

Lightning talk, NANOG 87

Improving Inference of Sibling Autonomous Systems

Zhiyi Chen, Zachary Bischof, Cecilia Testart, Alberto Dainotti
Georgia Institute of Technology



Significance of accurate inferences of sibling ASes

Accurate inferences on sibling ASes and their owner organizations is essential for a variety of research topics.

1

BGP hijacking
suspicion-level
inference

2

AS relationship
inference

3

AS type
classification

4

Internet
censorship
monitoring

Data sources

Whois databases & CAIDA AS-to-organization dataset (CA20)


ASN: 3356

Source Registry	ARIN
Number	3356
Name	LEVEL3
Handle	AS3356
Registration	Fri, 10 Mar 2000 05:00:00 GMT (Fri Mar 10 2000 local time)
Last Changed	Tue, 20 Feb 2018 17:38:56 GMT (Tue Feb 20 2018 local time)
Self	https://rdap.arin.net/registry/autnum/3356
Alternate	https://whois.arin.net/rest/asn/AS3356
Port 43 Whois	whois.arin.net

Related Entities ▾ 1 Entity

Source Registry	ARIN
Kind	Org
Full Name	Level 3 Parent, LLC
Handle	LPL-141
Address	100 CenturyLink Drive Monroe LA 71203 United States
Roles	Registrant

An example of Whois data from ARIN

Org name	Level 3 Parent, LLC		
country	United States 		
Org rank	1		
Customer Cone	48871 asns	733637 prefixes	2107947909 addresses
ASN degree	7215 transit	7313 global	
ASN members	14 observed	40 total	

AS Rank ▲	AS Number ▾	AS Name
Level 3 Parent, LLC		
1	3356	LEVEL3
35	3549	LVLT-3549
328	202	CENTURYLINK-LEGACY-...

An example of CA20 inference

Example of incorrect sibling relations

```

aut-num:          AS9426
as-name:          WESTPAC-AS-AP
descr:            Westpac Bank
descr:            Retail Bank in Australia
org:              ORG-SOPL2-AP
organisation:    ORG-SOPL2-AP
org-name:         SingTel Optus Pty Ltd
country:         AU
    
```

Actual owner

Whois data from APNIC

AS number	9426
AS name	WESTPAC-AS-AP
organization	SingTel Optus Pty Ltd

Inference from CA20

Inferred owner

Org name	SingTel Optus Pty Ltd		
country	Australia 🇦🇺		
Org rank	86		
Customer Cone	499 asns	19974 prefixes	35220841 addresses
ASN degree	284 transit	309 global	
ASN members	18 observed	63 total	

Inaccurate sibling inferences in CA20

Example of missing sibling relations

CA20 maps the ASes to **different** organization objects

Number	19047	Kind	Org
Name	AWS-01	Full Name	Amazon.com, Inc.
Handle	AS19047	Handle	AMAZON-4
Number	40045	Kind	Org
Name	AWS	Full Name	Amazon.com, Inc.
Handle	AS40045	Handle	AMAZO-4

Whois data from ARIN

Org name	Amazon.com, Inc.		
country	United States 🇺🇸		
Org rank	3371		
Customer Cone	6 asns	8528 prefixes	60456704 addresses
ASN degree	218 transit	367 global	
ASN members	3 observed	3 total	

AS Rank ▲	AS Number ▼	AS Name
Amazon.com, Inc.		
3768	16509	AMAZON-02
13691	7224	AMAZON-AS
17926	19047	AWS-01

Org name	Amazon.com, Inc.		
country	United States 🇺🇸		
Org rank	5634		
Customer Cone	3 asns	718 prefixes	16338432 addresses
ASN degree	0 transit	1 global	
ASN members	1 observed	3 total	

AS Rank ▲	AS Number ▼	AS Name
Amazon.com, Inc.		
11369	14618	AMAZON-AES
74809	399834	AWS
74809	40045	AWS

Inferences from CA20

❌ Automatic and accurate research efforts



Our contributions

0

Leveraged PeeringDB as an additional source and discovered CA20 inaccuracies.

1

Identified **two** root causes of inaccuracies in Whois datasets.

2

Generated a manually improved dataset, corrected inferences of more than 1K organizations and around 4K ASes.

3

Proposed an automatic approach for efficiently reconstructing the manually improved dataset with 95% of fidelity.



Manual research

- Contacted operators
- Reached out to RIRs
- Searched on Google
- Checked organization websites
- Keyword matching
- ...

Two root causes

**APNIC LIR
Issue**

**Multi-orgID
Issue**

APNIC LIR Issue



APNIC LIR Issue

APNIC LIRs apply for AS numbers **on behalf of** their **customers**

OrgID field in Whois: the **LIR's orgID**

Descr field in Whois: the **actual owner's** name

CA20 considers orgID as the only identifier of organizations.

Multi-orgID Issue



Multi-orgID Issue

Reasons:

Subsidiaries and working groups

Registration of ASes under different RIRs

Acquisitions and mergers



Our findings

We corrected mappings for 1,028 CA20 organizations,
associated with around 4K ASes.



APNIC LIR Issue

Incorrect sibling relations

15%



Multi-orgID Issue

Missing sibling relations

85%

Automatic approach

Efficiently generated an updated improved dataset for each new version of the CA20 dataset.

Improve Inference of Sibling Autonomous System

Passive and Active
Measurement
Conference 2023

Github repo of our improved dataset

<https://github.com/InetIntel/Improving-Inference-of-Sibling-ASes>

Contact

zchen798@gatech.edu





Thank you!