AS Relationships Inference from the Internet Routing Registries

Akmal Khan, *Hyun-chul Kim, **Ted "Taekyoung" Kwon

The Islamia University of Bahawalpur, *Sangmyung University, **Seoul National University

akmal_shahbaz@yahoo.com,hyunchulk@gmail.com;tkkwon@snu.ac.kr

ABSTRACT

We present a methodology to infer business relationships between ASes using routing polices stored in the Internet Routing Registries (IRR), which are a set of databases used by ASes to register their inter-domain routing policies. We show that the overall accuracy of our algorithm is comparable (95% for p2c, 92% for p2p links) to the existing algorithms, which infer AS relationships using BGP AS paths. We highlight that the IRR is a strong complementary source for better understandings of the structure, performance, dynamics, and evolution of the Internet since it is actively used by a large number of operational ASes in the Internet.

CCS CONCEPTS

• **Networks → Routing protocols**.

KEYWORDS

Inter-domain Routing, Border Gateway Protocol (BGP), Internet Routing Registries

1 INTRODUCTION

Accurate knowledge of business relationships between autonomous systems (ASes) is relevant to both technical aspects (e.g., network robustness, traffic engineering) and economy-based modeling of the evolution of Internet [9–13]. However, as business relationships between ASes are generally not publicly disclosed, considerable effort has been made to infer AS relationships between ASes [9–13].

The seminal work by Gao [10] infers relationships between ASes based on the valley-free property of AS paths, i.e., each AS path consists of an uphill segment of zero or more c2p or sibling links, zero or one p2p links at the top of the AS path, followed by a downhill segment of zero or more p2c or sibling links. More recently, Luckie *et al.* [11] proposed a method based on less restrictive valley-free property rules and validated 34.6% of their inferred AS relationships.

Similar to Nemecis [9], we highlight that the policies stored in the IRR can be used to infer AS relationships. A similar approach

(i.e., relying on the availability of routing policies from both side ASes of an AS link) was used by Luckie *et al.* [11] for possibly more accurate AS relationship inference, which resulted in extracting only 6.5 K p2c relationships from the IRR. In contrast, we show that a larger number of AS relationships for both p2c and p2p types can be accurately inferred even when only one side AS of an AS link has made their routing policies available in the IRR. Since most of other proposed AS relationship inference methods [10–13] use information of BGP AS paths, we demonstrate that inferring AS relationships from the IRR can help to cross-validate the inferences that are made by BGP AS paths. We show that the overall accuracy of our algorithm is comparable (95% for p2c, 92% for p2p links) to the existing algorithms, which infer AS relationships using BGP AS paths.

2 AS RELATIONSHIP INFERENCE

We present a methodology to infer business relationships between ASes using routing polices stored in the Internet Routing Registries (IRR) [1–7], which are a set of databases used by ASes to register their inter-domain routing policies. The routing policy information in the IRR is registered using a standard language, Routing Policy Specification Language (RPSL) [8]. In RPSL, a **mntner** object is used to register an authorized entity to add, delete, or modify objects related to an AS. When an AS needs to create and specify routing policies for a set of neighboring ASes, **as-set** objects are used. The as-set objects are hierarchical in nature as they can refer to other as-set objects. For registering import and export policies towards neighboring ASes, **aut-num** objects are used.

AS links observed in aut-num objects: Table 1 shows the following three policy registration practices of an AS (ASx) that can be used to infer its relationships with a neighboring AS (ASy): (1) ASx does not register the keyword **ANY** in its import and export polices towards ASy, then we classify the link as of type peer-to-peer (p2p). In other words, in a p2p relationship, ASes import only objects (e.g., as-set objects) maintained by their neighbor ASes and export only objects maintained by themselves. (2) If ASx registers the keyword **ANY** in its export policy for ASy but accept only ASy in its import policy, then we classify the AS link as of type provider-to-customer (p2c), i.e., ASes send all routes to their customer ASes in a p2c relationship. (3) If ASx registers the keyword **ANY** in its import policy from ASy but announce only ASx in its export policy, then we classify the AS link as of type customer-to-provider (c2p); ASes accept all routes from their provider ASes in a c2p relationship. After inferring c2p relationships, similarly to other AS relationship datasets [11], we reverse the direction of the AS link and store it as of type p2c.

Corresponding Authors: Ted "Taekyoung" Kwon (tkkwon @snu.ac.kr) and Hyun-chul Kim (hyunchulk@gmail.com)

Table 1: An example of AS relationship inference.

ASx's routing policy for ASy		
Import	**Export**	**Relationship**
1. ASy	ASx	p2p
2. ASy	ANY	p2c
3. ANY	ASx	c2p

AS links observed in as-set objects: Since we do not find routing policy annotations for the AS links only observed in as-set objects, we use the following three as-set objects naming conventions to infer AS relationships: (i) ASes name their as-set objects to specify whether the as-set object is composed of their customer ASes or peer ASes. Thus, AS links from the as-set objects whose name contains texts like "customer", "downstream", or "client" are classified as of type p2c. (ii) ASes name their as-set objects to specify the location of their BGP peerings, e.g., as-set object "AS2:AMS-IX" specifies the peering ASes of AS2 at AMS-IX. Thus, we classify links observed in as-set objects containing abbreviations of IXP names as of type p2p. Most ASes setup p2p relationships at IXPs though other type of relationships are also possible [11–13]. (iii) ASes name their as-set objects with a text like "upstream" to specify their provider ASes. Thus, links from such as-set objects are classified as c2p.

For as-set objects with no hints (in the name) about any AS relationships, if the exporter-AS (or referrer-AS) of an as-set object exists in the as-set object as a member AS, then the as-set object consists of customer ASes of the exporter-AS (or referrer-AS). More specifically, due to similar routing policies for customer ASes and its own AS, ASes often register their own ASes as a member AS in an as-set object containing customer ASes. However, since policies can be different for peers and provider ASes, ASes do not register their own AS as a member AS in an as-set object containing its peer or provider ASes. Consequently, all the AS links in the as-set object are classified as of type p2c. Note that we do not consider s2s relationships, as we find only a very small fraction (0.08%) of IRR AS links that are of type s2s. Thus, we only consider p2c and p2p type of AS relationships.

3 EVALUATION

In this section, we evaluate our proposed AS relationship inference method (described in Section 2) against existing methods that are based on BGP AS paths. We also compare our results with two ground-truth datasets shared by Luckie *et al.* [11].

CAIDA AS Relationships: Luckie *et al.* [11] refined existing AS relationship inference methods that are based on AS paths in BGP, and validated a large number of inferred AS relationships by collecting ground truth information (i) directly reported by network operators, (ii) extracted from the IRR RPSL objects, and (iii) obtained from BGP community values in BGP traces. We find 132,565 p2c links and 227,470 p2p links in the dataset of 1st. Jan. 2020.

ProbLink: Jin *et al.* [13] developed a probabilistic algorithm, ProbLink, to infer AS relationships by overcoming the challenges in inferring hard links, such as nonvalley-free routing, limited visibility, and non-conventional peering practices. We find 93,920 p2c links and 205,095 p2p links in ProbLink shared dataset of 1st. Jan. 2020.

Table 2: IRR AS Relationships compared with others.

Name	Matching p2c	Matching p2p
ProbLink	49,204 (94.6%)	53,196 (91.9%)
CAIDA	69,687 (94.8%)	48,561 (90.5%)
GT-RPSL	5,191 (97.8%)	N/A
GT-Comm	11,927 (92.5%)	8,417 (88.8%)

GT-RPSL and GT-Comm: These are the ground truth datasets shared by Luckie *et al.* [11]. For GT-RPSL, they extracted 6,530 p2c relationships from routing policies registered in the RIPE IRR dataset of Apr. 2012. For GT-Comm, they extracted 41,604 relationships (16,248 p2p and 23,356 p2c) by using a dictionary of 1,286 BGP community values from 224 ASes, which is constructed from the BGP traces of Apr. 2012.

3.1 Results

We infer 389,451 p2p and 220,556 p2c AS relationships from the IRR dataset of 1st. Jan. 2020 [2–7]. Table 2 shows the fraction of the inferred AS relationships matching with those of the other existing algorithms.

We observe that a high fraction (92.5-94.8%) of p2c relationships are consistently matched with the other datasets. For the results with p2p relationships, we find 1% improvement of ProbLink method over CAIDA, which has been achieved by relying less on the valley-free property; some AS paths in the BGP do not follow the valley-free property due to BGP mis-configurations, poisoned paths, or special routing policies [11].

Interestingly, 97.8% of our inference results are matched with GT-RPSL, which is extracted from the RIPE IRR by evaluating the policies of both-end ASes in an AS link. We further highlight that inferring AS relationships from the policy of a single AS in an AS link is also highly accurate. For the mismatching 2.2% relationships (i.e., 116 AS links), we find that 96 of the reported p2c relationships in GT-RPSL has recently been changed from p2c to p2p, which also could have been correctly inferred by our inference method, had it not been changed since Apr. 2012. The remaining 20 (out of 116 AS links) p2c links in GT-RPSL do not match as these links are of type s2s as reported by our AS relationship inference method. Note that GT-RPSL does not contain any p2p links. However, we show that p2p links inferred by our method are also matched well with the ones inferred by other methods. Finally, we find that the accuracy of our methods against the ground-truth dataset extracted from BGP communities (GT-Comm) is also significantly high; 92.5% for p2c and 88.8% for p2p.

In our future work, we are working to further improve our matching results of AS relationship inference, performing analysis to find the reasons behind AS relationship mis-matches between BGP and IRR based datasets, and comparing AS topology from the IRR with BGP and traceroute based datasets.

4 ACKNOWLEDGEMENTS

REFERENCES

[1] Internet Routing Registry. http://www.irr.net.
[2] RIPE DB. ftp://ftp.ripe.net/ripe/dbase.
[3] RADb. ftp://ftp.radb.net/radb/dbase.
[4] AfriNIC FTP. ftp://ftp.afrinic.net/.
[5] APNIC FTP. ftp://ftp.apnic.net/.
[6] ARIN FTP. ftp://ftp.arin.net/.
[7] LACNIC FTP. ftp://ftp.lacnic.net/.
[8] L. Blunk, J. Damas, F. Parent, A. Robachevsky. Routing Policy Specification Language. *RFC 4012*, Mar, 2005.
[9] G. Siganos and M. Faloutsos. Analyzing BGP Policies: Methodology and Tool. In *IEEE INFOCOM*, Mar, 2004.
[10] L. Gao. On Inferring Autonomous System Relationships in the Internet. *IEEE/ACM TON*, 2001.
[11] M. Luckie, B. Huffaker, k. claffy, A. Dhamdhere, and V. Giotsas, , "AS Relationships, Customer Cones, and Validation", in *ACM IMC*, Oct 2013.
[12] G. Feng, Srinivasan Seshan, and Peter Steenkiste. 2019. UNARI: an uncertainty-aware approach to AS relationships inference. In *CoNEXT*, Feb 2019. Proceedings of the 15th International Conference on Emerging Networking Experiments And Technologies (CoNEXT 19). Association for Computing Machinery, New York, NY, USA, 272-284. DOI:https://doi.org/10.1145/3359989.3365420
[13] Y. Jin, C. Scott, A. Dhamdhere, V. Giotsas, A. Krishnamurthy, and S. Shenker, "Stable and Practical AS Relationship Inference with ProbLink", In USENIX Symposium on Networked Systems Design and Implementation *NSDI*, Feb 2019, pp. 581–597.