prefixes found in BGP data. To do so, we first build IP delegation trees based on WHOIS records and then search in the tree for all the covering prefixes to establish the delegation chain of a prefix.

Building IP Delegation Tree: To preserve the hierarchical nature of prefix delegations, we construct radix prefix trees with IP address blocks obtained from WHOIS records, one for IPv4 and one for IPv6. We parse WHOIS records to extract IP address prefixes, organization name, and allocation type, as described in §4.2. In cases where two WHOIS records exist for the same prefix but with different allocation types, as in the Figure 1 example, we identify the Direct Owner and the Delegated Customer using the mapping from allocation type to ownership level (see Table 1).

Finding Direct Owner and Delegated Customer: For every routed prefix, we query the ownership tree to retrieve the WHOIS record corresponding to the most specific address block that matches or contains the prefix and extract the allocation type. If the allocation type indicates a Direct Owner, we designate the corresponding organization as both the Direct Owner and Delegated Customer, since no further sub-delegations have been made. Conversely, if the allocation type corresponds to a Delegated Customer, we designate the organization as the Delegated Customer. If there are multiple records for the same prefix with different allocation types but all of Delegated Customers, we identify the hierarchy of allocations using the allocation types (*e.g.,* a chain of *Allocation* to *Reallocation* to *Reassignment* in ARIN) and report the list in hierarchical order. Then, we move up the ownership tree to potential intermediate delegations that may still be Delegated Customers, until we reach the Direct Owner delegation. We assign the organization of the Direct Owner delegation as the Direct Owner of the BGP prefix and save all the information of intermediate prefixes. Thus, for every prefix in our BGP dataset, we extract *(i)* the Direct Owner, and *(ii)* the (most specific) Delegated Customer, and *(iii)* all potential intermediate Delegated Customers. We identify Direct Owners and Delegated Customers (if any) for 99.97% of routed prefixes.

[3]less than 3% of prefixes have more than one Delegated Customer.

5.3 Aggregating Prefixes from the same Organization

Many organizations register their direct IP delegations under different names reflecting related legal entities or subsidiaries, *e.g., Google LLC* and *Google Cloud*. Therefore, to provide a better picture of the address space under the control of the same organization, we design a methodology based on string processing techniques, BGP routing data and RPKI certificates to aggregate prefixes whose Direct Owners are the same organization but have different variations of the name.

Aggregating prefixes by the same Direct Owner provides a global view of IP address control. Direct Owners are at the highest level of management of address space and have the largest influence in deciding prefix sizes, connectivity, and whether or not to enable RPKI for their own and their Delegated Customers' usage. In contrast, Delegated Customers have to return their address space if they do not agree with the contractual terms of the address space they receive.

Global aggregation challenge: At a global level, different organizations may have a very similar name, and at the same time, the same company may use different variations of its name for IP resources used in different geographic regions. For instance, the *Telefonica* group from Spain has many subsidiaries in Latin America and has IP addresses under *CTC. CORP S.A. (Telefonica Empresas)*, *Telefonica Chile SA*, *Telefonica del Peru S.A.A.*, and *Telefonica Moviles del Uruguay SA*. However, *Telefonica Celular de Bolivia S.A.* and *Telefonica del Sur S.A.* are two other non-related companies that operate in the telecommunication market of that region. Thus, having a similar name is not enough to identify which IP addresses are controlled by the same organization.

IP addresses sharing operation and management: Even though we cannot simply use similar names to identify IP addresses from the same organization globally, when IP addresses with similar names exhibit shared operations (*i.e.,* are originated by the same ASN in BGP), or management (*i.e.,* they are in the same Resource Certificate), in all likelihood the IP addresses are from the same organization and registered under variations of the name. Prefix2Org leverages this key intuition to aggregate prefixes whose Direct Owners are from the same organization.

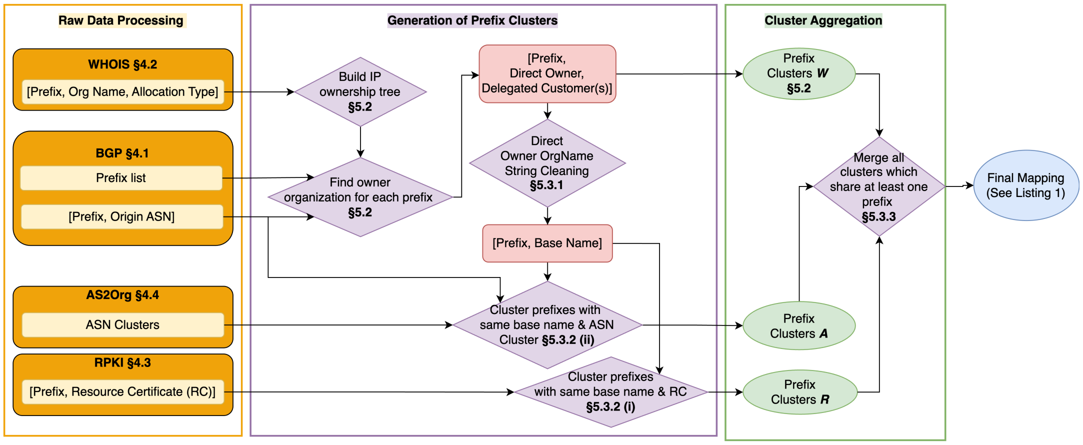

Figure 2: Data sources and methodology to generate Prefix2Org.

Prefix2Org Dataset Pipeline: Figure 2 provides a diagram of our pipeline. It starts with four sources of data (the leftmost column in Figure 2): *(i)* WHOIS IP addresses delegation records (§4.2) which allow us to identify Direct Owners and Delegated Customers (§5.2), *(ii)* BGP data for the list of routed prefixes and their origin ASN (§4.1), *(iii)* AS2Org and AS sibling datasets to create ASN clusters (§4.4), and *(iv)* RPKI resource certificates with the list of IPv4 and IPv6 prefixes in certificates (§4.3). Then, we group prefixes by their Direct Owner name to create the first set of prefix clusters $\mathcal{W}$ (§5.3.1). We refer to these clusters as the *Default Clusters*, as they rely solely on exact string matching after basic string processing[4] of the Direct Owner names. Next, following a heuristics-based string cleaning process (§5.3.1), we extract a *base name* from all Direct Owner names. Using the base name, we group the prefixes *(i)* originated by ASNs from the same *ASN cluster* (§5.3.2), creating the $\mathcal{A}$ clusters; and *(ii)* listed in the same RPKI Resource Certificate (§5.3.2), creating the $\mathcal{R}$ clusters. Finally, we merge the clusters sharing membership (§5.3.3) to create the final mapping.

The rule-based approach we iteratively designed specifically to extract *base names* from WHOIS records performed better than the many approximate string-matching algorithms we tried. We tested fuzzy string matching algorithms such as *thefuzz* [52], similarity scores between two strings based on the Levenshtein distance, and we explored entity resolution and de-duplication algorithms, such as Dedupe [21]. However, we find that they all yielded suboptimal results. The character-level approaches lack semantic understanding of the words being compared and assign a low similarity score to similar strings known to signify the same (*e.g., Telecom* and *Telecommunications*). Even the entity resolution approach is too general to capture the inherent structure of the naming schemes found in WHOIS records. As a result, the rule-based matching algorithm we iteratively designed, leveraging common patterns in the naming conventions found in WHOIS data, identifies more name variations used by the same organization. Below, we dive into the details of each step of the methodology after extracting the Direct Owner of routed prefixes.

[4]We transform the strings to lower case and remove extra spaces.

5.3.1 WHOIS Organization Name String Cleaning:

To identify all Direct Owner names used by an organization, we extract the *base name* from the exact WHOIS names of Direct Owners. We use a rule-based string-cleaning approach designed to handle name variations in WHOIS data. Our approach consists of four steps developed iteratively through extensive analysis of the WHOIS text corpus and manual tuning to enhance robustness and accuracy.

(i) *Initial cleaning and formatting:* We convert organization name strings to a uniform case and use regular expressions to clean special characters, punctuation marks, incorrect encoding, and noisy information such as street addresses and generic remarks (*e.g.,* "IP pool reserved for").

(ii) *Standardizing spelling:* We create a list of common typos and spelling variations (*e.g.,* "Center" vs "Centre", "Telecomunication" vs "Telecommunication") and standardize these terms in the names.

(iii) *Removing frequent words:* We remove legal entity endings and frequent words from organization names when they do not appear as the first word of the organization name string. We compile a list of legal entity endings (*e.g., LLC* and *Inc*) by scraping the Wikipedia page listing legal entity types by country [59]. We compute word frequencies across the WHOIS corpus and remove any word from the organization name string exceeding 100 occurrences[5]. This includes common words such as *the*, *data*, and *customers*.

(iv) *Filtering geographic information:* We remove geographical names when they do not appear as the first word of the organization name string. We utilize the ISO-3166 list [26] of 249 countries and scrape the Wikipedia list of cities with over one million inhabitants [60]. We manually add popular endonyms and country names in different languages such as *Deutschland* for Germany.

Finally, given short names provide insufficient information and can lead to false associations, if the processed name generated after the 4 steps contains fewer than three characters, we revert to the name derived after removing the legal entity ending in step *(iii)*. In Table 2 we list the number of unique organization names resulting after each step of the progressive string cleaning process.

[5]Varying the threshold between 50 and 200 words did not yield significant changes in the final results. We picked a threshold of 100 based on manual inspection of the list.

	# unique names
Original	81,425
Basic Cleaning	81,184
Regex drop	80,707
Corporate words drop	79,783
Frequent words drop	70,963
Geographic words drop	69,065
Refilling words with length <= 3	71,398

Table 2: Number of organization names after each step of the string cleaning process. The process resulted in a 12% reduction in the number of Direct Owner organization names found in the WHOIS records.

Through this process, we end up with 71,398 unique base names, representing a 12% reduction compared to the original 81,184 WHOIS organization name strings (after basic cleaning, see Table 2). We note that the same base name might be shared by two (or more) distinct organizations. For instance, *Fastly Inc.* is a well-known CDN operating in many countries globally, and *Fastly Network Solution Company* is a Vietnamese company that offers VPS and VM hosting. As shown in Table 3, both companies share the same base name *fastly*. However, using routing and RPKI data, we can see that the prefixes with that base name are originated by different ASNs and are in different RPKI Resource Certificates. In the next paragraphs, we describe how Prefix2Org leverages RPKI and routing data to properly cluster prefixes from the *same* organization.

5.3.2 Clustering prefixes with shared operation or management: We use base names to cluster prefixes that share routing operation or management. Specifically, we group prefixes with the same base name that either have the same origin ASN in BGP or are listed in the same RPKI Resource Certificate.

(i) *Clustering prefixes by RPKI Resource Certificates:* As described in §4.3, prefixes in the same Resource Certificate are managed at some level by the same resource management account. Most Resource Certificates issued by RIRs list the prefixes directly delegated to an organization, but there are notable exceptions. For instance, RIRs may issue a Resource Certificate to a National Internet Registry (NIR) with the list of resources delegated to that NIR[6]. NIRs can then use their Resource Certificates in two main ways: they may issue child-Resource Certificates to their customers, listing the prefixes delegated to them and providing the necessary keys for customers to issue further certificates themselves[7], or they directly issue Route Origin Authorizations (ROAs) on behalf of their customers when requested, without granting customers the ability to issue ROAs independently.[8] Further, in RIPE, legacy address space not linked to a member's account is present in a single Resource Certificate, resulting in that Resource Certificate containing address space from multiple organizations. In addition, RIPE also allows for the use of *sponsoring-orgs*—registered Local Internet Registries, that facilitate non-RIPE members in obtaining RIPE services such as requesting IP addresses and ASNs [48]. We observed that resources assigned to different organizations by the same *sponsoring-org* are often included within the same Resource Certificate in RIPE.

Although there is not always a one-to-one mapping between RPKI Resource Certificates and resources delegated to a single organization, all resources within the same Resource Certificate share a common delegation path. This makes it unlikely that two distinct organizations would have such similar names that their *base name* would be identical. Conversely, if organizations with distinct exact names but the same base name have prefixes in the same Resource Certificate, it is highly likely that these distinct names refer to the same organization. Therefore, we use the prefix lists from Resource Certificates to aggregate prefixes with the same base name, resulting in the set of clusters $\mathcal{R}$, which contains 68,888 clusters.

We note that 88% of the routed IPv4 (and 96.7% of the IPv6) prefixes are found in the RPKI Resource Certificates. The missing prefixes are mostly due to the fact that ARIN does not issue RPKI Resource Certificates for address space whose holder has not signed a service agreement with them (*e.g.,* , legacy holders) or has not opted in to RPKI.

(ii) *Clustering prefixes by Origin AS Siblings:* Prefixes originated by the same ASN share a level of operational decision-making, making them more likely to be related. Importantly, the choice of the origin ASN is orthogonal to IP delegation and management, offering an independent source of information for likely related prefixes. Even if ASes originate prefixes from many customers, it is again unlikely that two customers would have similar organization names, leading to identical base names. In contrast, prefixes originated by the same ASN and sharing the same base name are in all likelihood from the same organization.

Sometimes an organization manages multiple ASNs to route prefixes. In this work, instead of an individual origin ASN, we use *ASN Clusters*—a set of ASNs inferred to be owned by the same organization (see §4.4). By grouping prefixes with the same base name and originating from the same ASN cluster, we obtain the set of clusters $\mathcal{A}$, which contains 136,275 clusters.

5.3.3 Merging Prefix Clusters: RPKI RCs and origin ASN clusters offer two independent sources of information about prefixes likely related. Indeed, prefixes from the same organization may be originated by different ASN clusters, and conversely, prefixes from the same organization may be in different RCs (*e.g.,* prefixes are from different RIR zones or coming from a merger or acquisition not yet unified). Thus, we merge together prefix clusters that share membership in an RPKI cluster or ASN cluster, as shown in Figure 3. More specifically, we merge any prefix clusters $W_i, W_j \in \mathcal{W}$ if there exists a prefix-RPKI group $R_i \in \mathcal{R}$ or a prefix-ASN group $A_i \in \mathcal{A}$ such that W_i and W_j have at least one prefix in R_i or A_i (either $\{R_i \cap W_i \neq \emptyset$ and $R_i \cap W_j \neq \emptyset\}$ or $\{A_i \cap W_i \neq \emptyset$ and $A_i \cap W_j \neq \emptyset\}$). We identify 78,475 aggregated clusters, each comprising prefixes that share the same base name and are listed in the same RPKI Resource Certificate or are originated by the same ASN cluster.

Table 3 presents an excerpt from the Prefix2Org dataset, illustrating how prefixes are progressively grouped into clusters. Prefixes P1-P4 are each directly allocated to distinct entities within the broader Verizon Business organization. The string processing

[6]Eight out of nine NIRs—JPNIC, TWNIC, KRNIC, CNNIC, IDNIC, NIC.br, IRINN, and VNNIC—manage their own resource systems independently. NIC.MX has integrated its system with LACNIC, so its delegations follow LACNIC's procedures.
[7]JPNIC, TWNIC, KRNIC, CNNIC, IDNIC, NIC.br follow this approach.
[8]IRINN, VNNIC take this approach.

No.	Prefix	Direct Owner	Base Name	RPKI Cluster	ASN Cluster	Final Cluster
P1	210.80.198.0/24	Verizon Japan Ltd	verizon	**(verizon,0E:65:A4...)**	(verizon,18692)	verizon-I
P2	2404:e8:100::/40	Verizon Asia Pte Ltd	verizon	**(verizon,0E:65:A4...)**	(verizon,701)	verizon-I
P3	203.193.92.0/24	Verizon Hong Kong Ltd	verizon	**(verizon,0E:65:A4...)**	**(verizon,395753)**	verizon-I
P4	65.196.14.0/24	Verizon Business	verizon	(verizon,29:92:C2...)	**(verizon,395753)**	verizon-I
P5	2a04:4e40:8440::/48	**Fastly, Inc.**	fastly	(fastly,8E:AD:ED...)	(fastly,54113)	fastly-I
P6	172.111.123.0/24	**Fastly, Inc.**	fastly	(fastly,0F:DD:01...)	(fastly,54113)	fastly-I
P7	103.186.154.0/24	Fastly Network Solution	fastly	(fastly,16:7C:3B:...)	(fastly,63739)	fastly-II

Table 3: Aggregation of Verizon and Fastly prefixes in the Prefix2Org Dataset. All four Verizon prefixes are in the same cluster, despite having different Direct Owner names, due to a common base name and a non-empty intersection between the RPKI and ASN prefix clusters. In contrast, Fastly Inc. and Fastly Network Solutions are separate clusters.

```
"63.80.52.0/24": {
    "RIR": "ARIN",
    "Direct Owner (DO)": "Verizon Business",
    "DO Prefix": "63.64.0.0/10"
    "DO Allocation Type": "ALLOCATION",
    "Delegated Customer(s) (DC)": ["Bandwidth.com Inc.",
         "Ceva Inc"],
    "DC Prefix(es)": ["63.80.52.0/24", "63.80.52.0/24"],
    "DC Allocation Type(s)": ["REALLOCATION", "
         REASSIGNMENT"],
    "Base name": "verizon",
    "RPKI Certificate": "29:92:C2:35:B0:89...",
    "Origin ASN Cluster": "701",
    "Final Cluster": "verizon—076541",
}
```

Listing 1: Prefix2Org data for 63.80.52.0/24.

methodology described in §5.3.1 results in the base name "verizon." Because prefixes P1, P2, and P3 share this base name and are covered by prefixes in the same RPKI certificate (but different origin ASN), they are grouped into the same prefix RPKI group, denoted as $\mathcal{R}_i$:= (verizon, 0E:65:A4...). Similarly, prefixes P3 and P4, which also share the base name and originate from the same ASN cluster (but different RPKI Certificate), are grouped into the same prefix ASN group, $\mathcal{A}_i$:= (verizon, 395753). Since prefix P4 is a member of both $\mathcal{R}_i$ and $\mathcal{A}_i$, the cluster aggregation process (Figure 3) ultimately merges prefixes P1-P4 into a single final cluster labeled *verizon-I*. Listing 1 provides all the Prefix2Org data for Verizon prefix 63.80.52.0/24.

Similarly, prefixes P5 and P6 are grouped together in the final cluster because they share the same Direct Owner base name *fastly* and are in the same ASN cluster. However, prefix P7, although its base name is also *fastly*, is not included in this cluster. This exclusion occurs because P7 does not share the same RPKI certificate or the same origin ASN cluster as P5 and P6. It is owned by "Fastly Network Solution", an entity unrelated to "Fastly, Inc."

Figure 3: Aggregation of WHOIS prefix clusters via intersecting RPKI prefix clusters & ASN prefix clusters.

Metric	Count
IPv4 Prefixes	1,166,416
IPv6 Prefixes	260,467
Direct Owners	81,184
Delegated Customers	133,370
Base Names	71,398
Origin ASN	84,310
Prefix RPKI Groups	68,888
Prefix ASN Groups	136,275
Base Cluster	81,184
Base Cluster with RPKI Groups	79,759
Base Cluster with ASN Groups	79,484
Final Cluster	**78,475**
No. of Clusters with multiple org names	1853
% IPv4 prefixes in multi-org-name clusters	21.5%
% IPv6 prefixes in multi-org-name clusters	26.4%
% IPv4 addr space in multi-org-name clusters	36.9%

Table 4: Prefix2Org maps 1.16M IPv4 and 260k IPv6 prefixes to 78.5k organizations. Multi-Org-Name Cluster refers to clusters with multiple Direct Owner names.

6 Results

In this section, we describe the resulting Prefix2Org dataset for September 2024, with key metrics in Table 4. We apply the method outlined in §5.2 to all BGP routed IPv4 and IPv6 prefixes (1.4M prefixes) and use the WHOIS dataset described in §4.2. We find 138k unique organization names, which are either Direct Owners or Delegated Customers of the routed prefixes. We identify 81.2k Direct Owners holding direct address space delegations from RIRs, 79.1k (57% of all organization names) of which are also Delegated Customers and 57.2k organizations (41% of all organization names) that are only Delegated Customers but are not Direct Owners of any prefix.

We find that a larger fraction of the IPv6 prefixes, compared to IPv4, are only delegated to Direct Owners and do not have Delegated Customers, while IPv4 prefixes are heavily re-delegated among organizations. This phenomenon is caused by how IPv4 addresses were delegated [42] and by IPv4 address exhaustion [34, 35]. Specifically, we find respectively 31.7% (370k) and 17% (44k) of IPv4 and IPv6 prefixes, for which the Delegated Customer is not the same organization of the Direct Owner, *i.e.,* the Delegated Customer does not have authoritative control over the prefix.

After identifying the 81.2k Direct Owners of all routed prefixes, we further aggregate them using the methodology described in §5.3.3, obtaining 78.5k *prefix clusters*. This 3.3% aggregation (2,709 organizations) is significant because these organizations are large enterprises with subsidiaries, having a presence in multiple locations. Indeed, these 2,709 organizations, which were absorbed into 1,853 clusters in our dataset, administer 38% of the entire routed IPv4 address space and more than a quarter of all routed IPv6 prefixes.

We now dig into the distinct contributions of each component in our aggregation methodology (§5.3.3 and Figure 3) and examine the effect of our aggregation at a prefix granularity as well. When using only the default clustering method, *i.e.,* mapping prefixes to the exact organization name of the Direct Owner in WHOIS records, we obtain the $\mathcal{W}$ set of clusters. In Prefix2Org, however, we also group prefixes by the *base names* that are either *(i)* present in the same RPKI certificate, thus obtaining the $\mathcal{R}$ prefix clusters, or *(ii)* originated by an ASN in the same Origin ASN Cluster, obtaining the $\mathcal{A}$ prefix clusters. Finally, we take the union of these prefix clusters (Figure 3). Thus, each final cluster can have multiple WHOIS names. Specifically, we obtain 948 $\mathcal{R}$ clusters with more than one WHOIS name, which cover 4.8% IPv4 prefixes (10.5% of the routed IPv4 address space) and 3% IPv6 prefixes. Similarly, we get 1,509 $\mathcal{A}$ clusters with more than one WHOIS name, which cover 16.1% IPv4 prefixes (32% of the routed IPv4 space) and 10.4% IPv6 prefixes. The 4.8% increase in prefix-level aggregation resulting from the IPv4 clustering due to $\mathcal{R}$ clusters complements the 16.1% increase due to $\mathcal{A}$ clusters, resulting in a combined increase of 21.5% in the final dataset.

Top 100 Clusters. Our dataset enables us to more accurately identify the amount of address space held by large organizations, as many of them typically have multiple organization names in WHOIS. The largest 100 IPv4 clusters (by amount of address space) are cumulatively the Direct Owners of 65.3% of the routed IPv4 address space. Similarly, the largest 100 IPv6 clusters cover 75.9% of the routed IPv6 prefixes. *DoD Network Information Center* and *Verizon Business* are the largest organizations in terms of IPv4 space; *Comcast Cable Communications* and *China Education and Research Network* are the largest Direct Owners of IPv6 address space.

Looking at the number of Delegated Customers that the largest 100 IPv4 (IPv6) clusters have, we find that these organizations have 12k (8.2k) delegated customers, with *Charter Communications*, *PSINet Inc*—acquired by *Cogent*—and *Level 3 Parent LLC* being the organizations with highest number of Delegated Customers.

In Figure 4, we plot the cumulative fraction of routed IPv4 address space for the top 100 clusters (by number of addresses) by Prefix2Org (blue line), by the default method of aggregating by WHOIS organization name (red), and by the address space originated by ASes that are siblings according to state of the art AS2Org methodologies (green) [13, 16, 18]. The graph shows that through Prefix2Org, we identify a significantly larger amount of space owned by the largest organizations (blue line), compared to the default method of aggregating by WHOIS names (red line). On the other hand, aggregating prefixes simply by considering the ASes that

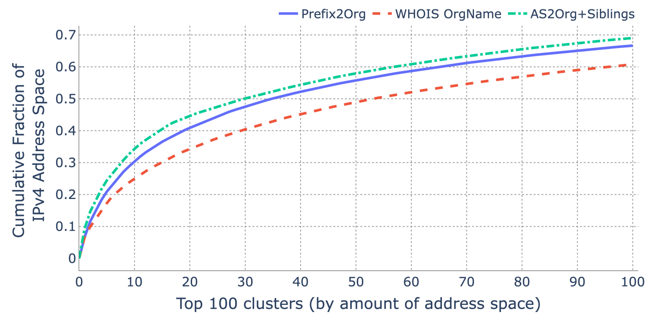

Figure 4: Cumulative fraction of address space covered by top 100 prefix clusters by WHOIS OrgNames vs Prefix2Org vs AS2Org with sibling datasets; Top 100 Prefix2Org clusters cover 6.2% more address space than the clustering by WHOIS organization names.

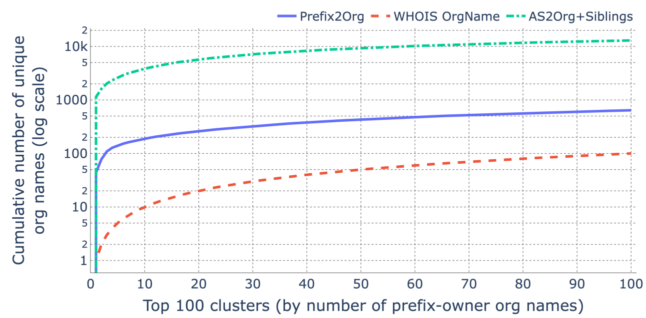

Figure 5: Cumulative number of unique prefix-owner organization names in top 100 clusters created by WHOIS OrgNames vs Prefix2Org vs AS2Org with sibling datasets; WHOIS OrgName clusters contain one name per group while top 100 Prefix2Org clusters encompass more than 600 names.

originate them in BGP and aggregating those ASes by sibling relationship (using state-of-the-art AS2Org approaches—green line), erroneously aggregates prefixes by assigning them to organizations that are not necessarily the Direct Owners.

Similarly, in Figure 5, we plot the cumulative number of WHOIS unique names used by the top 100 clusters. The gap between the blue linegraph (Prefix2Org) and the red one (WHOIS OrgName method) highlights the significant amount of further aggregation resulting from our method. These unique names are fundamental to accurately map organizations' ownership of resources. We also show the equivalent result that would be obtained by (erroneously) aggregating the WHOIS unique names of prefixes that are originated by sibling ASes (green line, using state-of-the-art AS2Org approaches).

7 Validation

Obtaining ground-truth data at scale for a prefix-to-organization mapping is inherently challenging. Thus, in this section, we describe our evaluation of Prefix2Org against ground-truth data collected from various public and private sources:

- **Public IP range lists:** Many organizations, particularly large cloud and hosting service providers, publish lists of their IP address ranges used for customer-facing services. These public lists are often used by their customers to identify egress traffic and

configure firewall rules accordingly. We also find a public report listing IP prefixes held by educational and research institutions affiliated with Internet2 [23], most of which are small organizations. Further details about these institutions and their use in our analysis are discussed in §7.2.
- **Bot lists:** We scrape the list of IP ranges used by bots run by companies that constantly scrape the Internet, such as Google, OpenAI, Anthropic, and Censys.
- **Communication with organizations:** We communicated with organizations to obtain ground truth data or validate our inferences. Specifically, we obtained from Cloudflare and Internet Initiative Japan (IIJ), a large ISP, the list of their prefixes. In our effort to also validate our results for small organizations, we contacted 82 small organizations by email, and five of them replied, as further described in §7.2.

In total, our validation efforts covered 3.1% of the routed IPv4 prefixes, representing 9.3% of the IPv4 address space and 5.5% of the routed IPv6 prefixes.

Companies	True Prefix	Pred Prefix	TP	FN	Precision	Recall
Amazon	8076	8483	7955	121	93.78	98.50
Anthropic	1	1	1	0	100	100
Apple	1351	1400	1351	0	96.50	100
Censys	8	15	8	0	53.33	100
Cloudflare	1477	1873	1477	0	100	100
Comcast	3080	7365	3080	0	41.82	100
Google	3452	3925	3352	100	85.40	97.10
Fastly	299	404	299	0	74.01	100
IIJ	133	143	133	0	100	100
Internet2-affiliates	1422	10869	5502	13	40.58	99.76
Meta	286	391	284	2	72.63	99.30
Microsoft	567	969	565	2	58.31	99.65
Oracle	747	1739	746	1	42.90	99.87
Zscaler	72	536	68	4	12.69	94.44
Email-respondents	5	5	0	0	100	100
Total	**20,976**	**38,118**	**24,826**	**243**	**66.55**	**99.03**

Table 5: Validation of IPv4 prefixes with IP collected range information (TP: True Positive, FN: False Negative). Overall recall score is 99%. The lower precision is due to incomplete ground truth data in the public IP range lists.

Companies	True Prefix	Pred Prefix	TP	FN	Precision	Recall
Amazon	4007	4051	3962	45	97.80	98.88
Anthropic	1	1	1	0	100	100
Censys	4	8	4	0	50	100
Cloudflare	1638	1939	1638	0	84.48	100
Comcast	18	4393	18	0	0.41	100
Google	211	519	211	0	40.66	100
Fastly	642	761	642	0	84.36	100
Facebook	314	407	314	0	77.15	100
IIJ	3	7	3	0	42.86	100
Internet2 affiliates	172	1596	256	4	31.26	98.46
Microsoft	31	72	31	0	43.06	100
Total	**7041**	**13754**	**7080**	**49**	**54.56**	**99.31**

Table 6: Validation of IPv6 Prefixes with collected IP range information (TP: True Positive, FN: False Negative). Overall recall score is 99.3%.

7.1 Validation using IP lists

While public IP range lists provide a useful reference for partial validation, they are typically non-exhaustive, *i.e.,* they sometimes omit addresses reserved for internal use or other services, which does not allow for ascertaining false positives. Our evaluation thus primarily focuses on recall scores, measuring the percentage of IP ranges correctly attributed to their corresponding organizations by Prefix2Org. For Cloudflare and IIJ, we are also able to evaluate false positives and thus precision, as we have received exhaustive lists from them privately. We find that Prefix2Org achieves a 100% precision score for these two organizations. Table 5 shows details for the IP ranges we collected from 14 enterprise organizations (where the row of Internet2 is actually referring to its affiliated organizations, most of which are small) to use as ground-truth data for validation.

Since Prefix2Org focuses only on BGP-routed prefixes, we exclude any prefixes from these datasets that are not present in the BGP routing tables. We then query Prefix2Org and extract the set of prefixes attributed to these organizations (referred to as *Predicted Prefixes*). Subsequently, we calculate the following metrics: the set of predicted prefixes correctly identified as true prefixes (True Positives), the set of true prefixes not attributed to the organization by Prefix2Org (False Negatives), and the corresponding Recall score.

Table 5 and Table 6 summarize the performance of Prefix2Org in attributing IPv4 and IPv6 prefixes to organizations respectively (refer to Table 13 and Table 14 in Appendix C for false positive results). Seven of the 13 organizations assessed achieved a recall score of 100%, demonstrating that Prefix2Org successfully identified all their true prefixes. The average recall score across all organizations is 98.8%, with 10 companies achieving a recall score of more than 99%. Similarly, we achieve a 99.3% recall for IPv6 prefixes. We investigate the significant variation in precision scores observed across different organizations, from 96.5% for Apple to only 12% for Zscaler. Our analysis reveals that the lower precision scores are primarily due to the information available in public data sources not including *all* IP ranges used by organizations. Many organizations do not utilize all of their IP resources for public-facing services and, as a result, their publicly available IP range lists do not include all their routed prefixes. This omission leads to a higher false positive, thereby resulting in lower precision scores.

Prefix2Org has high recall scores, and for the two large organizations for which we obtained comprehensive ground truth, it generates no false positives. Our manual investigation of false negatives in Prefix2Org revealed that some misattributions arise from complex organizational relationships, such as partnerships (*e.g.,* Amazon's collaboration of infrastructure with other companies to offer services in China) or subsidiaries with different legal names (*e.g.,* Meta's *Edge Network Services Ltd*). We add further details in Appendix D.

7.2 Validation for Small Organizations

Internet2, a service provider for academic, educational institutions in the US, publishes the RPKI Ready Report [23], which lists the IP prefixes of its affiliated organizations that lack RPKI ROA coverage. The report includes 810 organizations, most of which are small, with 64% holding only a single prefix and 98.1% holding fewer than 10 prefixes.

775 of these organizations collectively hold 1,422 IPv4 prefixes. For this set, Prefix2Org achieves a recall of 99.8% (referred as *Internet2-affiliates in Table 5*). For IPv6, 162 organizations hold a total of 172 prefixes, with Prefix2Org attaining an overall recall of 98.5%. The median recall scores among the Internet2-affiliates are 100% for both IPv4 and IPv6 prefixes.

To broaden our validation of small organizations beyond educational institutions in North America, we attempted to communicate with small organizations across all RIRs—those managing only one routed /24 IPv4 prefix—to obtain direct confirmation regarding the IP address under their management. Since abuse email addresses listed in WHOIS records are intended solely for reporting technical incidents, abuse, threats, or legal inquiries, we refrained from using them for outreach. Instead, we limited our search to small organizations that also hold an ASN, as these entities are more likely to maintain an active domain and have an IT team capable of providing authoritative ground truth information. We utlized CAIDA AS2Org dataset [16] to identify organizations with one /24 IPv4 prefix that also manage an ASN. We then utilized the domain information in WHOIS records, PeeringDB, and IPinfo [24, 41] to obtain the corresponding websites. We compiled a list of 27.7k Direct Owners holding only one prefix. Of these, 19.3k also hold an ASN, and we were able to locate public websites for 16.9k organizations. From this group, we randomly sampled 20 prefixes per RIR, resulting in a total of 100 organizations. We manually searched their websites for contact email addresses, prioritizing IT department contacts and examining contact, privacy, and terms and conditions pages. In total, we successfully sent 82 emails and received responses from five organizations, all of which confirmed the accuracy of our results. Although the response rate is low, this approach offers further validation of Prefix2Org's accuracy for small organizations. Overall, the results presented in §7.1 and §7.2 demonstrate the reliability of the Prefix2Org dataset across organizations of various sizes.

8 Case Studies

This section presents two case studies that demonstrate the utility of the Prefix2Org dataset for analyzing organizational behavior in Internet routing. The first case study explores organizations that do not own an ASN, a group often overlooked in traditional BGP-based analyses, and investigates their role in address space utilization and business relationships. The second case study examines the differences in ROA coverage when viewed from an AS-centric and prefix-centric perspective, highlighting how organizational practices and administrative boundaries can impact security measurements.

8.1 Studying organizations without ASes

In this case study, we leverage Prefix2Org to identify organizations that have not been delegated an ASN. Such organizations are typically absent from BGP measurement studies, as they do not appear as origin AS in BGP announcements. However, these organizations are directly responsible for managing and utilizing their address space. Therefore, including them in studies provides a more accurate picture of how address space is used on the Internet.

Origin ASN	Organization Name	Own Prefix ROA %	Origin Prefix ROA %
4515	PCCW IMS Limited	100	20
55933	Cloudie Limited	100	32.52
3301	Telia Company AB	100	37.91
3320	Deutsche Telekom AG	100	38.14
4713	NTT Comms	100	38.3
1221	Telstra Limited	100	39.88
2116	Globalconnect AS	100	41.36
2497	Internet Init. Japan	98.51	43.68
852	Telus Comm.	88.46	42.84
6461	Zayo Bandwidth	100	55.7
18747	IFX Corporation	0	80.4
396356	Latitude SH	0	73.85
37228	KT Rwanda Network	0	59.51
21127	JSC Transtelecom	0	58.89
62240	Clouvider Ltd	32.12	85.61

Table 7: Selected ASNs for which we find a significant disparity in ROA Coverage, when measured using Prefix2Org vs Prefix-to-Origin-ASN and AS2Org.

To identify the organizations that do not hold an ASN, we extract all organization names listed in Prefix2Org and exclude those that are also present in the CAIDA AS2Org dataset [16]. Note that we use CAIDA AS2Org only to extract the organization names corresponding to an ASN. Through this process, we identify 16,801 organizations (21.41% of all organizations in our dataset) that do not own an ASN. These organizations collectively own 8.0% of the routed IPv4 prefixes and 6.75% of the routed IPv6 prefixes.

Several organizations without their own ASN manage a substantial amount of address space. For example, Wireless Data Service Provider Corporation holds 944 IPv4 prefixes (15.2M IPv4 addresses), Aliyun Computing Co., Ltd controls 366 prefixes (13.4M IPv4 addresses), and Cloud Innovation Ltd manages 6017 prefixes (10M IPv4 addresses). Cloud Innovation Ltd, an IP leasing company, owns 6017 IPv4 prefixes that are originated from 362 different autonomous systems. The organizations whose ASes originate the most number of Cloud Innovation prefixes include IT HOSTLINE LTD (AS44559), Clouvider Limited (AS62240), and Tcloudnet (AS399077). Using Prefix2Org, we further determine that IT HOSTLINE LTD and Tcloudnet are Delegated Customers of prefixes where Cloud Innovation is the Direct Owner, originating reassigned prefixes via their own ASes.

In contrast, the 944 prefixes from Wireless Data Service Provider Corporation do not have any Delegated Customers. These prefixes are originated via 19 ASes owned by AT&T (AS20057, AS7287) and Verizon Business (AS6167, AS22394). Thus, Prefix2Org enables us to gain detailed insights into the business relationships and operational practices of organizations, even when they do not directly participate in BGP as origin ASes.

8.2 AS-Centric vs Prefix-Centric view of RPKI Adoption

In this section, we evaluate an organization's ROA coverage using two distinct perspectives: the traditional AS-centric view and the more granular prefix-centric view enabled by the Prefix2Org dataset. The goal of this analysis is to demonstrate how RPKI ROA adoption varies when assessed using standard ASN-to-Organization (AS2Org) data versus Prefix2Org data, which identifies the true organizational holder of the address space.

The analysis of organizations using these AS-centric and Prefix-centric views reveal a significant disparity in measured ROA coverage (see Table 7 for selected examples). For organizations in the CAIDA AS2Org dataset, we compute: *(i)* for the AS-Centric view, the ROA coverage for all prefixes originated by the organization's ASN (*Origin ASN* in Table 7); and *(ii)* for the Prefix-Centric view, the ROA coverage for all prefixes originated by the organization's ASN *and* for which the organization is also the Direct Owner according to Prefix2Org. This difference arises from the common operational practice where organizations, often Internet Service Providers, originate prefixes in BGP on behalf of their customers. The key finding lies in the administrative authority: organizations can issue ROAs for prefixes directly delegated to them, *i.e.,* for which they are Direct Owners, but they lack the authority to create ROAs for prefixes that are customer-owned. Even organizations that proactively adopt RPKI can only secure their own address space, potentially leaving customer prefixes without ROAs, as they lack the administrative control to issue them.

The AS-centric perspective often incorrectly leads to the attribution of low ROA coverage to the originating ASN. In this view, the lack of ROAs in customer prefixes lowers the overall ROA coverage associated with the ASN. In contrast, the prefix-centric analysis provides a more accurate assessment, potentially revealing that the organization has issued ROAs for its own prefixes and that the lack of ROAs is primarily due to customer prefixes that remain unsecured. To validate our findings, we communicated with Internet Initiative Japan (IIJ), who confirmed that IIJ's ROA coverage is indeed close to 100%, thereby supporting the analysis using the Prefix2Org dataset. This distinction highlights the utility of Prefix2Org in revealing organizations that may appear to lag in security adoption from an AS-centric view, but are in fact limited by the boundaries of their administrative authority. These insights are valuable for guiding network operators and security researchers toward more effective efforts to improve internet security practices.

9 Limitations and Future Work

Prefix2Org takes a best effort approach to map prefixes to organization(s) and then group prefixes by Direct Owner organizations. However, even when using RPKI and routing data to detect inconsistencies in WHOIS records, the distributed, dynamic and ever-evolving landscape of IP delegations inherently limit the accuracy of the mapping.

WHOIS Data: Prefix2Org heavily relies on WHOIS data, the canonical source of administrative information for Internet resources. Despite it being a critical source of registration information, WHOIS data is inherently limited by outdated or incomplete records. A significant challenge in linking prefixes to organizations also lies in processing the noisy WHOIS information to extract the base names for organizations. Furthermore, the denoising and string processing techniques employed in this paper are based on domain-specific heuristics developed through an extensive analysis of organization names in WHOIS. Over time, the string processing and denoising might need to be updated and improved.

Mergers and Acquisitions: Corporate entities frequently undergo reorganizations such as mergers, acquisitions, and the creation of subsidiaries. These operations introduce naming complexities that cannot be captured solely through string processing. We will explore leveraging third-party datasets of mergers and acquisitions, as well as Internet measurement datasets such as state-of-the-art AS-to-organization mappings, to uncover more sibling relationships. In addition, with the latest advancements in Large-Language Models (LLMs), which constantly index the Internet, we can develop frameworks to track historical and ongoing acquisitions and mergers.

Third-Party IP Leasing: Prefix2Org can help identify organizations that hold specific IP address blocks and further sub-delegate them, which may aid in detecting addresses involved in the IP leasing market. However, Prefix2Org might not identify all leased address space and leasing can appear in different ways on the dataset. Indeed Prefix2Org does not explicitly distinguish between lessors, leasing entities, and lessees, they are treated as Direct Owners or Delegated Customers based on observable WHOIS delegation records. Du et al. [20] inferred that 4.1% of routed IPv4 prefixes were involved in IP leasing as of April 2024. We leave as future work a thorough study of how that address space appears in Prefix2Org and whether Prefix2Org combined with BGP data could be used to infer IP leasing activity.

10 Conclusion

This paper presents Prefix2Org, the first publicly available prefix-to-organization dataset encompassing almost all prefixes publicly announced in BGP. Using WHOIS records, routing data and RPKI certificates, Prefix2Org maps IP prefixes to organizations based on the operational rights divided in two macro-levels of control over IP address space: Direct Owners and Delegated Customers. Through case studies, we demonstrate how we can identify organizations not well-represented in measurement studies due to their lack of AS ownership. These organizations do not appear in widely used datasets, such as CAIDA AS2Org and PeeringDB, but are represented in Prefix2Org. We also demonstrate how Prefix2Org provides a different (more accurate) view of the Internet in contrast to mapping prefixes to ASes and ASes to organizations.

We plan to periodically update and release our dataset to the public. Periodic snapshots would allow researchers to perform longitudinal analyses and study the dynamics of prefix ownership, such as address transfers, leasing activities, and the evolution of business relationships between organizations.

Acknowledgements

This work is partly supported by the National Science Foundation grant OAC-2419735 and a Google Research gift. 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 or Google.

References

[1] 2012. Managing Internet Number Resources. Retrieved Nov 13, 2024 from https://www.nro.net/wp-content/uploads/Managing-Internet-Number-Resources.pdf

[2] AFRINIC. 2020. *AFRINIC Certification Practice Statement for the Resource PKI.* https://rpki.afrinic.net/policy/CPS.pdf

[3] AFRINIC. 2020. Consolidated Policy Manual (v1.6). Retrieved Nov 13, 2024 from https://afrinic.net/policy/manual

[4] AFRINIC. 2020. *Legacy Resource Holders.* https://afrinic.net/membership/legacy-resource

[5] Amazon. [n. d.]. Amazon Web Services in China. Retrieved Nov 15, 2024 from https://www.amazonaws.cn/en/about-aws/china/

[6] American Registry for Internet Numbers. 2025. Legacy Resources at ARIN. https://www.arin.net/resources/guide/legacy/ Accessed: 2025-09-24.

[7] APNIC. [n. d.]. *APNIC RPKI CPS.* https://www.apnic.net/wp-content/uploads/2021/05/CPS.pdf

[8] APNIC. 2024. APNIC Internet Number Resource Policies. Retrieved Nov 13, 2024 from https://www.apnic.net/community/policy/resources

[9] ARIN. 2024. *CERTIFICATION PRACTICE STATEMENT FOR RESOURCE CERTIFICATION.* https://www.arin.net/resources/manage/rpki/cps.pdf

[10] ARIN. 2024. Number Resource Policy Manual. Retrieved Nov 13, 2024 from https://www.arin.net/participate/policy/nrpm/

[11] ARIN. 2025. *Legacy Resources at ARIN.* https://www.arin.net/resources/guide/legacy/

[12] ARIN. 2025. Resources Under Registration Services Agreement (RSA) Report. Retrieved June 14, 2025 from https://www.arin.net/reference/research/statistics/resources/

[13] Augusto Arturi, Esteban Carisimo, and Fabián E. Bustamante. 2023. *as2org+*: Enriching AS-to-Organization Mappings with PeeringDB. In *Passive and Active Measurement*, Anna Brunstrom, Marcel Flores, and Marco Fiore (Eds.). Springer Nature Switzerland, Cham, 400–428.

[14] 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*. ACM, Melbourne Australia, 199–205. https://doi.org/10.1145/1879141.1879166

[15] CAIDA. [n. d.]. Routeviews Prefix to AS mappings Dataset for IPv4 and IPv6. https://www.caida.org/catalog/datasets/routeviews-prefix2as/.

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

[17] Japan Network Information Center. 2024. *Guidelines for countermeasures against unauthorized routes on the Internet using RPKI ROA.* https://www.nic.ad.jp/doc/jpnic-01324.html#2.1.3

[18] Zhiyi Chen, Zachary S. Bischof, Cecilia Testart, and Alberto Dainotti. 2023. Improving the Inference of Sibling Autonomous Systems. In *Passive and Active Measurement*, Anna Brunstrom, Marcel Flores, and Marco Fiore (Eds.). Springer Nature Switzerland, Cham, 345–372.

[19] DB-IP. 2024. *DB-IP.* https://db-ip.com

[20] 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* (Madrid, Spain) *(IMC '24)*. Association for Computing Machinery, New York, NY, USA, 328–336. https://doi.org/10.1145/3646547.3689010

[21] Derek Eder Forest Gregg. [n. d.]. *Dedupe.io.* https://dedupe.io

[22] ICANN. [n. d.]. *Internet Assigned Numbers Authority (IANA) Policy For Allocation of IPv4 Blocks to Regional Internet Registries.* https://www.icann.org/resources/pages/allocation-ipv4-rirs-2012-02-25-en

[23] Internet2. 2024. *RPKI-Ready Networks with No ROAs.* https://route-reports.ns.internet2.edu/rpki_ready_report.html

[24] IPInfo. 2024. *IPInfo.* https://ipinfo.io

[25] IRINN. [n. d.]. *IRINN: How to raise request for ROA registration.* https://www.irinn.in/resourceCertDo

[26] ISO. 2025. *ISO 3166 Country Codes.* https://www.iso.org/iso-3166-country-codes.html

[27] Tanya Jawab. [n. d.]. *IDNIC: Resource Public Key Infrastructure (RPKI) - FAQ.* https://idnic.id/membership/faq/category/resource-public-key-infrastructure-rpki/OA==

[28] KISA. 2021. *Korea Internet and Security Agency (KISA) Routing Certification (RPKI) Support Guide.* https://íŢœê𝜇ŋìİÿíĎřëĎůìăŢëşťìĎijíĎř.íŢœê𝜇ŋ/jsp/resources/rpki.jsp

[29] LACNIC. [n. d.]. *LACNIC: Resource Certification (RPKI).* https://www.lacnic.net/640/2/lacnic/resource-certification-rpki

[30] LACNIC. 2010. *LACNIC CPS.* https://www.lacnic.net/innovaportal/file/641/1/lacniccpsv1.pdf

[31] LACNIC. 2023. LACNIC POLICY MANUAL v2.19. Retrieved Nov 13, 2024 from https://www.lacnic.net/innovaportal/file/680/1/manual-politicas-en-2-19.pdf

[32] LACNIC. 2025. *Legacy Resources.* https://www.lacnic.net/1022/2/lacnic/legacy-resources

[33] Suqi Liu, Ian Foster, Stefan Savage, Geoffrey M. Voelker, and Lawrence K. Saul. 2015. Who is .com? Learning to Parse WHOIS Records. In *Proceedings of the 2015 Internet Measurement Conference* (Tokyo, Japan) *(IMC '15)*. Association for Computing Machinery, New York, NY, USA, 369–380. https://doi.org/10.1145/2815675.2815693

[34] Ioana Livadariu, Ahmed Elmokashfi, and Amogh Dhamdhere. 2017. On IPv4 transfer markets: Analyzing reported transfers and inferring transfers in the wild. *Computer Communications* 111 (2017), 105–119. https://doi.org/10.1016/j.comcom.2017.07.012

[35] Ioana Livadariu, Ahmed Elmokashfi, Amogh Dhamdhere, and kc claffy. 2013. A first look at IPv4 transfer markets. In *Proceedings of the Ninth ACM Conference on Emerging Networking Experiments and Technologies* (Santa Barbara, California, USA) *(CoNEXT '13)*. Association for Computing Machinery, New York, NY, USA, 7–12. https://doi.org/10.1145/2535372.2535416

[36] RIPE NCC. 2023. *Routing Information System (RIS).* https://www.ripe.net/analyse/internet-measurements/routing-information-service-ris

[37] RIPE NCC. 2024. *RIPE NCC Certification Practice Statement (CPS) for the Resource Public Key Infrastructure (RPKI).* https://www.ripe.net/publications/docs/ripe-824/

[38] NetworksDB. 2024. *NetworksDB.* https://networksdb.io

[39] University of Oregon. 2022. *University of Oregon Route Views Project.* https://www.routeviews.org/routeviews/

[40] Chiara Orsini, Alistair King, Danilo Giordano, Vasileios Giotsas, and Alberto Dainotti. 2016. BGPStream: A Software Framework for Live and Historical BGP Data Analysis. In *Proceedings of the 2016 Internet Measurement Conference* (Santa Monica, California, USA) *(IMC '16)*. Association for Computing Machinery, New York, NY, USA, 429–444. https://doi.org/10.1145/2987443.2987482

[41] PeeringDB. 2023. *PeeringDB.* https://www.peeringdb.com

[42] Philipp Richter, Mark Allman, Randy Bush, and Vern Paxson. 2015. A Primer on IPv4 Scarcity. *SIGCOMM Comput. Commun. Rev.* 45, 2 (April 2015), 21–31. https://doi.org/10.1145/2766330.2766335

[43] RIPE. [n. d.]. *Can I convert my legacy space to Allocated PA or Assigned PI?* https://www.ripe.net/manage-ips-and-asns/legacy-resources/legacy-internet-resources-faqs/can-i-convert-my-legacy-space-to-allocated-pa-or-assigned-pi/

[44] RIPE. 2015. RIPE NCC Services to Legacy Internet Resource Holders. RIPE Document. https://www.ripe.net/publications/docs/ripe-639/ Document number RIPE-639. Accessed on September 24, 2025.

[45] RIPE. 2023. RIPE Resource Transfer Policies. RIPE Document. https://www.ripe.net/publications/docs/ripe-807/

[46] RIPE. 2024. IPv4 Address Allocation and Assignment Policies for the RIPE NCC Service Region. Retrieved Nov 13, 2024 from https://www.ripe.net/publications/docs/ripe-826/

[47] RIPE. 2025. *RIPE NCC Services to Legacy Internet Resource Holders.* https://www.ripe.net/manage-ips-and-asns/legacy-resources/ripe-ncc-services-to-legacy-internet-resource-holders/

[48] RIPE Network Coordination Centre. 2024. Sponsoring LIR Agreement (Independent Assignment Request and Maintenance Agreement for non-Legacy Internet Resources). https://www.ripe.net/manage-ips-and-asns/resource-management/number-resources/independent-resources/independent-assignment-request-and-maintenance-agreement/ Accessed: 2025-09-30.

[49] RIPE Network Coordination Centre. 2025. RIPE NCC Services to Legacy Internet Resource Holders. https://www.ripe.net/manage-ips-and-asns/legacy-resources/ripe-ncc-services-to-legacy-internet-resource-holders/ Accessed: 2025-09-24.

[50] Sanjaya. 2021. *RPKI services now available to APNIC historical resource holders.* https://blog.apnic.net/2021/03/26/rpki-services-now-available-to-apnic-historical-resource-holders/

[51] Jan Schaumann. 2024. Whose CIDR Is It Anyway? Retrieved Nov 13, 2024 from https://labs.ripe.net/author/jschauma/whose-cidr-is-it-anyway/

[52] SeatGeek. [n. d.]. *TheFuzz.*

[53] SEC. 2024. *LIST OF SUBSIDIARIES FACEBOOK, INC.* https://www.sec.gov/Archives/edgar/data/1326801/000132680117000007/fb-12312016x10kexhibit211.htm

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

[55] Florian Streibelt, Martina Lindorfer, Seda Gürses, Carlos H. Gañán, and Tobias Fiebig. 2023. Back-to-the-Future Whois: An IP Address Attribution Service for Working with Historic Datasets. In *Passive and Active Measurement*, Anna Brunstrom, Marcel Flores, and Marco Fiore (Eds.). Springer Nature Switzerland, Cham, 209–226.

[56] John Sweeting. 2023. ARIN and IPv4 Address Leasing. https://www.arin.net/blog/2023/02/23/ipv4-leasing/

[57] VNNIC. [n. d.]. *VNNIC: How to use RPKI to perform authentication in routing operations.* https://academy.vnnic.vn/dam-bao-an-toan-dinh-tuyen-bang-phuong-phap-xac-thuc-tai-nguyen-mang-su-dung-rpki

[58] Time Wang. 2025. *The Deployment Experience on TWNIC ROA ROV.* https://2025.apricot.net/assets/files/APAC945/the-deployment-exper_1740438775.pdf

[59] Wikipedia. 2024. *List of legal entity types by country.* https://en.wikipedia.org/wiki/List_of_legal_entity_types_by_country

[60] Wikipedia. 2025. *List of cities with over one million inhabitants.* https://en.wikipedia.org/wiki/List_of_cities_with_over_one_million_inhabitants

[61] Yuedong Zhang. 2019. *RPKI in China.* https://blog.apnic.net/2019/02/14/rpki-in-china/

[62] Maya Ziv, Liz Izhikevich, Kimberly Ruth, Katherine Izhikevich, and Zakir Durumeric. 2021. ASdb: A System for Classifying Owners of Autonomous Systems. In *Proceedings of the 21st ACM Internet Measurement Conference* (Virtual Event) *(IMC '21)*. Association for Computing Machinery, New York, NY, USA, 703–719. https://doi.org/10.1145/3487552.3487853

A Ethics

This work does not raise any ethical issues.

B Allocation Types and their properties

B.1 From Allocation Types to IP Delegation Rights:

To determine which allocation types grant which of the rights R_1 (provider independence), R_2 (sub-delegation) and R_3 (RPKI certificate issuance), we carefully scrutinize RIR IP delegation policies [3, 8, 10, 31, 46], RIR RPKI Certification Practice Statements (CPS) as well as RPKI and WHOIS records. By integrating top-down (policy to data) and bottom-up (data to policy) approaches, we are able to identify the specific rights associated with each allocation type found in the WHOIS database of each RIR. We also reach out to RIRs to verify our understanding when needed. In Tables 8 to 12, we provide detailed mappings of all allocation types per RIR.

Provider Independence (R_1): Some allocation types allow holders to switch upstream providers without returning the address block, *i.e.,* holders are free to choose how to get connectivity to the Internet. As connectivity has traditionally been a key aspect of IP delegation, this right is well described in RIR IP allocation policy manuals. Allocation types with the keywords *Portable* or *Provider Independent (PI)* indicate that the address space holder can choose any upstream provider for connectivity, thus ensuring Provider Independence. In contrast, allocation types with the keywords *Provider Aggregatable (PA)* or *Non-Portable*, require the holder to use the delegating entity as their upstream provider, as the upstream provider may aggregate its allocations for routing purposes. Provider independence is also always present in all direct IP delegations from RIRs or NIR, whether *Allocations* or *Assignments*.

Sub-delegation of address space (R_2): To identify the allocation types that allow sub-delegation, we start from RIR IP allocation policy manuals and complement the analysis with a data-driven approach: we construct a prefix tree from WHOIS records and examine which allocation types are associated with records that further re-delegate prefixes. This approach enables us to empirically validate our understanding of allocation types that permit address holders to re-delegate all or portions of their address space. We observe that allocations with the keyword *Assign* do not allow further sub-delegation, with one exception—LACNIC directly *Assigned* address blocks can be further *Reassigned*, though the option is only seldom used.

RPKI certificate issuance (R_3): A key property of RPKI is that all prefixes listed in the same Resource Certificate at a given hierarchical level use the same cryptographic key and therefore the same management account in the RIR system for issuing cryptographic certificates. This implies, that these prefixes are under a common administrative authority in the RPKI system, which is valuable for mapping prefixes to organizations. Following the Certificate chain in the RPKIviews snapshot, we extract and use the child-most RC in which a prefix is present to identify prefixes that share an RPKI management account.

We investigate the specific allocation types that grant organizations the ability to issue RPKI certificates (*e.g.,* ROAs). Our analysis of RIR policies [3, 8, 10, 31, 46], RPKI Certification Practice Statements (CPS) [2, 7, 9, 30, 37], and RPKI certificates chains (see §4.3), and then further communication with RIRs reveals that only organizations with direct IP delegations from RIRs can directly issue RPKI certificates for the delegated address space. For instance, organizations holding *Allocation* in ARIN, *Allocated PA* in RIPE and *Assigned Portable* address blocks in APNIC can issue ROAs.

Organizations receiving sub-delegations can only issue RPKI certificates if the direct delegation holder either issues the certificate on their behalf, or provides the infrastructure for delegating and hosting RPKI certificates of sub-delegations. It is not a requirement, and the organization may choose not to do it.

National Internet Registries: In APNIC and LACNIC, some countries have National Internet Registries (NIRs) that act as intermediates and delegate Internet resources in their country. NIRs are well-established registry entities that provide similar services to RIRs within their community. There are 9 NIRs in total, 7 in APNIC (JPNIC for Japan, TWNIC for Taiwan, KRNIC for Korea, CNNIC for China, IRINN for India, IDNIC for Indonesia, and VNNIC for Vietnam) and 2 in LACNIC (NIC.br for Brazil and NIC.mx for Mexico). Importantly, IP delegations from NIRs use the same allocation types as their parent RIR and have equivalent rights with respect to upstream provider choice (R_1) and sub-delegation (R_2), which is not the case of Local Internet Registries (LIRs) as LIR allocations are considered sub-delegations. We investigate whether direct delegations from the NIRs can all issue RPKI certificates just like direct delegations of RIRs. We inspect NIR's website, documentation, and the certificate chain of their resources. We find that all organizations with direct delegations from NIRs can issue RPKI certificates, although the technical system used can vary [17, 25, 27–29, 57, 58, 61]. RIRs issue a Resource Certificate to NIRs with the IP address blocks that the NIR has delegated or can delegate. NIRs use that certificate key to set up the infrastructure for their customer organization to issue their RPKI certificate, or directly sign certificates as requested by their customers, following established procedures. More importantly, all direct delegations from NIRs can issue RPKI certificates at will through an established mechanism. Thus, we infer that direct delegations from NIR have the same rights with respect to RPKI certificates as direct delegations from RIRs, thereby completing our ability to infer RPKI rights for all allocation types.

Legacy Address Space: About 30% of the routable IPv4 address space was delegated before the creation of RIRs, and therefore this *Legacy* address space is outside of regular RIR policies. RIRs provide basic registry services for legacy address space, allowing the registered organization to create WHOIS records, including sub-delegation records. All RIRs use their usual delegation and sub-delegation types for legacy address space, with the exception of RIPE, which has a dedicated allocation type (Legacy) for legacy address space. However, for legacy address space to gain full registry services, including the full management of WHOIS records and the ability to issue RPKI certificates, the holders need to become members or sign an agreement as requested by the RIRs [4, 11, 32, 47, 50]. In RIPE, when the holder of legacy address space becomes a member, the allocation type can be changed to one of the other

	ARIN		
Allocation Type	**Change Upstream (R_1)**	**Further sub-delegations (R_2)**	**Issue ROAs (R_3)**
Allocation	✔	✔	✔
†Allocation-Legacy	✔	✔	✗
Re-Allocation	✗	✔	✗
Reassignment	✗	✗	✗

† Modified Allocation Type in Prefix2Org

Table 8: Allocation Type values used by ARIN. In grey, the allocation types of prefixes of *Direct Owners*.

	LACNIC		
Allocation Type	**Change Upstream (R_1)**	**Further sub-delegations (R_2)**	**Issue ROAs (R_3)**
Allocated	✔	✔	✔
Reallocated	✗	✔	✗
Assigned	✔	✔	✔
Reassigned	✗	✗	✗

Table 9: Allocation Type values used by LACNIC. In grey, the allocation types of prefixes of *Direct Owners*.

	APNIC		
Allocation Type	**Change Upstream (R_1)**	**Further sub-delegations (R_2)**	**Issue ROAs (R_3)**
Allocated Portable	✔	✔	✔
Allocated Non-Portable	✗	✔	✗
Assigned Portable	✔	✗	✔
Assigned Non-Portable	✗	✗	✗

Table 10: Allocation Type values used by APNIC. In grey, the allocation types of prefixes of *Direct Owners*.

	RIPE		
Allocation Type	**Change Upstream (R_1)**	**Further sub-delegations (R_2)**	**Issue ROAs (R_3)**
Allocated PA	✔	✔	✔
Assigned PI	✔	✗	✔
Sub-Allocated PA	✗	✔	✗
Legacy	✔	✔	✔
†Legacy-Not-Sponsored	✔	✔	✗
Allocated-Assigned PA	✔	✗	✔
Assigned Anycast	✔	✗	✔
‡Allocated-By-RIR	✔	✔	✔
‡Allocated-By-LIR	✗	✔	✗
Assigned PA	✗	✗	✗
‡Assigned	✗	✗	✗
‡Aggregated-By-LIR	✗	✔	✗

‡ IPv6 only † Modified Allocation Type in Prefix2Org

Table 11: Allocation Type values in RIPE; PA - Provider Aggregatable, PI - Provider Independent. In grey the allocation types of prefixes of *Direct Owners*.

ones in RIPE [43]. In conclusion, legacy address space appears in RIR WHOIS records even though many legacy holders do not have full registry services, and they are not members of RIRs.

Most legacy address space is in the ARIN and RIPE administrative regions. ARIN has a list of legacy address space holders that have signed the registry service agreement [12]. 16% of prefixes in the ARIN zone, representing 4.4% of all routed IPv4 prefixes, have not signed a service agreement with ARIN. To distinguish prefixes without an agreement and as a consequence without the ability to issue RPKI certificates in ARIN, we mark prefixes utilizing ARIN's Basic Legacy Services as `Legacy-Allocation`. In RIPE, given `legacy` allocations are not necessarily members or sponsored

	AFRINIC		
Allocation Type	**Change Upstream (R_1)**	**Further sub-delegations (R_2)**	**Issue ROAs (R_3)**
Allocated PA	✔	✔	✔
Assigned PI	✔	✗	✔
Sub-Allocated PA	✗	✔	✗
Assigned Anycast	✔	✗	✔
‡Allocated-By-RIR	✔	✔	✔
Assigned PA	✗	✗	✗

‡ IPv6 only

Table 12: Allocation Type values in AFRINIC; PA - Provider Aggregatable, PI - Provider Independent. In grey, the allocation types of prefixes of *Direct Owners.*

by a member account, they may not have full registry service and thus the ability to issue RPKI certificates. However, in the RIPE RPKI certificate chain, the legacy prefixes that are not part of RIPE membership are in a distinct certificate. 4.4% of all routed RIPE IPv4 prefixes have the allocation type *Legacy*, of which 36.4% we infer are not full members of RIPE. To distinguish those cases in RIPE, we mark them as `Legacy-Not-Sponsored`.

B.2 From IP Delegation Rights to Two Levels of Ownership:

Through our study of rights granted by different types of address space, we realized that a key distinction in ownership level is whether the address space was directly delegated by the RIR or NIR, or if it was sub-delegated by an ISP (acting as a Local Internet Registry (LIR)). Holders of address space directly delegated by RIRs/NIRs can decide what rights they provide to further sub-delegation and are the ones with access to full registry services, including the ability to issue RPKI certificates. Also, when the direct delegation is of a type with more restricted rights about sub-delegation, *i.e.,* it is an *assignment* in RIPE or APNIC, the holder can always request the RIR a change of allocation type to the more permissive one without having to return the address space. Therefore, we decide to identify two macro levels of control over IP address space: *Direct Owner* for one for holders of direct delegations, and *Delegated Customer* for holders of sub-delegations.

Direct Owners have direct delegations granting them provider independence (R_1), usually the ability to further sub-delegate the address space (R_2), or they can request it, and the authority to issue RPKI certificates (R_3). Note that in the case of legacy address space, Direct Owners must enter into an agreement with the RIR first, but this is their choice. In contrast, Delegated Customers have sub-delegations that limit their ability to change upstream provider or sub-delegate the address block, depending on the contractual relationship with the Direct Owner, and that require the Direct Owner's support (if provided) to issue an RPKI certificate, *i.e.,* can only have RPKI certificates if the Direct Owner uses RPKI and provides a mechanism for Delegated Customers to have specific RPKI certificates for their delegation. Table 1 provides the mapping of 22 allocation types used across all RIRs into the categories of Direct Owner and Delegated Customer.

C Complete Validation Results

In Table 13 and Table 14 we provide detailed metrics of our validation results, including True Positives, False Positives, and False Negatives. Note that True Positives might exceed the number of True Prefixes if several inferred prefixes are sub-prefixes of a single prefix in the ground truth validation dataset.

D Investigation of False Negatives

We manually reviewed the false negatives in our evaluation and contacted three organizations for clarification, and asked them to provide any information on the False Negatives found in our dataset. For Amazon, we identified 121 prefixes attributed to non-Amazon entities: Ningxia West Cloud Data Technology Co. Ltd. (56 prefixes), Beijing Guanghuan Xinwang Digital Technology Co. Ltd. (56 prefixes), and Beijing Sinnet Technology Co. Ltd. (9 prefixes). Further investigation revealed that Amazon partners with these organizations in China to provide localized cloud services [5]. Thus, Amazon includes these IP addresses in its public IP range list, even though it does not directly hold these blocks.

For Meta Platforms, Prefix2Org attributes two prefixes to Edge Network Services Ltd, an organization different from Meta Platforms, Inc (false negatives). Upon investigation, we found that Edge Network Services Limited is a subsidiary of Meta registered with US SEC [53]. This highlights a scope for improvement in Prefix2Org, as it relies on string processing heuristics and is therefore unable to aggregate the prefixes of organizations with completely different names. We discuss the future scope of improvement in §9. Unfortunately, we were unable to gather information about the false negatives associated with Google and Microsoft.

A limitation of our validation procedure is the lack of a comprehensive and scalable public IP range dataset, which restricted our ability to validate more than a subset of Prefix2Org's results.

E IP leasing and Prefix2Org

IP leasing refers to organizations getting IP address usage agreements independent of connectivity services. This type of IP usage was not considered when RIRs established their IP delegation policies, and currently, only RIPE and ARIN permit it. RIPE allows temporary transfers where the IP addresses are expected to be returned to the original resource holder, which enables IP leasing [45]. RIPE also requires registration in its WHOIS database of any IP reassignment or reallocation, including those that occur under a leasing agreement. Similarly, ARIN allows organizations to lease their excess allocated IPv4 space, provided they record reassignments and reallocations in the ARIN WHOIS database. However,

Companies	True Prefixes	Predicted Prefixes	True Positive	False Positive	False Negative	Precision	Recall
Amazon	8076	8483	7955	528	121	93.78	98.50
Anthropic	1	1	1	0	0	100.00	100.00
Apple	1351	1400	1351	49	0	96.50	100.00
Censys	8	15	8	7	0	53.33	100.00
Cloudflare	1477	1873	1477	0	0	100.00	100.00
Comcast	3080	7365	3080	4285	0	41.82	100.00
Google	3452	3925	3352	573	100	85.40	97.10
Fastly	299	404	299	105	0	74.01	100.00
IIJ	133	143	133	0	0	100.00	100.00
Internet2 affiliates	1422	10053	4080	5973	8	40.58	99.80
Meta (Facebook)	286	391	284	107	2	72.63	99.30
Microsoft	567	969	565	404	2	58.31	99.65
Oracle	747	1739	746	993	1	42.90	99.87
Zscaler	72	536	68	468	4	12.69	94.44
Email Respondents	5	5	5	0	0	100.00	100.00
Total	**20976**	**38118**	**24826**	**12476**	**243**	**66.55**	**99.03**

Table 13: Validation results for IPv4 Prefixes in Prefix2Org

Companies	True Prefixes	Predicted Prefixes	True Positive	False Positive	False Negative	Precision	Recall
Amazon	4007	4051	3962	89	45.00	97.80	98.88
Anthropic	1	1	1	0	0.00	100.00	100.00
Censys	4	8	4	4	0	50.00	100.00
Cloudflare	1638	1939	1638	301	0.00	84.48	100.00
Comcast	18	4393	18	4375	0.00	0.41	100.00
Google	211	519	211	308	0.00	40.66	100.00
Fastly	642	761	642	119	0.00	84.36	100.00
Facebook	314	407	314	93	0.00	77.15	100.00
IIJ	3	7	3	4	0.00	42.86	100.00
Internet2 affiliates	172	1596	256	563	4	31.26	98.46
Microsoft	31	72	31	41	0	43.06	100.00
Total	**7041**	**13754**	**7080**	**5897**	**49**	**54.56**	**99.31**

Table 14: Validation results for IPv6 Prefixes in Prefix2Org

ARIN does not consider IP leasing an allowable justification for obtaining or requesting additional IPv4 address blocks [56].

IP leasing can appear in different ways in Prefix2Org, depending on the agreements between the lessors (holders of the address space offered on lease), the leasing entity (intermediate organization that brokers the IPs), and lessees (entities using the leased address space in accordance with the leasing contract): *(i)* in some cases, the Direct Owner recorded in Prefix2Org is the lessor or leasing entity itself, as they hold and monetize address space directly delegated to them in the leasing market. When these addresses are leased, the lessee may appear as a Delegated Customer in WHOIS and hence Prefix2Org. *(ii)* alternatively, the leasing entity may be listed as a Delegated Customer if the contractual arrangement provides them with the rights of an IP allocation type. In these situations, the leasing entity might further include the lessee's organization name or "Private Customer" in the delegation record, if it is articulated as an IP sub-delegation. *(iii)* it is also possible that neither the leasing entity nor the final lessee appears in any delegation record. This occurs when the lessor retains full management and control of the address space.

In summary, Prefix2Org does not explicitly distinguish between lessors, leasing entities, and lessees; they are treated as Direct Owners or Delegated Customers based on observable WHOIS delegation records.