Poster: Understanding the confounding factors of inter-domain routing modeling

Savvas Kastanakis
Lancaster University

Vasileios Giotsas
Lancaster University

Neeraj Suri
Lancaster University

1 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. Even though the Internet AS-level topology and the AS-level paths' inference have been long-standing problems for the past two decades, 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 further taking geography into consideration, we can refine the accuracy up to 94.6%.

2 EXPERIMENTAL SETUP AND RESULTS

In this project, we infer AS-level paths between arbitrary sources and destinations, using the simulator of Sermpezis and Kotronis [9], which offers a Python implementation of the Gao-Rexford model [4]. Additionally, we collect the AS-paths followed in practice through the BGPStream API [7]. Finally, we define metrics which help us capture the efficacy of the Gao-Rexford model against the ground truth paths. The metrics are as follows:

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

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

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

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

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.

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

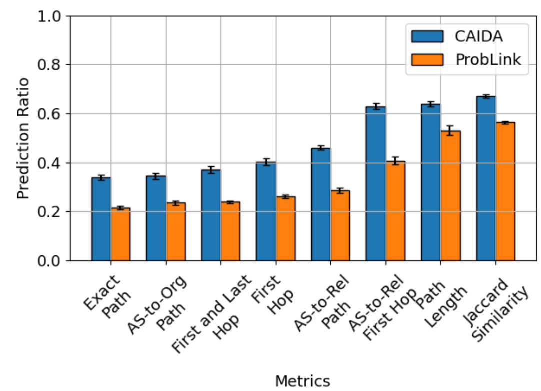

Figure 1: Average inference accuracy (vanilla model).

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

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

To infer the AS-paths between the above AS-pairs we conduct Monte-Carlo simulations using the Gao-Rexford model and the current state-of-the-art AS-level topologies, CAIDA and ProbLink [2, 6]. We perform 3 rounds of simulations. 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 study the confounding factors that affect the inference accuracy.

In Fig. 1, we present the accuracy of the model across all performance metrics using both topology datasets. Overall, using the CAIDA topology we achieve at least 10% better accuracy across all metrics, compared against the ProbLink topology. Regarding the main metric of our analysis, *Exact Path Match*, the model scores 33,8% using the CAIDA topology and 21,4% using the ProbLink topology. 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.*

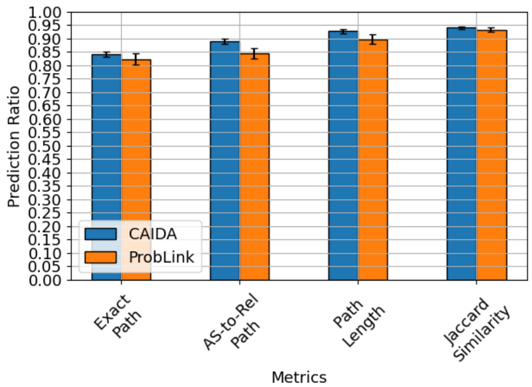

Figure 2: Average accuracy when first-hop known.

To simulate scenarios in which we have knowledge over the first-hop, we only consider entries from our ground-truth dataset, for which we can predict the first-hop in the AS-path correctly, and deal with the problem of predicting the rest of the path. In Fig. 2, we plot the inference accuracy of the first-hop aware model and we observe that it yields a 2,5 times higher accuracy than the vanilla model (84,1% for CAIDA and 82,1% for ProbLink). This is a strong indicator that **(1) the ability to determine the first-hop (2.5x increase in accuracy) can significantly affect the overall inference**.

Finally, we study the reasons behind the exact path misses, by identifying the first *broken link* in the inferred path, i.e., the first AS in the inferred path which is different in the actual path. Then we classify it in one of the following categories:

Missing AS-relationship: In this class, the *broken link* occurs because the AS-relationships' dataset does not include a link for the respective ASes in the observed path.

Valley-free violation: In this class, the *broken link* occurs because the actual path has a valley [5], therefore, the Gao-Rexford model by default cannot produce this route or the AS-relationships, which are inferred using the valley-free model, are incorrect.

Local preference violation: A local-pref violation happens when the model selects a route through a *less preferable neighbor* than the inferred one.

Shortest path violation: A shortest path violation occurs when the model selects a *shorter* path than the actual one.

Geo-agnostic path selection: Due to the flattening of the Internet topology [3], it is reasonable to consider the geographical distance between two ASns when infering AS-paths. Yet, neither the actual BGP best path algorithm [8] nor the Gao-Rexford model consider geography. Hence, in this occasion, the *broken link* is a consequence of the geo-agnostic path selection.

	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 %
Geo-agnostic selection	**8.54** %	**10.31** %

Table 1: Confounding factors

From Table 1, we can see that **(2) having resolved the first-hop problem, the most important factor that affects the inference process is the *geo-agnostic path selection***. For this project, we neither plan to fix the missing AS-relationships (0.95% of exact path misses), nor replace the valley-free model (4.36% of exact path misses), hence, the maximum achievable accuracy we can score with the current approaches, is 94.69%.

As part of our future research, we work on identifying the specific use-cases and what-if scenarios that can be addressed with current models, and explore the promising benefits we can achieve by designing geo-aware prediction models.

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] Yi-Ching Chiu, Brandon Schlinker, Abhishek Balaji Radhakrishnan, Ethan Katz-Bassett, and Ramesh Govindan. 2015. Are We One Hop Away from a Better Internet?. In *Proceedings of the 2015 Internet Measurement Conference (IMC '15)*. Association for Computing Machinery, New York, NY, USA, 523–529. https://doi.org/10.1145/2815675.2815719

[4] Lixin Gao and Jennifer Rexford. 2001. Stable Internet routing without global coordination. *IEEE/ACM Transactions on networking* 9, 6 (2001), 681–692.

[5] 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.

[6] 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.

[7] 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.

[8] 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

[9] 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.