MVP: Measuring Internet Routing from the Most Valuable Points

Thomas Alfroy
University of Strasbourg
talfroy@unistra.fr

Thomas Holterbach
University of Strasbourg
tholterbach@unistra.fr

Cristel Pelsser
University of Strasbourg
pelsser@unistra.fr

ACM Reference Format:
Thomas Alfroy, Thomas Holterbach, and Cristel Pelsser. 2022. MVP: Measuring Internet Routing from the Most Valuable Points
. 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/10.1145/3517745.3563031

1 INTRODUCTION

Scrutinizing BGP routes is part of the everyday tasks that network operators and researchers conduct to monitor their networks and measure Internet routing. This task is facilitated by the expansion of routing information services such as RIPE RIS [2] and RouteViews [3] that collect BGP routes from an increasing number of Vantage Points (VPs). Unfortunately, while more data is often beneficial, in the case of BGP, it involves downloading and processing large volumes of route updates that exhibit a high level of redundancy. Today with more than one billion route updates collected every day, users often have no other option than to focus on a subset of the VP. Because of the highly skewed location of the VP, randomly selecting them may result in a lot of missing information.

In this paper, we present MVP, a system that exploits the similarity between the VPs to help users find which piece of data to use. We design MVP to be generic: regardless of the user's objective, our system allows finding more useful information with fewer data. MVP relies on three core components. First, it quantifies, for each VP, the topological changes seen for past BGP events. Then, it clusters the VPs based on these labels. Finally, it generates a set of dissimilar VPs by selecting VPs in different clusters. We evaluate MVP on several common use-cases, such as Internet topology mapping, and show that MVP addresses these objectives with $\approx 1.6x$ fewer data compared to the existing baselines.

2 MONITORING INTERNET ROUTING

As of April 2022, RIPE RIS deployed >1k vantage points. RIPE RIS [2] and RouteViews [3] aim at collecting BGP routes and making them available to users. To capture as much routing information as possible, these services collect routes from more and more BGP Vantage Points (VP), i.e., routers that mirror their BGP routes to a collector. Especially since 2015, the number of VPs greatly increased over time and reached 1050 in 2021. RouteViews exhibits a similar order of magnitude with no less than 950 VPs in 2022.

RIPE RIS collects >45 million updates every hour. Users can download either a RIB dump, i.e., a snapshot of the BGP routes seen by the VPs at particular times, which results in ≈920k routes [1] for 32% of the VPs that mirror all their routes. Alternatively, users may download every single BGP update observed by the VPs over time, which currently results in >45 million updates every hour for RIPE RIS and exceeds the billion of updates a day when considering both RIPE RIS and RouteViews.

Redundancy between data keep increasing over time. Due to the highly skewed location of the VP in the Internet, most of the VPs have a similar partial view of the Internet, and thus collect redundant data. In fact, 62% of the RIPE RIS VPs are located in the top 5k ASes according to the CAIDA asrank algorithm. Randomly selecting VP leads to a skewed selection, resulting in a lot of redundant information about the core of the Internet.

3 OUR SOLUTION

We propose MVP, a system that returns a list of dissimilar BGP Vantage Points (VPs) that enables users to better measure Internet routing from fewer data. MVP relies on a VPs selection strategy that strikes the best balance between the volume of data to process and its utility, regardless of the user's objective. In this section, we give an overview of MVP's core components (§3.1) and key insights (§3.2).

3.1 Core components

Our solution includes three core components that we outline in the following paragraphs.

Quantification of the VPs' observations through time and topological space. Our objective for choosing these two dimensions is for the vantage point selection to be as generic as possible, and applicable for a large variety of applications. More precisely, we first take BGP events for which we have ground truth: we know their time and location in the AS-level topology. We ensure to consider a large number of events that happened at different locations and times in the AS topology. Then, MVP computes the change these events induced on on graph features that capture local and global properties. MVP then uses the feature values to compute similarity between the VPs.

Calculation of similarity between the VPs. For every BGP event, MVP groups the VPs that collect similar information using a tailored clustering algorithm. From these per-event clusters, our system then computes an overall (i.e., across all the events) similarity score for

---

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

ACM ISBN 978-1-4503-9259-4/22/10.
https://doi.org/10.1145/3517745.3563031

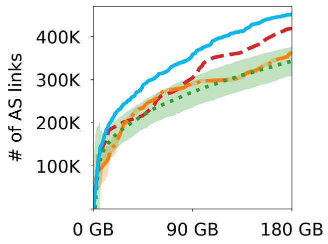

Figure 1: MVP outperforms other baselines.

every VP pair. Basically, the more often two VPs end up in the same cluster, the higher is their overall similarity score.

Generation of a set with the most dissimilar VPs. Finally, the last step is to exploit the computed pairwise VP similarity scores to return a set of dissimilar VPs. MVP relies on a greedy algorithm to build the set of dissimilar VPs. More precisely, it starts with a singleton that includes the most dissimilar VP. Then, it greedily adds the VP that is the most dissimilar with the VPs already in the set. Here, we extend the pairwise similarity score to estimate the compound similarity between a VP and a set of VPs.

3.2 Key insights

We now explain two key insights that allow MVP to perform well in many real-world scenarios.

MVP uses topological features for genericity. There exists several features that quantify the observation of a VP. However, because MVP aims to be as generic as possible, we cannot rely on metrics that quantify the VP observation for some specific use cases. We thus decide to use topological features to avoid MVP overfitting on some particular use cases. These features aim to quantify how BGP events change the AS-level topology. Altogether, they compute generic changes from a local and global perspective without optimizing a specific objective.

MVP considers the volume of data collected by every VP. It turns out that the VPs collect a highly different volume of data. Thus, MVP considers the average volume of data collected by every VP when generating a set of dissimilar VPs. More precisely, at every iteration of our greedy algorithm, MVP adds in the set of VPs the VP that generates the less data among the ones that are the most dissimilar with the VPs currently in the set. This strategy aligns with our initial goal, which is to strike the best balance between the volume of data (and not the number of VPs) and its utility.

4 EVALUATION

In this section, we demonstrate the ability of MVP to generate a set of VPs for which the collected routes provide high utility yet with low volume on three different use-cases, namely Internet topology mapping, transient path detection, and anomaly detection. Because the measured performance of MVP is comparable across these three uses-cases, we only present MVP's performance for the topology mapping use-case. This objective consists in building an accurate view of the Internet at the AS level. We map the AS topology using one RIB dump as well as the route updates collected during 2k periods of 5 minutes randomly selected.

Methodology. We compare MVP against the three following baselines. Each aims to build a set of VPs.
(i) Random. At every iteration, a VP is selected randomly and added to the set. The location of the VP in the Internet is skewed, and so may be the resulting set of VPs.
(ii) AS-distance. The first VP is selected randomly. Then, at every iteration, the VP that maximizes the sum of all the pairwise distances is added to the set. As opposed to the random strategy, the generated set of VPs is not skewed towards the core of the topology.
(iii) Greedy. At every iteration, the VP that discovers the highest number of new AS links is added. Thus, it is the solution that provides the best tradeoff between the number of VPs and the number of discovered AS links. We found this baseline in the literature [4].

Fig. 1 shows the performance of MVP and the three baselines for the three objectives. For this evaluation, we select VPs until the volume of data processed during the 2000 periods of 5 minutes reaches 180GB. We do not select more VPs after this point because the *Greedy* baseline, which requires recomputing at every iteration all the links that every non-selected VPs would discover, is too time-consuming. Because the *Random* and *AS-distance* baselines are probabilistic, we show their maximum, medium and minimum performance, computed from a set of 100 runs and represented by the light green and orange areas around their respective curve.

MVP outperforms every baseline when mapping the Internet Topology. More precisely, with 90GB of processed data, MVP discovers 55K more AS links than the *Greedy* baseline in the medium case. Most of the time, MVP even outperforms the best-case scenario of the *Random* baseline. Using MVP, 300K links can be observed with 51GB of data, whereas 84GB of data are required with *Greedy* (1,6x more). MVP exhibits similar results for the other two objectives.

The fact that MVP even outperforms the *Greedy* baseline that aims to detect as many AS links as possible clearly demonstrates the ability of MVP to find the best balance between the utility and the volume of the selected VPs.

Our tool is available online with the commands:

```
curl -d 'volume=5GB' http://5.161.124.63/mvp
curl http://5.161.124.63/help
```

5 ACKNOWLEDGMENT

Thomas Alfroy is funded by ArtIC project "Artificial Intelligence for Care" (grant ANR-20-THIA-0006-01) and co funded by Région Grand Est, Inria Nancy - Grand Est, IHU of Strasbourg, University of Strasbourg and University of Haute-Alsace. Thomas Holterbach is partially funded by ISOC, the MANRS Observatory and by Région Grand Est.

REFERENCES

[1] [n.d.]. CIDR REPORT Feb 2021. https://www.cidr-report.org/as2.0/.
[2] 2016. RIPE RIS Raw Data. https://www.ripe.net/data-tools/stats/ris/.
[3] University of Oregon. 2016. Route Views. www.routeviews.org/.
[4] Ying Zhang, Zheng Zhang, Z. Morley Mao, Y. Charlie Hu, and Bruce M. Maggs. 2007. On the impact of route monitor selection. In *IMC '07*.