Revisiting IP-to-AS mapping for AS-level traceroute

Baobao Zhang[1,2,3] zhang-bb10@mails.tsinghua.edu.cn

Jun Bi[1,2,3] Junbi@tsinghua.edu.cn

Yangyang Wang[1,2,3] Wangyy-06@mails.tsinghua.edu.cn

Yu Zhang[4] yu zhang@hit.edu.cn

Jianping Wu[1,2,3] jianping@cernet.edu.cn

1, Network Research Center, Tsinghua University; 2, Department of Computer Science, Tsinghua University
3, Tsinghua National Laboratory for Information Science and Technology (TNList)
4, Department of Computer Science and Technology, Harbin Institute of Technology, China

ABSTRACT

On the way to obtaining accurate AS-level traceroute paths, lots of efforts have focused on the improvement of the original IP-to-AS mapping table which was extracted from BGP routing tables. One of those efforts is called pair matching which refines the original mapping table by maximizing the number of matched pairs of traceroute and BGP AS paths from the same AS to the same destination. However, in the existing pair-matching-based method, the granularity for mapping is prefix, i.e. that the IP addresses in the same /24 prefix always belong to the same AS or set of ASes, which may yield ambiguity and does not hold in some cases. In this paper, we revisit the IP-to-AS mapping with IP-address granularity by allowing IP addresses in the same prefix to be mapped to different ASes.

Keywords

IP-to-AS Mapping, IP-granularity, Traceroute, BGP

1. INTRODUCTION

This paper aims to get the accurate AS-level traceroute path, so we need one accurate IP-to-AS mapping, which can map the IP address to the AS (autonomous system) that is using the address. It is of great significance to diagnose network failure and discover the AS topology. However, there has not been a completely accurate IP-to-AS mapping up to now. There are two kinds of approaches to refine the IP-to-AS mapping: (1) IP-router-AS method, which is based on alias resolution [1], firstly maps the IP address to the router, then maps the router to AS; (2) pair-matching method [2], which is based on the assumption that the traceroute path is consistent with the BGP AS path [3], modifies the IP-to-AS mapping to maximize the number of matched traceroute-BGP path pairs.

Our paper focuses on the pair-matching method. The IP-router-AS method has received lots of interests, while the pair-matching method has not been improved since the first work was published in 2004. However recently, [4][5] utilized the pair matching to quantify the pitfalls of using traceroute. Then in this paper, we wish to improve the mapping by training the mapping table with path pair data.

ACM CoNEXT Student Workshop, December 6, Tokyo, Japan

The existing best pair-matching method is [2]. [2] takes /24 prefix as granularity to modify IP-to-AS mapping. It is not true in real world: IP addresses in the same /24 prefix may not belong to the same ASes, especially for IP addresses of routers. The reason is as follows: (1) border AS router. It may use IP addresses assigned to adjacent ASes; (2) Internet Exchange Point (IXP). It is used to exchange traffic between different ASes. IP addresses in the same IXP prefixes are used by different ASes. Although [2] allows that prefixes are mapped to one set of ASes, it yields ambiguity. So we call [2]'s method Prefix-granularity method with ambiguity (PGMA). As is known to us, one IP address is only used by one only AS, so the IP-to-AS mapping should be deterministic rather than ambiguous.

One simple idea: the prefix granularity is not fine-grained enough to refine the IP-to-AS mapping, so we use IP address granularity, i.e. IP-granularity method (IGM). The idea is simple, but the improvement is large. This paper revisits the IP-to-AS mapping with IGM, and compares IGM with PGMA comprehensively. Besides, we can use the IP-to-AS mapping refined by IGM to infer IXP prefixes.

2. METHODOLOG Y

PGMA and IGM both modify mappings based on the original mapping extracted from routing tables.

(1) Prefix-granularity Method with ambiguity (PGMA)

It takes /24 prefix as granularity to modify mappings, i.e. [2]'s method. It uses threshold to restrict the number of mapping ASes of prefixes. Prefix-granularity method (PGM) and Limit method are two extreme conditions of PGMA: In PGM, the prefix is mapped to one only AS while in limit, the prefix is mapped to its all possible ASes.

(2) IP-granularity Method (IGM)

It takes IP address as granularity to modify mappings. The algorithm thought and frame are the same with PGMA.

3. EVALUATION

We collected the BGP routing tables on 2010-04-22 from Routeviews[6] and Ripe[7](ten collectors).We then extracted the original IP-to-AS mapping from routing tables. We collected the traceroute probes on 2010-04-22 from CAIDA[8]. To process into pairs, it is required that the traceroute monitor and its BGP monitor should be located in the same AS. The monitors that satisfy the requirement are as shown in Table 1. The traceroute path and its corresponding BGP AS path forms one pair.For consistency with [2], we also consider ambiguous match as match, but we mark ambiguous match out.We use three of the four monitors as the training dataset, and the left as the test dataset. So we can get four groups of training datasets and test datasets as shown in Table 2.

Table 1: Data Source Information

Monitor	AS number	# pair	#IP	#prefix
nrt-jp	7660	280K	373K	164K
she-cn	4538	243K	380K	167K
jfk-us	6939	312K	296K	125K
lax-us	2152	282K	374K	164K

Table 2: Training Datasets and Test Datasets

ID	Training Dataset	Test Dataset
1	AS4538,AS6939,AS2152	AS7660
2	AS7660,AS6939,AS2152	AS4538
3	AS7660,AS4538,AS2152	AS6939
4	AS7660,AS4538,AS6939	AS2152

3.1 Training Match Ratio

We run the IP-to-AS refining method on the training dataset, the match ratio on which is called training match ratio, as shown in FIG. 1. The IGM achieves the higher match ratio than PGMA. Although Limit has highest match ratio, the matched pairs are almost ambiguous matched.

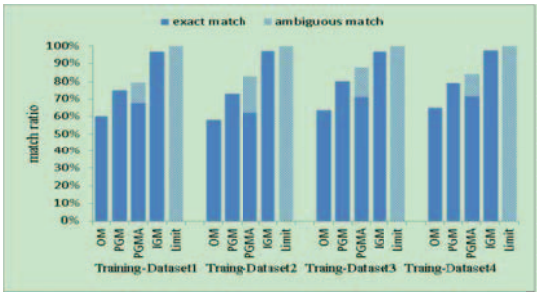

Figure 1: Training match ratio.

3.2 Test Match Ratio

We applied the trained mapping to its test dataset, the match ratio on which is called test match ratio, as shown in FIG.2. Similar with training match ratio, IGM achieves the higher match ratio than PGMA. Limit has large ambiguity.

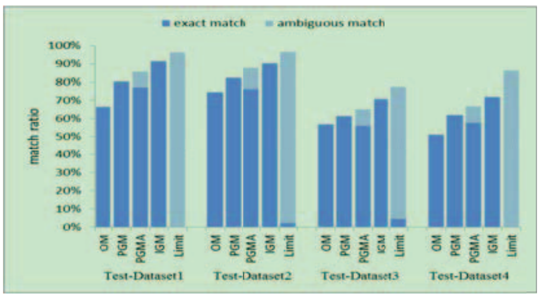

Figure 2: Test match ratio.

Acknowledgement: This research is supported by National Science Foundation of China under Grant 61073172, Specialized Research Fund under Grant 200800030034, and National Basic Research Program ("973" Program) of China under Grant 2009CB320501. Besides, we thank CAIDA, Routeviews and Ripe for supporting data.

3.3 Inferring IXP

The characteristic of IXP prefixes: IP addresses in IXP prefixes are used by many different ASes. We can use the IP-granularity mappings refined by IGM to infer IXP prefixes. The simple reference method is: firstly, we count the number of mapping ASes for every /24 prefix; then if the number of its mapping ASes is larger than 4, we say that the /24 prefix is IXP prefix. We take training dataset3 for example: there are 28 /24 prefixes whose number of mapping ASes is larger than 4, and 21 of these 28 prefixes have been validated by IXP prefixes that we collected from PeeringDB[9], PCH[10] and Euro-IX [11] on 2009-03-09. In table 3, we list the top three IXP prefixes we referred.

Table 3: Inferring IXP

Prefix	Mapping ASes num.	Internet Exchange Point Name
80.81.192.0/24	102	DE-CIX
195.66.224.0/24	100	LINX Brocade LAN
195.69.144.0/24	66	AMS-IX

4. CONCLUSION

IGM can achieve higher match ratio than PGM and PGMA. We have known that IP address granularity is important to refine the IP-to-AS mapping. In the future, we will consider the method that combines IP address with prefix in order to complement each other's strengths. Besides, we can use IGM to infer IXP prefixes. The reference method is preliminary, and we will improve it in the future.

5. REFERENCES

[1] http://www.caida.org/publications/papers/2008/alias_resolution_techreport/

[2] Z. M. Mao, D. Johnson, J. Rexford, J. Wang, and R. H. Katz, "Scalable and accurate identification of AS-level forwarding paths," in INFOCOM 2004, 2004.

[3] Z.Morley Mao, Jennifer Rexford, Jia Wang, and Randy Katz, "Towards an Accurate AS-level Traceroute Tool," in Proc. ACM SIGCOMM, September 2003.

[4] Yu Zhang, Ricardo Oliveira, Hongli Zhang, Lixia Zhang , "Quantifying the Pitfalls of Traceroute in AS Connectivity Inference", PAM 2010, April 2010.

[5] Yu Zhang, Ricardo Oliveira, Yangyang Wang, Shen Su, Baobao Zhang, Hongli Zhang, Lixia Zhang, "A Framework to Quantify the Pitfalls of Traceroute in AS-level Topology Measurement", IEEE Journal of Selected Areas in Communications

[6] http://archive.routeviews.org/

[7] http://www.ripe.net/data-tools/stats/ris/ris-raw-data

[8] https://topo-data.caida.org/team-probing/

[9] https://www.peeringdb.com

[10] http://www.pch.net/ixpdir

[11] http://www.euro-ix.net