



ARTICLE

Exploring the Temporal Degradation and Drift of AS Path Inference

Xionglve Li¹, Changsheng Hou^{2,*}, Yuzhou Huang³, Zhenyu Qiu¹, Gang Hu¹, Bingnan Hou¹, Wei Dong¹ and Zhiping Cai¹

¹College of Computer Science and Technology, National University of Defense Technology, Changsha, China

²Academy of Military Sciences, Beijing, China

³The Army Logistics Department, Beijing, China

*Corresponding Author: Changsheng Hou. Email: houchangsheng@nudt.edu.cn

Received: 10 February 2026; Accepted: 08 April 2026; Published: 15 June 2026

ABSTRACT: The Internet inter-domain paths, i.e., the AS paths, are important for network management, traffic engineering, and security. Due to business confidentiality, security, and privacy, the AS path information is non-public. Due to limited measurement resources, obtaining AS path information by measurement-based approaches is not scalable. Therefore, path inference approaches are proposed to broaden the availability of path information. These approaches assume that AS paths remain stable over a certain period of time, yet conflicting research findings question this assumption. Furthermore, the duration of the “certain period of time” is not clearly defined. Thus, we aim to address the following question: “*How do the performance and temporal drift of path inference approaches evolve over time?*” In this paper, we conduct a quantitative validation study and a temporal drift analysis to examine the evolution of AS path inference performance over time. The quantitative validation study shows that the minimal performance degradation is only 2.09% over eight weeks. The temporal drift analysis shows that, among the three evaluated methods, KnownPath exhibits the slowest drift, GMPI shows a moderate drift rate, and ProbInfer drifts the fastest under the current decision rule. The results provide preliminary evidence on how historical data can be leveraged despite limited measurement resources and can inform refresh-frequency decisions for path inference services under computational constraints.

KEYWORDS: Network measurement; AS path inference; temporal drift

1 Introduction

The Internet is composed of numerous Autonomous Systems (ASes), and the Border Gateway Protocol (BGP) is the de facto standard for inter-domain routing. The AS path information is important for network management [1,2], traffic engineering [3], and security [4,5]. It is also closely related to routing incidents, anomalous path propagation, and inter-domain attack analysis [6–8]. However, BGP path information is not directly available to the public and can only be obtained through measurement or path inference. Measurement-based approaches include traceroute and collecting BGP data from routers. Due to limited measurement resources, however, these measurement-based approaches are not scalable. Therefore, path inference approaches have been proposed to broaden the availability of path information.

Internet inter-domain path inference has been studied for nearly two decades; approaches can be broadly categorized into heuristic policy-aware inference and data-driven inference. The classic heuristic policy-aware inference approaches are based on the valley-free principle and AS relationships, such as KnownPath [9], iPlane [10], and iNano [11]. Data-driven inference approaches, on the other hand, leverage machine learning techniques to infer paths based on observed data, such as Sibyl [12], ProbInfer [13], and

the recent PathRadar [14]. More recently, our prior work proposed a generative and measurable process for fine-grained AS path inference (GMPI) [15], and further extended this line toward personalized and adaptive inference (PA-GMPI) [16]. We have also systematically reviewed the methodological landscape, applications, and open challenges of Internet inter-domain path inference [17].

These studies have substantially improved the ability to infer unmeasured AS paths, primarily by constructing more accurate inference models under limited or incomplete observations. However, they typically evaluate inferred paths against ground-truth data collected at the same time as the inference inputs or models, and therefore mainly emphasize instantaneous inference accuracy. **In practical deployment, path inference services are often used at a later time, when routing conditions may already have changed and the previously inferred paths may no longer remain fully representative of the current routing state.**

On the other hand, prior studies on routing stability do not present a fully uniform picture. Green et al. [18] reported substantial persistence of primary inter-domain paths over multi-month observations, whereas Bakhshaliyev et al. [19] showed that stable-path ratios decrease as the observation horizon expands from hours to months. Comarela et al. [20] further demonstrated that routing states evolve measurably over long timescales rather than remaining static. Taken together, these studies suggest that inter-domain paths may appear highly stable over certain observation horizons, yet exhibit noticeable temporal variation over others. However, these studies characterize routing behavior itself rather than the usable lifetime of path inference outputs. As a result, the literature still lacks a direct evaluation of what these mixed stability findings imply for the validity and usable lifetime of AS path inference results.

Motivated by this gap between AS path inference and routing-stability research, this paper aims to systematically answer the following question: **How do the performance and temporal drift of AS path inference approaches evolve over time?** In this sense, our work is positioned not as another inference method, but as an evaluation study on the temporal degradation and temporal drift of inferred AS paths. Answering this question is essential for understanding the reliability and practical applicability of path inference results in dynamic inter-domain routing environments.

In practical applications, users expect path inference services to return paths that are representative of the current network state, even though the underlying routing data may have been collected earlier. Therefore, some degree of performance degradation over time is inevitable, and understanding its extent is crucial because it informs two practical decisions: (1) the backward time range of the data that can be used for path inference and (2) the update frequency of path inference services. A fundamental limitation inherent in path inference methods is the lack of measurement data. Therefore, the temporal degradation pattern of path inference methods can help indicate how far back in time the data can be used for path inference. Path inference services are computationally intensive, so a trade-off between update frequency and computational resources is often necessary.

Despite the practical importance of temporal degradation in path inference, there is still limited understanding of how quickly stale inference results become outdated. Thus, in this paper, we conduct a quantitative validation study and a temporal drift analysis to examine the evolution of path inference performance over time. The quantitative validation study uses real-world data from a specific day and evaluates the inferred paths over the following eight weeks to quantify the performance degradation of path inference approaches over time. Thus, it helps to determine how far back in time the data can be used for current path inference. The temporal drift analysis further identifies when the inference results derived from the baseline snapshot are no longer representative of the current network state, thereby informing refresh scheduling decisions under the evaluated setting.

The contributions of this paper are as follows:

- The validation study reveals how representative path inference approaches degrade over time. The results show that the minimal performance degradation of the selected path inference approaches is only 2.09% in nearly two months. The maximum observed degradation is 7.44% in nearly two months, which occurs for ProbInfer in our evaluation, a data-driven path inference approach.
- The temporal drift analysis provides an operational comparison of refresh horizons across methods. Under the current decision rule, material drift is first detected on day 20 for KnownPath, on day 10 for GMPI, and on day 2 for ProbInfer. This comparison indicates that the three evaluated methods have different refresh horizons and temporal sensitivities under the evaluated setting.
- This research provides preliminary evidence on how historical data can be leveraged despite limited measurement resources and can inform refresh-frequency decisions for path inference services under computational constraints.

2 Related Work

2.1 Traceroute Methods

Traceroute is a commonly used tool for measuring the path between two hosts. The tool sends a series of packets to the destination, with each packet containing an increasing Time-to-Live (TTL) value. When a packet reaches an intermediate router, the router decrements the TTL value by one. When the TTL value reaches zero, the router discards the packet and sends an ICMP Time Exceeded message back to the source. The source then records the IP address of the router and the round-trip time of the packet. This process is repeated with increasing TTL values until the packet reaches the destination. The recorded IP addresses are used to obtain the path between the source and the destination.

In [21], the authors summarized the limitations (including “loops”, “cycles”, and “diamonds”) of traditional traceroute and proposed Paris Traceroute to provide more accurate path information. To extend path measurement to the reverse direction, revtr 2.0 [22] combines novel measurement techniques with a large-scale deployment and significantly improves throughput, accuracy, and coverage for Internet-scale reverse-path exploration. Recent studies continue to improve traceroute-based topology discovery and active probing efficiency, especially in IPv6 networks. 6Search [23] proposes a reinforcement learning-based traceroute approach for efficient IPv6 topology discovery, while TNet [24] further improves the efficiency of active IPv6 network discovery.

Though multiple traceroute methods have been proposed, these methods share the same limitation: the measurement resources are too limited to obtain the path information between any two ASes. On the other hand, mapping IPs to ASes is also a challenging task [25–28]. As a result, path inference approaches are proposed to address these challenges.

2.2 AS Path Inference

Heuristic-based path inference approaches. KnownPath [9] improves AS path inference by incorporating the valley-free principle [29] to constrain feasible routing paths. In iNano [11], in addition to constructing an Internet topology and adhering to the valley-free principle, the authors proposed the creation of a routing preference database. This database stores AS-level routing preferences in 3-tuples, such as (AS1, AS2 > AS3), which indicates that AS1 favors a path through AS2 over AS3 when both paths are of equal length. When inferring a path, iNano initially searches for the shortest paths that comply with the valley-free principle. It then selects paths that align with the routing preferences. In [30], a method called PredictRoute was introduced. This technique utilizes Traceroute data acquired through active detection to infer paths.

The researchers developed distinct directed acyclic graphs for each destination prefix and subsequently employed a destination-specific probabilistic Markov model to derive the inferred paths. In iPlane [10], segments of the measured Internet paths that converged at a common point (which could be an AS, a PoP, or a router) were stitched to infer paths between two target points. The HyperPath proposed in [31] also obtains paths between two ASes by concatenating segments of measured paths. This paper initially establishes that the structure of the AS-level network topology can be conceptualized as a tree, providing a theoretical foundation for stitching path segments to infer paths. Then, a heuristic algorithm is proposed to stitch path segments to infer paths between two ASes.

Data-driven path inference approaches. Due to the intricate nature of the Internet routing system, refined heuristic rules derived from collected measurement data often lack the accuracy needed to model it effectively. Heuristic methods frequently encounter issues with redundant paths. To tackle this challenge, researchers have put forth data-driven inference techniques. Sibyl [12] and ProbInfer [13] both involve stitching path segments to deduce inferred paths, although they employ distinct approaches to minimize redundancy. Sibyl employs a supervised machine learning model called RuleFit [32] to select the most optimal path. In contrast, ProbInfer introduces a probability model (decision tree) to mitigate redundancy. More recently, PathRadar [14] proposed a fine-grained AS path inference framework with a progressive learning process that applies different learning models to different AS categories. The process of Generative and Measurable Path Inference (GMPI) [15] utilizes heuristic algorithms to generate paths for an AP. Subsequently, a dual attention neural network is utilized to extract features from AS paths and estimate the likelihood of these generated paths. In [33], the authors proposed a method called RouteInfer, which infers AS paths by inferring the routing policies of ASes. First, a three-layer policy model is proposed to extract the routing policies for ASes, but it cannot obtain policies for all ASes. Hence, a learning-based approach is proposed to mitigate this limitation. Recent studies have also extended this line toward personalized and adaptive inference settings [16] and survey-style syntheses of inter-domain path inference methods and applications [17]. Taken together, these studies mainly focus on how to construct more accurate or more flexible inference models. In contrast, the present work focuses on how the validity of inferred paths evolves as the routing system changes over time.

2.3 AS Relationship Inference

AS relationship inference is closely related to AS path inference because inferred business relationships provide key constraints and signals for path generation, valley-free filtering, and routing-preference modeling. Gao [29] established the classic relationship model underlying many later inference and routing studies. More recent work has focused on improving relationship inference under incomplete and biased observations. ProbLink [34] improves inference stability and practicality on hard links, while TopoScope [35] recovers AS relationships from fragmentary observations. Prehn and Feldmann [36] further show that available validation data can be substantially biased, which directly affects the evaluation of relationship inference algorithms. Recent learning-based extensions include multiclass relationship inference with graph convolutional networks [37] and HELA [38], which combines empirical and learning-based components to improve accuracy and stability. These studies provide important support for path inference, but they mainly aim to infer the relationships themselves rather than evaluate how path inference outputs degrade over time.

2.4 Mixed Evidence on Internet AS-Level Path Stability

Prior studies on Internet AS-level path stability do not present a fully uniform picture. Green et al. [18] analyze multi-month BGP observations and show that primary inter-domain paths can remain highly persistent over time. In contrast, Bakhshaliyev et al. [19] report that the stable-path ratio decreases from 97%

over one hour to 89% over one day, 76% over one week, 54% over one month, and 44% over two months. Comarela et al. [20] further show that routing states evolve measurably over long timescales rather than remaining static. Taken together, these studies suggest that Internet AS-level paths may appear highly stable over some observation horizons, yet exhibit noticeable temporal variation over others. However, these studies characterize routing behavior itself rather than the usable lifetime of path inference outputs. Therefore, the literature still lacks a direct evaluation of how temporal routing dynamics affect the validity of inferred AS paths. This mixed picture is exactly why it is necessary to conduct a validation study on the temporal degradation and temporal drift of path inference approaches.

3 Validation Methodology

To quantify the temporal degradation and drift of path inference approaches, we conduct a validation study using real-world data. The validation process uses data from a specific day to evaluate inferred paths over the following eight weeks. The temporal behavior of path inference approaches is characterized by both performance degradation over time and a temporal drift analysis of stale inference results.

3.1 Preliminaries

As defined in [15], AS path inference focuses on identifying a feasible AS path ($s \rightarrow d$) satisfying underlying and unrevealed routing policies for any unmeasured AS pair (AP, two different ASes) $Q = (s, d)$. For ease of understanding, we first introduce some basic concepts and notations in path inference.

- *AS Pair (AP)*: An AP is a pair of two different ASes, denoted as $Q = (s, d)$, where s is the traffic source AS and d is the destination. If the path between s and d is known, it is called a measured AP; otherwise, it is called an unmeasured AP. The AP of an AS path refers to the AP composed of its first and last ASes.
- *Starting and ending AS*: The starting AS of an AP (AS path) is the first AS of the AP (AS path), and the ending AS is the last AS of the AP (AS path).

3.2 Input Data

Three types of data are used in the validation study: AS paths derived from BGP routing tables, AS relationship data, and AS Rank data.

3.2.1 AS Paths Derived from BGP Routing Tables

The routing tables are collected from the Route Views project [39] and the RIPE RIS project [40], two widely used public platforms for BGP data collection. These platforms collect routing information through peering sessions with external networks and provide routing-table snapshots for research use. Route Views offers data from multiple route collectors, while RIPE RIS operates a globally distributed set of Remote Route Collectors, many of which are located at Internet Exchange Points.

In this paper, we use 20 routing-table snapshots collected at 8 a.m. on selected dates from October to November 2023, as summarized in Table 1. The fixed collection time is used to reduce diurnal effects and ensure comparability across snapshots. From these routing tables, we extract AS paths and use them to construct the validation dataset for each snapshot. Following common practice in prior work, the extracted AS paths are treated as ground-truth routing outcomes observed at the control plane.

At the same time, BGP routing tables on any given day may still contain localized abnormal events or temporary policy shifts, and 1 October 2023 is not necessarily an exception. For this reason, the present study should be understood as a large-scale aggregated analysis rather than as evidence that a single snapshot is universally representative. When the evaluation covers a very large AP set and multiple subsequent

snapshots, limited local fluctuations are more likely to introduce noise or mild bias than to invalidate the overall temporal trend, unless they affect a large proportion of the observed routing state. In our manual review of publicly reported Internet operational conditions around 1 October 2023, we did not identify evidence of a large-scale Internet-wide disruption on that date.

Table 1: Statistics of the evaluation data.

Snapshot	Type of AP	# AP	# Starting AS	# Ending AS
20231001	Measured	16824494	1168	70665
	Unmeasured	6289556	1031	67402
20231002	Unmeasured	6289556	1031	67402
20231003	Unmeasured	6289556	1031	67402
20231004	Unmeasured	6289556	1031	67402
20231005	Unmeasured	6289556	1031	67402
20231006	Unmeasured	6289556	1031	67402
20231007	Unmeasured	6289556	1031	67402
20231008	Unmeasured	6289556	1031	67402
20231009	Unmeasured	6289556	1031	67402
20231010	Unmeasured	6289556	1031	67402
20231011	Unmeasured	6289556	1031	67402
20231012	Unmeasured	6289556	1031	67402
20231013	Unmeasured	6289556	1031	67402
20231014	Unmeasured	6289556	1031	67402
20231021	Unmeasured	6289556	1031	67402
20231028	Unmeasured	6289556	1031	67402
20231104	Unmeasured	6289556	931	67402
20231111	Unmeasured	6289556	931	67402
20231118	Unmeasured	6289556	931	67402
20231125	Unmeasured	6289556	931	67402

To assess path inference approaches, both unmeasured APs and their corresponding ground-truth paths are required. In the snapshot of 20231001, the collected AS paths are divided into training and testing sets using a 70%–30% split. The split is performed randomly at the AP level to avoid bias toward specific AS pairs. Seventy percent of the APs and their paths are used as input to the path inference approaches, while the remaining 30% are used for evaluation.

As this study focuses on temporal behavior across multiple snapshots, we further construct a common evaluation set that remains observable throughout the full observation window. The evaluation data is obtained as follows:

- The APs corresponding to the testing set in snapshot 20231001 are treated as initially unmeasured APs, denoted as U' .
- The AP sets of the remaining 19 snapshots are denoted as U_1 to U_{19} . We compute the intersection $U = U' \cap U_1 \cap \dots \cap U_{19}$ to obtain the final set of unmeasured APs that remain observable across all snapshots, thereby ensuring temporal consistency of the evaluation set.
- For each snapshot, the AS paths corresponding to these APs are used as ground-truth data for both quantitative validation and temporal drift analysis.

As shown in [Table 1](#), the number of starting ASes is much smaller than the number of ending ASes. This is a direct consequence of the collector-based observation model: the starting ASes are constrained by the locations and peering relationships of the collectors and their connected vantage points, whereas the ending ASes span a much broader destination space.

3.2.2 AS Relationship Data

We use the CAIDA AS relationship dataset dated 1 October 2023 [41,42]. According to CAIDA, this dataset is inferred primarily from publicly available BGP data and captures inter-AS business relationships such as customer-provider and peer-peer links.

AS relationships are commonly classified into customer-to-provider, peer-to-peer, and sibling-to-sibling types [43]. Based on these relationships, Gao proposed the valley-free principle, which states that AS-level routing paths typically follow policy-compliant patterns [29]. In this study, the inferred AS relationships provide policy-related constraints for path inference, particularly for methods that rely on valley-free assumptions or relationship-aware path construction.

3.2.3 AS Rank Data

We use CAIDA's AS Rank dataset dated 1 October 2023 [44]. AS Rank orders ASes primarily based on customer-cone size, with additional attributes such as AS degree, country, and organization information. In this paper, these attributes are used in GMPI to characterize the relative position and structural importance of ASes, supporting both path generation and AS representation learning.

3.3 Selected Path Inference Approaches

Three approaches are selected to conduct the validation study: GMPI, KnownPath, and ProbInfer. GMPI, proposed in [15], represents the state-of-the-art path inference approach. On the other hand, KnownPath, introduced in [9], is a classic path inference approach.

GMPI: GMPI is the state-of-the-art path inference approach, which involves a generative and measurable path inference process. GMPI first generates paths for an AP using a heuristic algorithm. Then, some AS paths related to the AP are obtained from the collected data to provide latent routing preferences of the AP. Finally, a dual attention neural network is utilized to extract features from AS paths and estimate the likelihood of these generated paths. After that, the path with the highest likelihood is selected as the inferred path.

GMPI uses the BGP routing tables, the AS relationship data, and the AS Rank data to accomplish the path inference. In this paper, the parameters of GMPI are set to the same values as in the previous work [15].

KnownPath: It is a classic and widely used path inference approach, which is derived from the Bellman-Ford algorithm. For a destination prefix p and a set of ASes $A = \{a_1, \dots, a_n\}$, suppose that the path between a_i and p is known. Let $A' = \{a'_1, \dots, a'_m\}$ be the set of ASes for which the path between a'_i and p is unknown. The KnownPath approach first constructs an AS-level Internet topology using BGP data. Then, it expands

the paths between ASes in A and p to obtain the paths between ASes in A' and p based on the constructed topology and the AS relationships.

KnownPath uses the BGP routing tables and the AS relationship data to accomplish the path inference.

ProbInfer: ProbInfer is a data-driven path inference approach, which obtains inferred paths by stitching path segments and reducing redundancy with a decision-tree model. ProbInfer uses the BGP routing tables and the AS relationship data to accomplish the path inference.

3.4 Metrics

3.4.1 Metrics for Quantitative Validation

For an AP, the inference accuracy of its inferred path is defined as the Jaccard similarity between the inferred path P^{inf} and the ground-truth path P^{gt} . The Jaccard similarity is calculated as follows:

$$\text{Jaccard}(P^{\text{inf}}, P^{\text{gt}}) = \frac{|P^{\text{inf}} \cap P^{\text{gt}}|}{|P^{\text{inf}} \cup P^{\text{gt}}|} \quad (1)$$

where P^{inf} and P^{gt} denote the sets of ASes in the inferred and ground-truth paths, respectively.

Three metrics are used to measure the performance of path inference approaches in quantitative validation: *upper bound accuracy (UBA)*, *average accuracy (AA)*, and *exact same ratio (ESR)*.

Upper Bound Accuracy (UBA) measures the best achievable accuracy when multiple candidate paths are generated. For an AP, given k candidate paths $\{P_{i,1}^{\text{inf}}, \dots, P_{i,k}^{\text{inf}}\}$, UBA is defined as the maximum Jaccard similarity among these candidates. This metric reflects the upper bound performance of the path generation process.

Average Accuracy (AA) is defined as the average Jaccard similarity across all APs in a snapshot:

$$\text{AA} = \frac{1}{|U|} \sum_{i \in U} \text{Jaccard}(P_i^{\text{inf}}, P_i^{\text{gt}}) \quad (2)$$

where U denotes the set of APs in the evaluation dataset.

Exact Same Ratio (ESR) is defined as the fraction of APs whose inferred paths exactly match the ground-truth paths:

$$\text{ESR} = \frac{|\{i \in U \mid P_i^{\text{inf}} = P_i^{\text{gt}}\}|}{|U|} \quad (3)$$

3.4.2 Metrics for Temporal Drift Analysis

To characterize how inference results deviate over time, we analyze temporal drift relative to a baseline snapshot t_0 . For each subsequent snapshot t_x , we compute the following three complementary metrics based on the distribution of Jaccard similarities.

Mean Shift measures the directional change in average similarity relative to the baseline:

$$\Delta_\mu(t_x) = \bar{S}_{t_x} - \bar{S}_{t_0} \quad (4)$$

Cohen's d measures the standardized effect size of this change:

$$d(t_x) = \frac{\bar{S}_{t_x} - \bar{S}_{t_0}}{s_p(t_x)}, \quad s_p(t_x) = \sqrt{\frac{(n-1)\sigma_{t_0}^2 + (n-1)\sigma_{t_x}^2}{2n-2}} \quad (5)$$

First-order Wasserstein Distance measures the overall distributional difference between similarity distributions at t_0 and t_x .

These three metrics are reported together because they capture complementary aspects of temporal drift: mean shift reflects absolute change, Cohen's d captures standardized effect size, and Wasserstein distance characterizes global distributional deviation.

3.5 Quantitative Analysis of Temporal Degradation

To observe the daily and weekly performance degradation of the path inference approaches, a quantitative analysis is conducted. With the data from 1 October 2023, the paths between unmeasured APs are inferred for all 20 snapshots and validated using the ground-truth paths from each snapshot. The *AA* and *ESR* of the path inference approaches are calculated for each snapshot. For GMPI, the *UBA* of the path generation process is also observed.

We intentionally use a fixed baseline snapshot in this analysis. This design isolates how inference results derived from one reference state degrade as the network evolves over time, without introducing additional variation from changing the training snapshot itself. It also reflects a practical deployment scenario in which path inference results or models trained at time t_0 are reused for some period before being refreshed.

Accordingly, the reported degradation curves should be interpreted as fixed-baseline temporal behavior under the selected baseline setting, rather than as a claim that the same degradation pattern must hold for every possible starting date. The purpose of this section is to characterize how quickly performance decays when the baseline is held constant, whereas the question of whether the same trend is invariant across multiple baseline dates is a separate robustness issue discussed in [Section 5](#).

3.6 Temporal Drift Analysis of Path Inference

To understand how long path inference results derived from a baseline snapshot remain valid, we analyze the temporal drift of path similarity over time. Unlike conventional hypothesis testing that focuses on whether two samples are statistically different, our goal is to determine after how many days the inference results generated from a baseline snapshot t_0 become no longer representative of the current network state.

This formulation aligns with the notion of concept drift in data-driven systems, where previously learned knowledge may become outdated due to distributional changes in the underlying system.

For each AP i , let $p_i^{(t_0)}$ denote the path inferred using data collected at time t_0 , and let $g_i^{(t_x)}$ denote the ground-truth path observed at time t_x . The inference accuracy at time t_x is defined as:

$$S_i^{(t_x)} = J(p_i^{(t_0)}, g_i^{(t_x)}), \quad (6)$$

where $J(\cdot, \cdot)$ denotes the Jaccard similarity defined in [Section 3.4.1](#). The distribution of $\{S_i^{(t_x)}\}$ characterizes the validity of stale inference results at time t_x .

Drift quantification. To quantify the difference between the baseline distribution $S^{(t_0)}$ and the distribution at time t_x , we use three complementary metrics defined in [Section 3.4.2](#): mean shift, Cohen's d , and the first-order Wasserstein distance.

Let $S^{(t_0)} = \{S_i^{(t_0)}\}_{i=1}^n$ and $S^{(t_x)} = \{S_i^{(t_x)}\}_{i=1}^n$ be the sampled similarity sets for a repeated run, where n is the sample size. Let \bar{S}_t and σ_t denote the sample mean and standard deviation at time t , respectively.

These metrics capture complementary aspects of temporal drift: mean shift reflects absolute directional change, Cohen's d captures standardized effect size, and Wasserstein distance quantifies global distributional differences beyond mean-based statistics.

From statistical difference to material drift. Given the large number of APs, even small differences may appear consistently across repeated samples. Therefore, instead of relying solely on statistical significance, we jointly consider both the uncertainty of the mean shift and the magnitude of distributional drift.

We say that the stale inference results at time t_x become *materially different* from those at t_0 when both:

- the 95% empirical interval of the mean shift over repeated samples excludes zero, and
- the Wasserstein distance exceeds a practical threshold.

Expiration time. We define the expiration time T^* of inference results derived from t_0 as:

$$T^* = \min\{t_x : S^{(t_x)} \text{ is materially different from } S^{(t_0)}\}. \quad (7)$$

This time point provides an operational estimate, under the current decision rule, of when stale path inference results are no longer representative of the current network state.

Implementation details. In our experiments, each snapshot contains more than 10^6 APs. For computational efficiency and to reflect realistic query scenarios, we repeatedly sample 5000 APs from the common AP set between the baseline and target snapshots. This sampling process is repeated 50 times for each target snapshot.

We report the mean and 95% empirical interval of the mean shift over repeated samples, together with the average Cohen's d and Wasserstein distance across repeated samples. We use 0.01 as a practical Wasserstein cutoff to distinguish very small distributional changes from operationally meaningful drift in this study, rather than as a universal statistical constant. A snapshot is considered to exhibit material drift when the 95% empirical interval of the mean shift excludes zero and the average Wasserstein distance exceeds 0.01. While different practical cutoffs may change the exact estimated expiration day, they do not change the overall ordering of the three methods in our results.

4 Validation Results

The validation process is as follows: the training dataset (70% of the AS paths, the AS relationship data, and the AS Rank data) obtained from 1 October 2023 is used to infer paths for the unmeasured APs for all 20 snapshots. For GMPI, we observe not only the final path inference performance but also the path generation performance.

We use the same parameter settings for GMPI as in the original paper [15] to ensure that the observed temporal behavior reflects the method's performance under its intended configuration: the input/output layer size is set to $d = 512$, the learning rate is set to $lr = 0.00001$, dropout is set to 0.1 in the position encoding layer, and the maximum number of observed positive paths is set to $m = 200$. KnownPath and ProbInfer are also configured as in the original paper.

4.1 Quantitative Analysis Results

The performance of the three path inference approaches over time is shown in [Tables 2–4](#). From the tables, ProbInfer shows the largest performance degradation, followed by GMPI, while KnownPath exhibits the smallest degradation over the examined period.

To further observe the temporal behavior of the path generation process in GMPI, the *UBA* and *ESR* of the path generation process are calculated, as shown in [Figs. 1 and 2](#). The degradation of the path generation process is relatively small, suggesting that both the path generation and path selection components contribute to the overall temporal degradation of GMPI.

Table 2: Performance degradation of GMPI.

Date	Accuracy	ESR	Accuracy Degradation	ESR Degradation	Accuracy Cumulative Degradation	ESR Cumulative Degradation
20231001	85.91%	60.10%				
20231002	85.55%	60.03%	0.36%	0.07%		
20231003	85.34%	59.58%	0.21%	0.45%		
20231004	85.26%	59.47%	0.08%	0.11%		
20231005	85.16%	59.25%	0.10%	0.22%		
20231006	85.03%	58.98%	0.13%	0.27%		
20231007	84.88%	58.67%	0.15%	0.31%	1.03%	1.47%
20231008	85.01%	58.95%	-0.13%	-0.28%		
20231009	84.92%	58.72%	0.09%	0.23%		
20231010	84.79%	58.54%	0.13%	0.18%		
20231011	84.61%	58.17%	0.18%	0.37%		
20231012	84.69%	58.34%	-0.08%	-0.17%		
20231013	84.70%	58.39%	-0.01%	-0.05%		
20231014	84.70%	58.34%	0.00%	0.05%	1.21%	1.76%
20231021	83.73%	56.44%	0.97%	1.90%	2.18%	3.66%
20231028	83.41%	55.77%	0.32%	0.67%	2.50%	4.33%
20231104	83.34%	55.11%	0.07%	0.66%	2.57%	5.00%
20231111	83.22%	54.87%	0.12%	0.24%	2.69%	5.24%
20231118	82.89%	54.32%	0.33%	0.55%	3.02%	5.79%
20231125	82.67%	53.85%	0.22%	0.47%	3.24%	6.25%
Daily Degradation			0.06%	0.11%		

Table 3: Performance degradation of KnownPath.

Date	Accuracy	ESR	Accuracy Degradation	ESR Degradation	Accuracy Cumulative Degradation	ESR Cumulative Degradation
20231001	81.23%	56.94%				
20231002	81.13%	56.73%	0.10%	0.21%		
20231003	80.97%	56.44%	0.16%	0.29%		

(Continued)

Table 3 (continued)

Date	Accuracy	ESR	Accuracy Degradation	ESR Degradation	Accuracy Cumulative Degradation	ESR Cumulative Degradation
20231004	80.96%	56.42%	0.01%	0.02%		
20231005	80.87%	56.21%	0.09%	0.21%		
20231006	80.73%	55.89%	0.14%	0.32%		
20231007	80.54%	55.47%	0.19%	0.42%	0.69%	1.47%
20231008	80.67%	55.80%	-0.13%	-0.33%		
20231009	80.62%	55.65%	0.05%	0.15%		
20231010	80.47%	55.39%	0.15%	0.26%		
20231011	80.26%	54.97%	0.21%	0.42%		
20231012	80.34%	55.14%	-0.08%	-0.17%		
20231013	80.43%	55.34%	-0.09%	-0.20%		
20231014	80.42%	55.30%	0.01%	0.04%	0.81%	1.64%
20231021	79.75%	53.85%	0.67%	1.45%	1.48%	3.09%
20231028	79.53%	53.54%	0.22%	0.31%	1.70%	3.40%
20231104	79.36%	53.42%	0.17%	0.12%	1.87%	3.52%
20231111	79.14%	52.96%	0.22%	0.46%	2.09%	3.98%
20231118	78.92%	52.55%	0.22%	0.41%	2.31%	4.43%
20231125	78.82%	52.23%	0.10%	0.32%	2.41%	4.75%
Daily Degradation			0.04%	0.08%		

Table 4: Performance degradation of ProbInfer.

Date	Accuracy	ESR	Accuracy Degradation	ESR Degradation	Accuracy Cumulative Degradation	ESR Cumulative Degradation
20231001	64.09%	31.06%				
20231002	64.06%	30.97%	0.03%	0.09%		
20231003	61.90%	27.12%	2.16%	3.85%		
20231004	63.07%	29.34%	-1.17%	-2.78%		
20231005	62.67%	28.50%	0.40%	0.84%		

(Continued)

Table 4 (continued)

Date	Accuracy	ESR	Accuracy Degradation	ESR Degradation	Accuracy Cumulative Degradation	ESR Cumulative Degradation
20231006	61.57%	26.78%	1.10%	1.72%		
20231007	61.51%	26.63%	0.06%	0.15%	2.58%	4.43%
20231008	61.57%	26.85%	-0.06%	-0.22%		
20231009	63.22%	29.96%	1.65%	3.11%		
20231010	61.60%	26.99%	1.62%	3.97%		
20231011	61.31%	26.53%	0.29%	0.46%		
20231012	61.40%	26.70%	-0.09%	-0.17%		
20231013	61.51%	26.96%	-0.11%	-0.26%		
20231014	61.28%	26.48%	0.23%	0.48%	2.81%	4.58%
20231021	60.77%	26.06%			3.32%	5.00%
20231028	61.84%	28.74%			2.25%	3.32%
20231104	57.18%	23.83%			6.91%	7.23%
20231111	57.09%	23.88%			7.00%	7.47%
20231118	56.76%	23.37%			7.33%	7.97%
20231125	56.65%	23.23%			7.44%	8.23%
Daily Degradation			0.13%	0.15%		

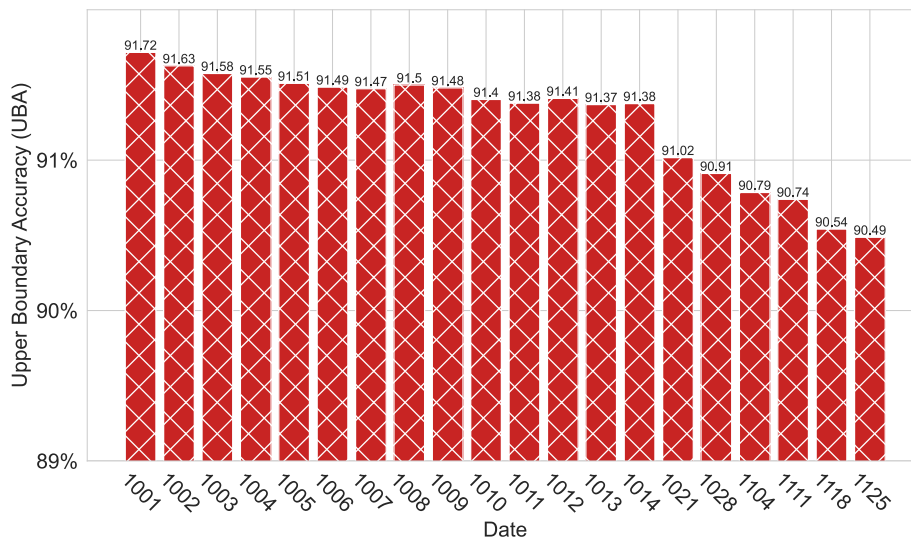


Figure 1: Average UBA of GMPT's path generation process across snapshots.

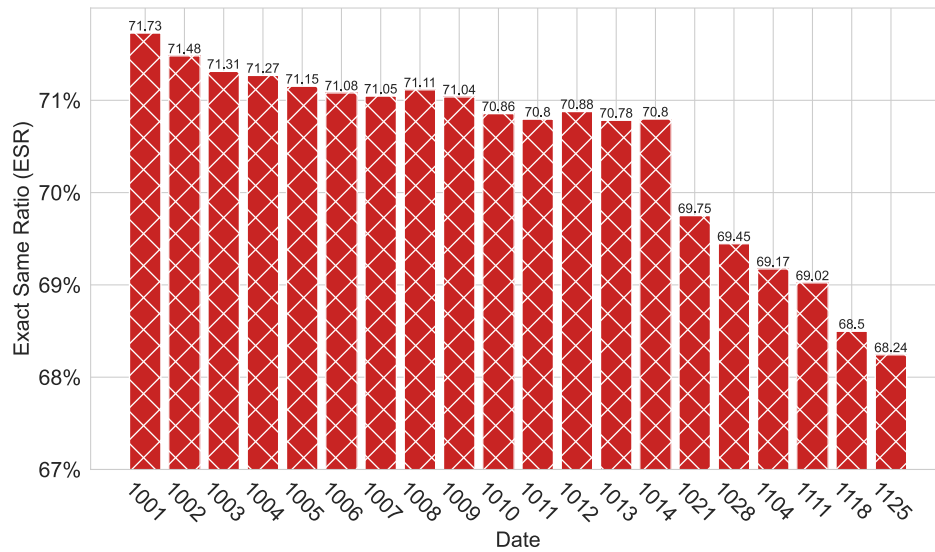


Figure 2: ESR of GMPI's path generation process across snapshots.

KnownPath: The validation results of KnownPath are shown in Table 3. KnownPath has the smallest performance degradation among the three methods, with a decrease of 2.09% and 4.47% in accuracy and ESR over eight weeks, respectively. This relatively small decrease suggests that many AS-level paths remain stable over the examined period.

ProbInfer: The validation results of ProbInfer are shown in Table 4. ProbInfer has the most significant performance degradation among the three methods, with a decrease of 7.44% and 8.23% in accuracy and ESR over eight weeks, respectively. This observation indicates that ProbInfer is less temporally robust than the other two evaluated methods in our dataset.

In the quantitative analysis, several snapshots show slight local rebounds relative to the immediately previous date, such as 20231008, 20231012, and 20231013. Similar local fluctuations also appear for ProbInfer on 20231028 and for KnownPath on 20231104. These observations suggest that the degradation process is not strictly monotonic at every snapshot. However, the current evidence is not sufficient to infer any broader recurring Internet-wide routing pattern. Their underlying drivers require dedicated longitudinal analysis and remain part of our future work.

The above results indicate that the performance of path inference approaches changes over time. In the current evaluation, KnownPath is the most temporally robust of the three methods, GMPI is intermediate, and ProbInfer exhibits the largest degradation.

4.2 Temporal Drift Analysis Result

The temporal drift analysis provides an operational view of when stale inference results cease to be representative of the current network state. The results are shown in Tables 5–7.

Table 5: Temporal drift analysis of GMPI relative to the baseline snapshot (1 October 2023).

Timescale (day)	Snapshot	Mean Shift	95% Empirical Interval of Mean Shift	Cohen's <i>d</i>	Wasserstein Distance	Conclusion
1	20231002	-0.25%	[-0.62%, 0.08%]	-0.013	0.0039	Not material
2	20231003	-0.42%	[-0.78%, 0.06%]	-0.021	0.0049	Not material
3	20231004	-0.55%	[-0.94%, -0.25%]	-0.028	0.0058	Not material
4	20231005	-0.60%	[-0.95%, -0.25%]	-0.030	0.0061	Not material
5	20231006	-0.74%	[-1.13%, -0.34%]	-0.037	0.0074	Not material
6	20231007	-0.87%	[-1.26%, -0.33%]	-0.044	0.0087	Not material
7	20231008	-0.74%	[-1.15%, -0.27%]	-0.037	0.0075	Not material
8	20231009	-0.82%	[-1.16%, -0.38%]	-0.041	0.0084	Not material
9	20231010	-0.94%	[-1.30%, -0.54%]	-0.047	0.0095	Not material
10	20231011	-1.16%	[-1.58%, -0.71%]	-0.058	0.0116	Material drift
11	20231012	-1.12%	[-1.59%, -0.78%]	-0.056	0.0112	Material drift
12	20231013	-1.07%	[-1.37%, -0.71%]	-0.054	0.0108	Material drift
13	20231014	-1.11%	[-1.43%, -0.72%]	-0.055	0.0111	Material drift
20	20231021	-2.04%	[-2.50%, -1.61%]	-0.101	0.0204	Material drift
27	20231028	-2.30%	[-2.79%, -1.91%]	-0.113	0.0230	Material drift
34	20231104	-2.58%	[-2.94%, -2.20%]	-0.127	0.0258	Material drift
41	20231111	-2.73%	[-3.07%, -2.40%]	-0.134	0.0273	Material drift
48	20231118	-3.05%	[-3.35%, -2.75%]	-0.149	0.0305	Material drift
55	20231125	-3.22%	[-3.54%, -2.96%]	-0.157	0.0322	Material drift

Table 6: Temporal drift analysis of KnownPath relative to the baseline snapshot (1 October 2023).

Timescale (day)	Snapshot	Mean Shift	95% Empirical Interval of Mean Shift	Cohen's <i>d</i>	Wasserstein Distance	Conclusion
1	20231002	-0.10%	[-0.18%, -0.02%]	-0.004	0.0010	Not material
2	20231003	-0.25%	[-0.36%, -0.17%]	-0.010	0.0025	Not material
3	20231004	-0.25%	[-0.37%, -0.08%]	-0.010	0.0025	Not material
4	20231005	-0.35%	[-0.49%, -0.25%]	-0.015	0.0036	Not material
5	20231006	-0.46%	[-0.64%, -0.24%]	-0.019	0.0046	Not material

(Continued)

Table 6 (continued)

Timescale (day)	Snapshot	Mean Shift	95% Empirical Interval of Mean Shift	Cohen's d	Wasserstein Distance	Conclusion
6	20231007	-0.67%	[-0.81%, -0.50%]	-0.028	0.0067	Not material
7	20231008	-0.56%	[-0.75%, -0.38%]	-0.023	0.0056	Not material
8	20231009	-0.65%	[-0.81%, -0.48%]	-0.027	0.0065	Not material
9	20231010	-0.75%	[-0.90%, -0.49%]	-0.031	0.0075	Not material
10	20231011	-0.94%	[-1.14%, -0.77%]	-0.039	0.0095	Not material
11	20231012	-0.93%	[-1.14%, -0.69%]	-0.038	0.0093	Not material
12	20231013	-0.76%	[-0.98%, -0.53%]	-0.031	0.0076	Not material
13	20231014	-0.80%	[-0.98%, -0.58%]	-0.033	0.0080	Not material
20	20231021	-1.46%	[-1.70%, -1.25%]	-0.060	0.0146	Material drift
27	20231028	-1.73%	[-2.07%, -1.35%]	-0.071	0.0173	Material drift
34	20231104	-1.92%	[-2.40%, -1.60%]	-0.079	0.0192	Material drift
41	20231111	-2.16%	[-2.50%, -1.76%]	-0.088	0.0216	Material drift
48	20231118	-2.32%	[-2.58%, -2.02%]	-0.095	0.0233	Material drift
55	20231125	-2.39%	[-2.85%, -1.92%]	-0.097	0.0239	Material drift

Table 7: Temporal drift analysis of ProbInfer relative to the baseline snapshot (1 October 2023).

Timescale (day)	Snapshot	Mean Shift	95% Empirical Interval of Mean Shift	Cohen's d	Wasserstein Distance	Conclusion
1	20231002	-0.05%	[-0.43%, 0.51%]	-0.002	0.0027	Not material
2	20231003	-1.92%	[-2.27%, -1.62%]	-0.081	0.0193	Material drift
3	20231004	-0.78%	[-1.24%, -0.15%]	-0.033	0.0081	Not material
4	20231005	-1.34%	[-1.73%, -1.04%]	-0.056	0.0134	Material drift
5	20231006	-2.21%	[-2.57%, -1.81%]	-0.092	0.0221	Material drift
6	20231007	-2.23%	[-2.80%, -1.67%]	-0.093	0.0223	Material drift
7	20231008	-2.13%	[-2.74%, -1.61%]	-0.089	0.0213	Material drift
8	20231009	-0.65%	[-1.05%, -0.11%]	-0.027	0.0066	Not material
9	20231010	-2.07%	[-2.44%, -1.55%]	-0.087	0.0208	Material drift
10	20231011	-2.36%	[-2.80%, -1.77%]	-0.099	0.0237	Material drift

(Continued)

Table 7 (continued)

Timescale (day)	Snapshot	Mean Shift	95% Empirical Interval of Mean Shift	Cohen's d	Wasserstein Distance	Conclusion
11	20231012	-2.26%	[-2.61%, -1.89%]	-0.094	0.0226	Material drift
12	20231013	-2.09%	[-2.63%, -1.61%]	-0.087	0.0209	Material drift
13	20231014	-2.34%	[-2.73%, -1.83%]	-0.098	0.0234	Material drift
20	20231021	-2.85%	[-3.46%, -2.31%]	-0.119	0.0285	Material drift
27	20231028	-1.56%	[-2.06%, -0.82%]	-0.065	0.0156	Material drift
34	20231104	-3.42%	[-4.03%, -2.84%]	-0.143	0.0342	Material drift
41	20231111	-3.41%	[-4.12%, -2.81%]	-0.142	0.0341	Material drift
48	20231118	-3.79%	[-4.42%, -3.25%]	-0.158	0.0379	Material drift
55	20231125	-3.86%	[-4.51%, -3.30%]	-0.161	0.0386	Material drift

Among the three approaches, KnownPath exhibits the slowest temporal drift, GMPI shows a moderate drift rate, and ProbInfer drifts the fastest. Under the current decision rule, material drift is first detected on day 20 for KnownPath, on day 10 for GMPI, and on day 2 for ProbInfer. This ordering is consistent with the quantitative analysis, where KnownPath shows the smallest long-term degradation and ProbInfer shows the largest.

The temporal drift pattern of ProbInfer is also less monotonic than those of GMPI and KnownPath. For example, its drift is material on day 2, not material on day 3 under the current threshold, and becomes material again in subsequent snapshots.

5 Discussion

While this study provides a systematic evaluation of the temporal degradation and drift of path inference approaches, several limitations and potential sources of bias should be considered when interpreting the results.

Limited data and observation bias: Our analysis relies on BGP routing-table data from Route Views and RIPE RIS, which provide control-plane observations from a limited set of vantage points (VPs). As a result, the observed AS paths may not fully reflect data-plane forwarding behavior, and the coverage of routing dynamics depends on the spatial distribution of VPs. This limitation may lead to incomplete visibility of certain routing changes, particularly in less well-observed regions of the Internet.

Evaluation set construction bias: To ensure temporal consistency, we construct a common AP set by intersecting APs across all snapshots. While this design enables controlled temporal comparison, it may introduce a bias toward APs that remain consistently observable over time. Such APs are more likely to correspond to relatively stable routing scenarios, which may lead to a conservative estimation of temporal drift in more dynamic parts of the network.

Single-baseline sensitivity: The current validation uses 1 October 2023 as the only baseline snapshot. In operational Internet routing data, limited abnormal events or temporary policy shifts may occur on many dates, and the selected baseline date is unlikely to be completely free of such local fluctuations. However,

because our evaluation is performed on a very large AP set and the degradation curves are observed across multiple subsequent snapshots, limited local anomalies are more likely to act as noise or mild bias than to dominate the aggregate trend, unless they affect a substantial portion of the routing state. Our manual review of publicly reported Internet operational conditions did not identify evidence of a large-scale Internet-wide abnormal event on 1 October 2023. Nevertheless, using a single baseline date still restricts the robustness and generalizability of the conclusions, because it does not test whether the same degradation pattern would be reproduced under other starting dates. A stronger design would repeat the analysis from multiple independent baseline snapshots and compare the resulting curves directly.

Limited method coverage: The evaluation focuses on three representative methods, namely GMPI, KnownPath, and ProbInfer. While these methods span several important inference styles, the conclusions may not fully generalize to approaches that rely heavily on active measurements or other fundamentally different inference paradigms.

Limited mechanism-level interpretation: The current study evaluates temporal degradation and drift at the method level, but it does not include a dedicated structural analysis of failed paths. Therefore, the present results do not support causal attribution of the observed degradation pattern to any specific internal mechanism of ProbInfer. Future work should analyze failed-path structures directly to examine such mechanism-level explanations.

Granularity limitations: This study focuses on AS-level path inference, which provides a coarse-grained abstraction of inter-domain routing. Such abstraction does not capture finer-grained routing dynamics at the prefix level, PoP level, or IP level, where routing decisions may differ due to traffic engineering, load balancing, or intra-AS policies. As a result, temporal variations observed at finer granularities may not be fully reflected in our evaluation.

Despite these limitations, the consistent temporal trends observed across large-scale datasets and multiple complementary metrics suggest that our findings provide a meaningful characterization of the temporal degradation and drift of path inference approaches under the evaluated setting. These limitations also highlight important directions for future work, including incorporating data-plane measurements, expanding method coverage, and exploring finer-grained path representations.

6 Conclusion

In this paper, we explore the temporal degradation and drift of path inference approaches through eight weeks of real Internet path data, three representative path inference methods, and quantitative analysis and temporal drift analysis. This provides preliminary evidence that can inform the deployment and update of path inference services under the evaluated setting.

Our experiments show that path inference methods do degrade over time, but the magnitude of degradation differs across methods and remains measurable rather than severe over the examined eight-week period. The temporal drift analysis further provides an operational reference for refresh scheduling by indicating when stale inference results are no longer sufficiently representative of the current network state under the evaluated setting.

Acknowledgement: The authors would like to express their gratitude to the editors and reviewers for their detailed review and insightful advice.

Funding Statement: This work is supported by the National Natural Science Foundation of China (No. 62472434), the Key Program of NSFC Hunan (2026JJ30028), and the China Postdoctoral Science Foundation (2023TQ0089).

Author Contributions: Xionglve Li conceived the study, designed the methodology, implemented the experiments, analyzed the data, and drafted the manuscript. Changsheng Hou contributed to the study design, supervised the research, and revised the manuscript. Yuzhou Huang and Zhenyu Qiu contributed to data preparation, experiment execution, and result validation. Gang Hu, Bingnan Hou, and Wei Dong contributed to the interpretation of the results and manuscript revision. Zhiping Cai contributed to the overall research design, supervision, and final revision of the manuscript. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: Not applicable.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Banerjee R, Razaghpahan A, Chiang L, Mishra A, Sekar V, Choi Y, et al. Internet outages, the eyewitness accounts: analysis of the outages mailing list. In: *Passive and active measurement*. Cham, Switzerland: Springer International Publishing; 2015. p. 206–19. doi:10.1007/978-3-319-15509-8_16.
2. Lutu A, Bagnulo M, Pelsser C, Maennel O, Cid-Sueiro J. The BGP visibility toolkit: detecting anomalous Internet routing behavior. *IEEE/ACM Trans Netw*. 2016;24(2):1237–50. doi:10.1109/tnet.2015.2413838.
3. Sundaresan S, Deng X, Feng Y, Lee D, Dhamdhare A. Challenges in inferring Internet congestion using throughput measurements. In: *Proceedings of the 2017 Internet Measurement Conference*; 2017 Nov 1–3; London, UK. p. 43–56. doi:10.1145/3131365.3131382.
4. Prehn L, Foremski P, Gasser O. Kirin: hitting the Internet with distributed BGP announcements. In: *Proceedings of the 19th ACM Asia Conference on Computer and Communications Security*; 2024 Jul 1–5; Singapore. p. 19–34. doi:10.1145/3634737.3657000.
5. Stöger F, Birge-Lee H, Giuliani G, Subira-Nieto J, Perrig A. BGP vortex: update message floods can create internet instabilities. In: *Proceedings of the 34th USENIX Security Symposium (USENIX Security 25)*; 2025 Aug 13–15; Seattle, WA, USA. p. 3613–29.
6. Testart C, Clark DD. A data-driven approach to understanding the state of internet routing security. In: *Proceeding of the Research Conference on Communication, Information and Internet Policy*; 2020 Feb 17–19; Washington, DC, USA. doi:10.2139/ssrn.3750155.
7. Jin Y, Renganathan S, Ananthanarayanan G, Jiang J, Padmanabhan VN, Schroder M, et al. Zooming in on wide-area latencies to a global cloud provider. In: *Proceedings of the ACM Special Interest Group on Data Communication*; 2019 Aug 19–23; Beijing, China. p. 104–16. doi:10.1145/3341302.3342073.
8. Gray C, Mosig C, Bush R, Pelsser C, Roughan M, Schmidt TC, et al. BGP beacons, network tomography, and Bayesian computation to locate route flap damping. In: *Proceedings of the ACM Internet Measurement Conference*; 2020 Oct 27–29; Virtual. p. 492–505. doi:10.1145/3419394.3423624.
9. Qiu J, Gao L. As path inference by exploiting known as paths. In: *Proceedings of the Global Communications Conference*; 2005 Nov 28–Dec 2; St. Louis, MO, USA.
10. Madhyastha HV, Isdal T, Piatek M, Dixon C, Anderson T, Krishnamurthy A, et al. iPlane: an information plane for distributed services. In: *Proceedings of the Symposium on Operating Systems Design and Implementation*; 2006 Nov 6–8; Seattle, WA, USA. p. 367–80.
11. Madhyastha HV, Katz-Bassett E, Anderson TE, Krishnamurthy A, Venkataramani A. iPlane nano: path prediction for peer-to-peer applications. In: *Proceedings of the Symposium on Networked System Design and Implementation*; 2009 Apr 22–24; Boston, MA, USA. p. 137–52.
12. Cunha Í, Marchetta P, Calder M, Chiu Y-C, Machado BVA, Pescapè A, et al. Sibyl: a practical internet route oracle. In: *Proceedings of the Symposium on Networked System Design and Implementation*; 2016 Mar 16–18; Santa Clara, CA, USA. p. 325–44.
13. Li X, Cai Z, Hou B, Liu N, Liu F, Cheng J. ProbInfer: probability-based AS path inference from multigraph perspective. *Comput Netw*. 2020;180(6):107377. doi:10.1016/j.comnet.2020.107377.

14. Jin Z, Shi X, Ma Q, Sun L, Wang Z, Yin X, et al. Which way to go? Inferring fine-grained AS paths with PathRadar. In: Proceedings of the IEEE INFOCOM 2025—IEEE Conference on Computer Communications; 2025 May 19–22; London, UK. p. 1–10. doi:10.1109/infocom55648.2025.11044440.
15. Li X, Zhou T, Cai Z, Su J. Realizing fine-grained inference of AS path with a generative measurable process. *IEEE/ACM Trans Netw.* 2023;31(6):3112–27. doi:10.1109/tnet.2023.3270565.
16. Li X, Wang C, Yang T, Shen A, Qiu Z, Hou B, et al. Realizing personalized and adaptive inference of AS paths with a generative and measurable process. *IEEE Trans Netw.* 2025;33(2):729–44. doi:10.1109/tnet.2024.3506156.
17. Li X, Wang C, Yang Y, Hou C, Hou B, Cai Z. Internet inter-domain path inferring: methods, applications, and future directions. *Comput Mater Contin.* 2024;81(1):53–78. doi:10.32604/cmc.2024.055186.
18. Green T, Lambert A, Pelsser C, Rossi D. Leveraging inter-domain stability for BGP dynamics analysis. In: *Passive and active measurement*. Cham, Switzerland: Springer International Publishing; 2018. p. 203–15. doi:10.1007/978-3-319-76481-8_15.
19. Bakhshaliyev K, Canbaz MA, Gunes MH. Investigating characteristics of Internet paths. *ACM Trans Model Perform Eval Comput Syst.* 2019;4(3):1–24. doi:10.1145/3342286.
20. Comarela G, Gürsun G, Crovella M. Studying interdomain routing over long timescales. In: Proceedings of the 2013 Conference on Internet Measurement Conference; 2013 Oct 23–25; Barcelona, Spain. p. 227–34. doi:10.1145/2504730.2504771.
21. Augustin B, Cuvellier X, Orgogozo B, Viger F, Friedman T, Latapy M, et al. Avoiding traceroute anomalies with Paris traceroute. In: Proceedings of the 6th ACM SIGCOMM Conference on Internet Measurement; 2006 Oct 25–27; Brazil: Rio de Janeiro. p. 153–8. doi:10.1145/1177080.1177100.
22. Vermeulen K, Gurmericliler E, Cunha I, Choffnes D, Katz-Bassett E. Internet scale reverse traceroute. In: Proceedings of the 22nd ACM Internet Measurement Conference; 2022 Oct 25–27; Nice, France. p. 694–715. doi:10.1145/3517745.3561422.
23. Liu N, Jia C, Hou B, Hou C, Chen Y, Cai Z. 6Search: a reinforcement learning-based traceroute approach for efficient IPv6 topology discovery. *Comput Netw.* 2023;235(5):109987. doi:10.1016/j.comnet.2023.109987.
24. Zhao J, Shi F, Xu C, Peng J, Ge M, Xue P, et al. TNet: efficient IPv6 active network discovery. *Comput Netw.* 2026;280(14s):112170. doi:10.1016/j.comnet.2026.112170.
25. Zhang B, Bi J, Wang Y, Zhang Y, Wu J. Refining IP-to-AS mappings for AS-level traceroute. In: Proceedings of the 2013 22nd International Conference on Computer Communication and Networks (ICCCN); 2013 Jul 30–Aug 2; Nassau, Bahamas. p. 1–7. doi:10.1109/icccn.2013.6614180.
26. Luckie M, Dhamdhere A, Huffaker B, Clark D, Claffy K. Bdrmap: inference of borders between IP networks. In: Proceedings of the 2016 Internet Measurement Conference; 2016 Oct 26–28; Santa Monica, CA, USA. p. 381–96. doi:10.1145/2987443.2987467.
27. Marder A, Smith JM. MAP-IT: multipass accurate passive inferences from traceroute. In: Proceedings of the 2016 Internet Measurement Conference; 2016 Oct 26–28; Santa Monica, CA, USA. p. 397–411. doi:10.1145/2987443.2987468.
28. Marder A, Luckie M, Dhamdhere A, Huffaker B, Claffy K, Smith JM. Pushing the boundaries with bdrmapIT: mapping router ownership at Internet scale. In: Proceedings of the Internet Measurement Conference 2018; 2018 Oct 31–Nov 2; Boston, MA, USA. p. 56–69. doi:10.1145/3278532.3278538.
29. Gao L. On inferring autonomous system relationships in the Internet. *IEEE/ACM Trans Netw.* 2001;9(6):733–45. doi:10.1109/90.974527.
30. Singh R, Tench D, Gill P, McGregor A. PredictRoute: a network path prediction toolkit. In: Proceedings of the Abstract Proceedings of the 2021 ACM SIGMETRICS/International Conference on Measurement and Modeling of Computer Systems; 2021 Jun 14–18; Virtual. p. 21–2. doi:10.1145/3410220.3460107.
31. Tao N, Chen X, Fu X. AS path inference: from complex network perspective. In: Proceedings of the 2015 IFIP Networking Conference (IFIP Networking); 2015 May 20–22; Toulouse, France. p. 1–9. doi:10.1109/ifipnetworking.2015.7145303.
32. Friedman JH, Popescu BE. Predictive learning via rule ensembles. *Ann Appl Stat.* 2008;2(3):916–54. doi:10.1214/07-aos148.

33. Wu T, Wang JH, Wang J, Zhuang S. RouteInfer: inferring interdomain paths by capturing ISP routing behavior diversity and generality. In: *Passive and active measurement*. Cham, Switzerland: Springer International Publishing; 2022. p. 216–44. doi:10.1007/978-3-030-98785-5_10.
34. Jin Y, Scott C, Dhamdhare A, Giotsas V, Krishnamurthy A, Shenker S. Stable and practical AS relationship inference with ProbLink. In: *Proceedings of the 16th USENIX Symposium on Networked Systems Design and Implementation (NSDI 19)*; 2019 Feb 26–28; Boston, MA, USA. p. 581–98.
35. Jin Z, Shi X, Yang Y, Yin X, Wang Z, Wu J. TopoScope: recover AS relationships from fragmentary observations. In: *Proceedings of the ACM Internet Measurement Conference*; 2020 Oct 27–29; Virtual. p. 266–80. doi:10.1145/3419394.3423627.
36. Prehn L, Feldmann A. How biased is our validation (data) for AS relationships? In: *Proceedings of the 21st ACM Internet Measurement Conference*; 2021 Nov 2–4; Virtual. p. 612–20. doi:10.1145/3487552.3487825.
37. Peng S, Shu X, Ruan Z, Huang Z, Xuan Q. Classifying multiclass relationships between ASes using graph convolutional network. *Front Eng Manag*. 2022;9(4):653–67. doi:10.1007/s42524-022-0217-1.
38. Shi X, Jin Z, Xiong B, Huang X, Xi X, Li D, et al. HELA: inferring AS relationships with a hybrid of empirical and learning algorithms. *IEEE Trans Netw*. 2025;33(5):2648–63. doi:10.1109/ton.2025.3572148.
39. RouteViews. Oregon route views project. 2026 [cited 2026 Jan 1]. Available from: <http://www.routeviews.org>.
40. RIPENCC. RIPE RIS. 2026 [cited 2026 Jan 1]. Available from: <http://www.ripe.net/ris/>.
41. AS relationships (serial-1). 2026 [cited 2026 Jan 1]. Available from: [https://catalog.caida.org/dataset/as\\$delimiter"0383383\\$_relationships\\$delimiter"0383383\\$_serial\\$delimiter"0383383\\$_1](https://catalog.caida.org/dataset/as$delimiter).
42. Center for applied internet data analysis. 2026 [cited 2026 Jan 1]. Available from: <https://www.caida.org/>.
43. Huston G. Interconnection, peering and settlements. *Internet Protoc J*. 1999;2(1):1–29.
44. CAIDA. AS rank. 2026 [cited 2026 Jan 1]. Available from: <http://as-rank.caida.org/>.