Understanding the confounding factors of inter-domain routing modeling

Savvas Kastanakis
Lancaster University

Vasileios Giotsas
Lancaster University

Neeraj Suri
Lancaster University

ABSTRACT

The Border Gateway Protocol (BGP) is a policy-based protocol, which enables Autonomous Systems (ASes) to independently define their routing policies with little or no global coordination. AS-level topology and AS-level paths inference have been long-standing problems for the past two decades, yet, an important question remains open: *"which elements of Internet routing affect the AS-path inference accuracy and how much do they contribute to the error?"*. In this work, we: (1) identify the confounding factors behind Internet routing modeling, and (2) quantify their contribution on the inference error. Our results indicate that by solving the first-hop inference problem, we can increase the exact-path score from 33.6% to 84.1%, and, by taking geolocation into consideration, we can refine the accuracy up to 94.6%.

CCS CONCEPTS

• **Networks** → **Network simulations**;

ACM Reference Format:
Savvas Kastanakis, Vasileios Giotsas, and Neeraj Suri. 2022. Understanding the confounding factors of inter-domain routing modeling. In *Proceedings of the 22nd ACM Internet Measurement Conference (IMC '22), October 25–27, 2022, Nice, France.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3517745.3563025

1 RELATED WORK

The pioneering and most classic approach to infer AS-paths was proposed by Gao and Rexford in 2001 [4]. In 2007, Muhlbauer et al. introduced a new abstraction: next-hop atoms [9] which correspond to per-neighbor path choices. In 2012, Gill et al. [6] developed a novel routing tree algorithm that computes paths between all source-destination pairs in an AS graph. Recently, We et al. [14] developed a learning-based technique, taking into consideration the node, link, and path features related to route decisions in practice.

2 EXPERIMENTAL SETUP AND RESULTS

In this project, we infer AS-level paths using the simulator proposed by Sermpezis and Kotronis [13], which offers a Python implementation of the Gao-Rexford model [4]. Additionally, we collect the AS-paths followed in practice through the BGPStream API [10]. BGPStream utilizes vantage points (VPs) from RouteViews and RIPE RIS projects, which provide a partial view of the Internet topology graph. Nonetheless, not all ASes share their routing tables, thus, researchers rely on inferences to study paths from non-VP ASes. Finally, we leverage the following well-known metrics to capture the inference accuracy of the model:

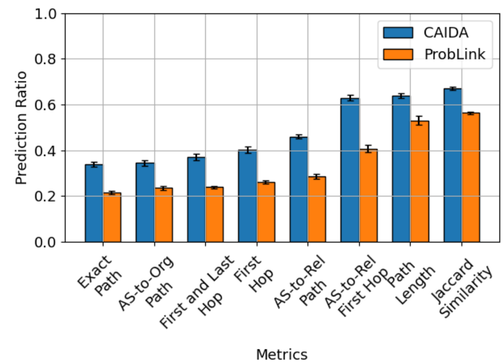

Figure 1: Average inference accuracy (vanilla model).

Exact AS-Path Match: The ratio of inferred AS-paths that are exactly the same as the observed AS-paths.

Path Length Match: The ratio of inferred AS-paths that have the same length as the observed AS-paths.

First-hop Match: The ratio of inferred AS-paths that have the same first-hop as the observed AS-paths.

First and Last-hop Match: The ratio of inferred AS-paths that have the same first and last hop as the observed AS-paths.

AS-to-ORG Path Match: The ratio of inferred AS-to-ORG paths[1] that are the same as the observed AS-to-ORG paths.

AS-to-Rel Path Match: The ratio of inferred AS-to-Rel paths[2] that are the same as the observed AS-to-Rel paths.

AS-to-Rel First-hop Match: The ratio of inferred AS-to-Rel paths that have the same first-hop as the observed AS-to-Rel paths.

Jaccard Similarity: The intersection of the inferred and observed AS-paths over the union of the inferred and observed AS-paths.

We conduct three rounds of Monte-Carlo simulations using the Gao-Rexford model and the current state-of-the-art AS-level topologies, CAIDA and ProbLink [2, 8]. In the first round, we study the performance of the vanilla Gao-Rexford model. In the second round, we explore the performance of the Gao-Rexford model given that we have knowledge over the first-hop. In the last round of simulations, we identify and quantify the confounding factors that affect the inference accuracy.

[1] A path of organization IDs.
[2] A path of AS-relationships.

IMC '22, October 25–27, 2022, Nice, France

ACM ISBN 978-1-4503-9259-4/22/10.
https://doi.org/10.1145/3517745.3563025

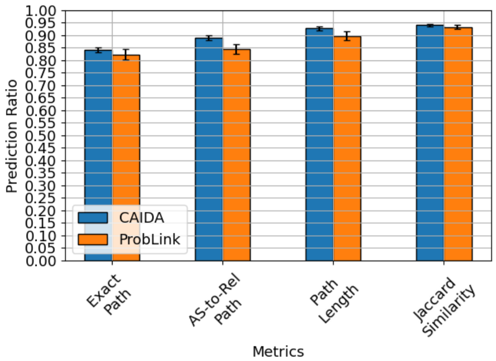

Figure 2: Average accuracy when first-hop known.

In Fig. 1, we plot the Gao-Rexford model inference accuracy. Overall, using the CAIDA topology we achieve at least 10% better accuracy across all metrics, compared against ProbLink. Regarding the main metric of our analysis, *Exact Path Match*, the model scores 33,8% and 21,4% using CAIDA and ProbLink respectively. This is evidence that the limitations addressed in recent work [1] still hold: *the Gao-Rexford model can predict AS-paths exactly as they are observed on the real Internet, only 1/3 of the times.*

To evaluate a first-hop aware model, we only consider inferences for which we can correctly predict the first-hop in the AS-path. In Fig. 2, we plot the inference accuracy of the first-hop aware model and observe that it yields 2,5 times higher accuracy than the vanilla model (84,1% and 82,1% for CAIDA and ProbLink respectively). This is a strong indicator that **(1) the ability to determine the first-hop can significantly affect the overall inference**.

Finally, we study the reasons behind the exact path misses, by identifying the first AS-link in the inferred path that differs from the actual path (see Table 1). The confounding factors are as follows:

Missing AS-relationship: The topology dataset does not include a link for the respective ASes in the observed path.

Valley-free violation: The actual path has a valley [7], hence, the model, by default, cannot predict this route.

Local preference violation: The model selects a route through a *less preferable neighbor* than the observed path.

Shortest path violation: The model selects a *longer* path than the observed path.

Location-agnostic path selection: Due to the Internet flattening [5], it is reasonable to consider the geographical distance between ASes in the BGP route selection process [12]. Yet, neither the actual BGP best path process [11] nor the Gao-Rexford model consider geolocation. We leverage country-level location information from the AS-rank API [3] and study whether the inferred paths go through longer distances than the observed paths.

From Table 1, we observe that **(2) having resolved the first-hop problem, the most important factor that affects the inference process is the *location-agnostic path selection*.** We neither plan to fix the missing links (0.95%), nor replace the valley-free model (4.36%), hence, the maximum achievable accuracy is 94.69%. Preliminary results show that we can improve the accuracy of existing inference techniques to 91.63%, given that we include a location-aware methodology in the inference process.

	CAIDA	ProbLink
Exact Path Match	**84.1** %	**82.1** %
Missing AS relationships	0.95 %	0.53 %
Valley-free violations	4.36 %	1.32 %
Local-pref violations	**7.86** %	**11.42** %
Shortest path violations	2.93 %	3.62 %
Loc-agnostic selection	**8.54** %	**10.31** %

Table 1: Confounding factors

Currently, we are working on identifying the specific what-if questions that can be addressed with existing models, and explore the benefits of location-aware prediction modeling.

ACKNOWLEDGMENTS

We thank Ioana Livadariu for her valuable comments. This research was supported, in part, by EC H2020 CONCORDIA GA No. 830927 and by the UKRI Trustworthy Autonomous Systems Node in Security [EPSRC grant EP/V026763/1].

REFERENCES

[1] Ruwaifa Anwar, Haseeb Niaz, David Choffnes, Ítalo Cunha, Phillipa Gill, and Ethan Katz-Bassett. 2015. Investigating interdomain routing policies in the wild. In *Proceedings of the 2015 Internet Measurement Conference.* 71–77.
[2] CAIDA. 2022. AS-Relationships Dataset. https://publicdata.caida.org/datasets/as-relationships/. (2022). [Online; accessed 16-May-2022].
[3] CAIDA. 2022. ASRank API. https://api.asrank.caida.org/v2/restful/doc. (2022). [Online; accessed 16-May-2022].
[4] Lixin Gao and Jennifer Rexford. 2001. Stable Internet routing without global coordination. *IEEE/ACM Transactions on networking* 9, 6 (2001), 681–692.
[5] Phillipa Gill, Martin Arlitt, Zongpeng Li, and Anirban Mahanti. 2008. The flattening internet topology: Natural evolution, unsightly barnacles or contrived collapse?. In *International Conference on Passive and Active Network Measurement.* Springer, 1–10.
[6] Phillipa Gill, Michael Schapira, and Sharon Goldberg. 2012. Modeling on quicksand: Dealing with the scarcity of ground truth in interdomain routing data. *ACM SIGCOMM Computer Communication Review* 42, 1 (2012), 40–46.
[7] Vasileios Giotsas and Shi Zhou. 2012. Valley-free violation in internet routing—analysis based on bgp community data. In *2012 IEEE International Conference on Communications (ICC).* IEEE, 1193–1197.
[8] Yuchen Jin, Colin Scott, Amogh Dhamdhere, Vasileios Giotsas, Arvind Krishnamurthy, and Scott Shenker. 2019. Stable and Practical AS Relationship Inference with ProbLink. In *16th USENIX Symposium on Networked Systems Design and Implementation (NSDI 19).* 581–598.
[9] Wolfgang Mühlbauer, Steve Uhlig, Bingjie Fu, Mickael Meulle, and Olaf Maennel. 2007. In search for an appropriate granularity to model routing policies. *ACM SIGCOMM Computer Communication Review* 37, 4 (2007), 145–156.
[10] 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.* 429–444.
[11] Yakov Rekhter, Tony Li, and Susan Hares. 2006. *A Border Gateway Protocol 4 (BGP-4).* RFC 4271. https://www.rfc-editor.org/rfc/rfc4271.txt
[12] Matthew Roughan, Walter Willinger, Olaf Maennel, Debbie Perouli, and Randy Bush. 2011. 10 lessons from 10 years of measuring and modeling the internet's autonomous systems. *IEEE Journal on Selected Areas in Communications* 29, 9 (2011), 1810–1821.
[13] Pavlos Sermpezis and Vasileios Kotronis. 2019. Inferring catchment in internet routing. *Proceedings of the ACM on Measurement and Analysis of Computing Systems* 3, 2 (2019), 1–31.
[14] Tianhao Wu, Jessie Hui Wang, Jilong Wang, and Shuying Zhuang. 2022. RouteInfer: Inferring Interdomain Paths by Capturing ISP Routing Behavior Diversity and Generality. In *International Conference on Passive and Active Network Measurement.* Springer, 216–244.