Chair of Network Architectures and Services
School of Computation, Information, and Technology
Technical University of Munich

TECHNICAL UNIVERSITY OF MUNICH

SCHOOL OF COMPUTATION, INFORMATION, AND TECHNOLOGY

INFORMATICS

MASTER'S THESIS IN INFORMATION SYSTEMS

Detecting Network Latency Anomalies Using Regular Traceroute Measurements

Kempec Halk

Technical University of Munich
School of Computation, Information, and Technology
Informatics

Master's Thesis in Information Systems

Detecting Network Latency Anomalies Using Regular Traceroute Measurements

Erkennung von Netzwerk-Latenz-Anamolien durch regelmäßige Traceroute-Messungen

Author:	Kempec Halk
Supervisor:	Prof. Dr.-Ing. Georg Carle
Advisor:	Tim Betzer, M. Sc.
	Michael Oberrauch, M. Sc.
Date:	September 22, 2025

I confirm that this Master's Thesis is my own work and I have documented all sources and material used.

Garching, September 22, 2025

Location, Date

Signature

The author additionally includes a Qualified Electronic Signature pursuant to EU Regulation 910/2014.

Abstract

Observing and understanding the intricate behavior of the Internet to detect abnormal events like infrastructure damage requires systematic analysis of large-scale traceroute measurements. Existing research in network anomaly detection primarily focuses on either path-based changes or end-to-end performance metrics like the Round-Trip-Time. However, topological methods may not detect performance degradations, while analyses based on performance fail to differentiate between true anomalies and regular routing changes.

To address this, we propose a systematic methodology that employs a comprehensive multistage analysis for large-scale traceroute data. We establish a dynamic and stable baseline profile to distinguish significant deviations from normal network behavior and calculate a detailed set of statistical and path-based metrics. Subsequently, we compare current sets of traceroute data against this baseline at multiple statistical and topological levels. Furthermore, we employ statistical tests to analyze the entire distributions of datasets. Finally, we correlate multiple anomaly indicators into composite events to provide high confidence in detection.

We validate the effectiveness of our methodology through several real-world case studies, including a major power outage on the Iberian Peninsula and multiple Baltic Sea submarine cable incidents. In our analysis, we successfully identified significant rerouting events with substantial performance degradation as well as more subtle and asymmetric impacts.

Conclusively, our work presents a robust and systematic method to detect and classify network anomalies beyond simple outage detection. By establishing a reliable baseline with a multistage analysis of traceroute data, our approach offers a highly detailed examination of network events with high confidence in findings. As a result, it presents an effective procedure to gain a deeper understanding of the health and stability of the global Internet infrastructure.

Zusammenfassung

Die Beobachtung und das Verständnis des komplexen Verhaltens des Internets zur Erkennung von Ereignissen wie Infrastrukturschäden erfordern die systematische Analyse von umfangreichen Traceroute-Messungen. Bestehende Forschung zur Anomalieerkennung in Netzwerken konzentriert sich hauptsächlich auf pfadbasierte Änderungen oder auf Ende-zu-Ende-Leistungsmetriken wie die Round-Trip-Time. Topologische Methoden können jedoch Leistungsverschlechterungen möglicherweise nicht erkennen, während auch leistungsbasierte Analysen nicht zwischen echten Anomalien und normalen Routing-Änderungen unterscheiden können.

Um dieses Problem zu lösen, schlagen wir eine systematische Methode vor, die eine umfassende, mehrstufige Analyse für große Mengen von Traceroute-Daten verwendet. Wir erstellen ein dynamisches und stabiles Baseline-Profil, um signifikante Abweichungen vom normalen Netzwerkverhalten zu unterscheiden, und berechnen einen detaillierten Satz statistischer und pfadbasierter Metriken. Anschließend vergleichen wir aktuelle Sätze von Traceroute-Daten auf mehreren statistischen und topologischen Ebenen mit dieser Baseline. Darüber hinaus verwenden wir statistische Tests, um die gesamten Verteilungen der Datensätze zu analysieren. Schließlich korrelieren wir mehrere Anomalie-Indikatoren zu zusammengesetzten Ereignissen, um eine hohe Sicherheit in der Erkennung zu gewährleisten.

Wir validieren die Wirksamkeit unserer Methode anhand mehrerer Fallstudien aus der Praxis, darunter ein großer Stromausfall auf der Iberischen Halbinsel und mehrere Vorfälle mit Unterseekabeln in der Ostsee. In unserer Analyse haben wir erfolgreich signifikante Rerouting-Ereignisse mit erheblicher Leistungsverschlechterung sowie subtilere und asymmetrische Auswirkungen identifiziert.

Zusammenfassend stellt unsere Arbeit eine robuste und systematische Methode zur Erkennung und Klassifizierung von Netzwerkanomalien vor, die über die einfache Ausfallerkennung hinausgeht. Durch die Etablierung einer zuverlässigen Baseline mit einer mehrstufigen Analyse von Traceroute-Daten bietet unser Ansatz eine sehr detaillierte Untersuchung von Netzwerkereignissen mit hoher Zuverlässigkeit der Ergebnisse. Als Resultat stellt dies ein effektives Verfahren dar, um ein tieferes Verständnis für den Zustand und die Stabilität der globalen Internetinfrastruktur zu gewinnen.

First and foremost, I would like to thank my partner for her amazing support during the entire period of working on my thesis. Furthermore, I want to thank my parents for standing behind me unconditionally during all the years of my studies, no matter my progress or the time it took me. Also, I am very thankful to everyone I met during my studies, who helped me through rough phases and times of uncertainty. I would not have been able to finish my degree without all of you.

I would also like to thank my advisors for their help with my thesis, be it feedback, issues during the research, or supplying me with any hardware I needed. Finally, I want to thank Prof. Carle for giving me the opportunity to work on this subject at his chair, enabling me to finish the final part of my studies.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The global Internet infrastructure depends on physical connections that are vulnerable to disruptions. Although the modern Internet is quite resilient, incidents like the submarine cable cuts in the Baltic Sea[1] highlight this vulnerability. Damage to this critical data infrastructure can isolate regions, disrupt economies and impact communication on a massive scale. However, the challenges extend beyond the immediate effects of such dramatic failures. The Internet is a decentralized network composed of interconnected systems, where traffic routing is governed by complex policies and commercial agreements rather than solely by optimal performance. This complexity makes it very difficult to observe and understand the actual paths that data travels. Additionally, it can be challenging to detect subtle degradations or instabilities before they develop into major outages. The routes are often asymmetric and can appear counterintuitive, highlighting the need for empirical observation to truly understand the operational state of the Internet.

In this thesis, we address the challenge of observing and understanding the intricate behavior of the Internet to better comprehend and detect the effects of events like infrastructure damage. To achieve this, we propose and implement a systematic methodology for analyzing large-scale traceroute measurements to detect and classify network anomalies. We utilize the extensive publicly available data from RIPE Atlas, a global network of thousands of measurement probes that provide a comprehensive view of Internet connectivity. Our approach employs the principle of dynamic baselines. Therefore, we establish a stable and reliable profile of normal network behavior, while considering

[1] Two Baltic submarine cables were damaged by a cut on the 18th of November 2024

predictable temporary variations, such as time of day. This allows us to accurately distinguish true anomalies from normal network fluctuations.

The core of our methodology is a comprehensive multistage analysis designed to process massive datasets of traceroute measurements. We collect raw data and store it in a high-performance ClickHouse database, optimized for large-scale analytical queries, which are essential for our research. After preprocessing the data, we generate a detailed set of statistical and path-based metrics that capture both the performance and topological characteristics of network traffic. Furthermore, instead of identifying anomalies based on static thresholds, we employ an elaborate method to detect significant deviations from the established baseline. Additionally, we apply statistical tests to compare entire data distributions, allowing us to detect subtle shifts in performance profiles that a sole aggregated value might miss. Finally, we correlate individual anomaly indicators into composite events, providing high confidence in detecting abnormal network events.

In the following chapter, we review existing research on network anomaly detection. In order to comprehend the whole topic, we then describe background information about the fundamental concepts and technologies relevant to this work in Chapter 3. In Chapter 4, we offer a detailed description of our methodology for traceroute processing and anomaly identification. Subsequently, in Chapter 5, we present the findings from applying our approach to real-world data, showcasing the detection of various network events. Next, we evaluate the effectiveness and accuracy of our methodology in Chapter 6. Finally, we conclude our topic and its contribution in Chapter 7 and present potential directions for future research.

Chapter 2

Related Work

Detecting and characterizing anomalies in Internet routing is an ongoing field of research. In this chapter, we review existing publications, their scope and their approach to identify anomalies.

2.1 Observational Studies of Network Events

The field of network anomaly detection has been significantly advanced by the availability of large-scale measurement platforms such as RIPE Atlas [1]. Research has successfully demonstrated the utility of this data to analyze disruptions in network connectivity and performance. A prime example of using RIPE Atlas data for event analysis is the research of the Baltic sea submarine cable cuts in November 2024 by Aben *et al.* at RIPE NCC. [5][6][4]

In an initial analysis published shortly after the C-Lion1 cable cut, Aben *et al.* used RIPE Atlas ping measurements between anchor nodes to provide a preliminary assessment of the impact due to the damage. They observed that while 30 % of measured paths between Finland and Germany experienced latency increases, there was no considerable increase in packet loss. This led to the conclusion that the Internet successfully routed around the damage with only a modest performance impact. [6]

A subsequent, more detailed investigation by Aben *et al.* provides a deeper analysis of the event by including traceroute measurement data. They investigated the specific changes in routing and latency across different levels. Their analysis revealed that Internet resilience manifests in multiple distinct layers. They quantified the impact at each level, finding that around 30 % of paths had inter-domain changes, almost half

experienced intra-domain rerouting and more than 21 % showed evidence of circuit-level changes. Furthermore, they found that a noteworthy number of the path shifts show a latency increase of over 3 ms in each category. [4]

2.2 Path-Based Anomaly Detection

Many research publications focus on investigating the network path itself, analyzing changes in topology to identify deviations in the routing data. These methods often utilize information from the routing protocol Border Gateway Protocol to detect anomalies such as route hijacks and misconfigurations.

Yang *et al.* propose a path-based routing anomaly detection algorithm that combines both topological and attribute information from BGP routing tables. They transform the anomaly detection into a graph theory problem, using a random walk sampling method to effectively capture the underlying structure of the routing table. By constructing path and attribute embeddings, their model creates routes as a comprehensive feature representation and utilizes those to identify anomalies in the routing table. The approach is designed to be more robust than methods that rely on fixed topological patterns, which can degrade in performance in complex network environments. [7]

Liu *et al.* introduce a transmitter-oriented approach using the network metric of betweenness centrality to characterize inter-domain routing events. By analyzing the variation of AS betweenness centrality over time, they can identify the temporal, topological and relational characteristics of route changes following disruptive events. [8]

A survey by Al-Musawi *et al.* provides a thorough overview of the field, classifying BGP anomalies and the various techniques used to identify them. It highlights the ongoing challenges in real-time detection and identification of the source of the anomalies. [9]

Although topological methods are effective for identifying routing anomalies, such as BGP hijacks, they have some limitations. These methods often rely on access to Border Gateway Protocol data streams, which may not always be available. Additionally, they can have difficulties detecting performance-related issues, like congestion, that do not result in changes in the path at the Autonomous System level.

2.3 Performance-Based Anomaly Detection

Another widespread approach to detect anomalies is the analysis of end-to-end performance metrics, most commonly the Round-Trip-Time. These methods are effective at detecting performance degradations that may not be visible to topological approaches.

Hou *et al.* propose an unsupervised two-step method to detect and characterize anomalies in large-scale RTT time series data. They first use a change-point detection algorithm to identify shifts in Round-Trip-Time time series. Subsequently, they apply a second change-point detection to this series to identify moments of correlated state changes, which are flagged as anomalous events. Their approach is designed to reduce the false alarms that can be caused by normal RTT fluctuations in highly variable network environments. [10]

In a publication from 2017, Fontugne *et al.* utilize the diversity of RIPE Atlas traceroute measurements to monitor in-network delays. They introduce a method for estimating delay changes on intermediate links by calculating differential RTTs from different vantage points. By monitoring the distribution of the Round-Trip-Times over time, they detect and locate abnormal delay changes. [11]

These performance-based methods are effective but have significant disadvantages. As noted by Hou *et al.*, methods solely based on RTT changes have difficulties distinguishing between anomalous network events and normal fluctuations caused by legitimate routing changes [10].

Chapter 3

Background

First, we illustrate the basics of Internet routing to understand how the path for traffic and individual packets is decided to reach the destination. Next, we explain the measurement tool that creates the data on which our analysis is based. Furthermore, we describe RIPE Atlas, the platform that provides foundational information and measurements for our research. Subsequently, we introduce ClickHouse, the database we utilize to store the large amount of data required in our investigation to detect anomalies. Afterwards, we detail how the statistical method Interquartile Range we utilize to identify outliers in a dataset works. Finally, we describe the statistical distribution tests Kolmogorov-Smirnov and Anderson-Darling.

3.1 Internet Routing

The global Internet is not a single, monolithic network but rather a decentralized network of networks. This architecture is composed of thousands of Autonomous Systems (ASes), which are independently operated by entities such as Internet Service Providers (ISPs), large companies, universities or government institutions. An AS is a set of routers under a single technical administration with a clearly defined routing policy to the Internet. [12]

To interconnect these networks, the routing protocol Border Gateway Protocol (BGP) is used to exchange routing and reachability information between the Autonomous Systems. BGP allows an AS to announce its IP address ranges, known as prefixes, to its neighbors, advertising them to the Internet. Furthermore, the Border Gateway Protocol is classified as a path-vector routing protocol and determines the best route for packets

based on factors including path details, network policies and reliability. It updates routes dynamically in response to network changes, ensuring Internet stability. [13][14][15]

Because of these policies, the path taken from a source to a destination can differ from the route chosen in the reverse direction. Moreover, the various factors influencing the path make predicting the route a packet takes challenging. Therefore, analyzing Internet traffic is highly complex and requires large amounts of data from measurement tools like traceroute to discover anomalies across networks.

3.2 Traceroutes

Traceroute is a network diagnostic tool used to map the sequence of router IP addresses, known as hops, and measure the latency to each one [16]. The fundamental mechanism involves sending probe packets to a destination with progressively increasing Time-to-Live (TTL) values. The TTL is a field in the IP header decremented by every hop forwarding the packet. When the TTL value reaches zero, the hop discards the packet and sends an ICMP Time Exceeded message back to the source. This mechanism is used as an iterative process of four steps, increasing the TTL on each stage, beginning with one, to build a map of the path. [17]

1. Per-Hop Probing: For each TTL value, starting from one, the probe sends a series of packets, typically three, towards the final destination.
2. ICMP Response: The hop in the path corresponding to the one decrementing the TTL value to zero returns an ICMP Time Exceeded message for each packet in the burst.
3. Per-Hop RTT: The source probe captures the response and records the Round-Trip-Time for each reply, resulting in multiple measurements for each hop.
4. TTL Increment: The value of the Time-to-Live is increased by one and the process is repeated until the current hop corresponds to the destination or a maximum hop count is exceeded.

Figure 3.1 shows a traceroute of a path with one hop between source and destination.

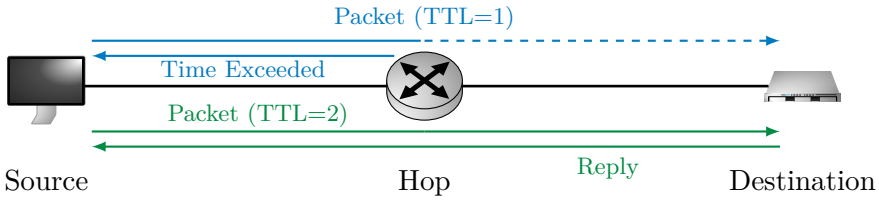


Figure 3.1: Traceroute

3.3 RIPE Atlas

The RIPE Atlas [1] project is a global Internet measurement platform operated by the RIPE NCC (Réseaux IP Européens Network Coordination Centre) [5]. The project aims to provide detailed information on the Internet infrastructure, usage and development. To do so, RIPE Atlas operates an extensive and strategically distributed network of devices known as probes at different geographical locations worldwide. Figure 3.2 shows a map of all active probes, represented as green dots, around the world. We can observe that the majority of RIPE Atlas probes are located in Central and Western Europe.

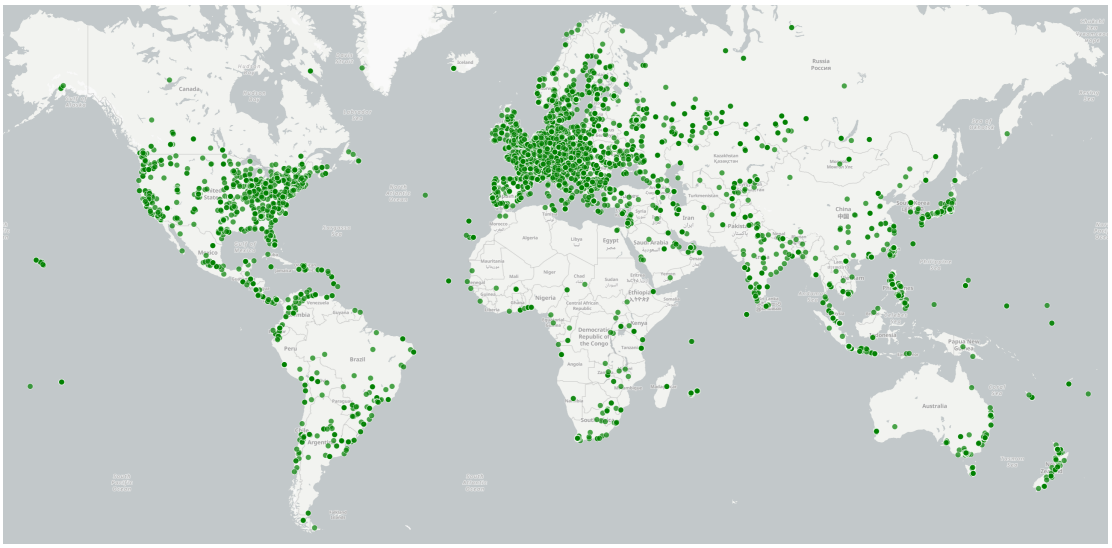


Figure 3.2: RIPE Atlas Probes

These probes conduct real-time measurements of a wide range of Internet metrics, such as latency, packet loss and reachability. Probes are categorized into two different types.

1. Probes are small hardware devices or software instances that can be deployed on existing infrastructure. They run measurements in the RIPE Atlas system and report their results to the data collection components.

   The geolocation of a standard probe is user-submitted and generally not verified by the RIPE NCC. While the platform performs automated tasks trying to validate locations, there is no assurance that a geolocation will be accurate. Additionally, RIPE Atlas reports semi-randomized coordinates for probes to enhance user privacy [1]. Consequently, the reliability of the geolocation depends on the truthworthiness of its hosts.

2. Anchors are an enhanced, more powerful type of probes. They can either be rack-mounted hardware devices or set up as virtual anchors in the form of a virtual machine [1]. Anchors are designed to be installed in a network with high availability and capacity. Therefore, they have a much higher measurement capability than standard probes and provide a high uptime and stable connectivity. In addition to the vast amount of measurements anchors perform, they also act as reliable targets for measurements from probes and other anchors [5].

   Hosting an anchor is exclusively available to vetted organizations, coordinated with the RIPE NCC to uphold network stability and integrity. The location of an anchor is thoroughly verified by the RIPE NCC, which involves a careful confirmation of the physical address. Therefore, we can use any anchor with high confidence in its reliability and accuracy of the provided coordinates.

Furthermore, the platform maintains an extensive collection of historical data, providing valuable data for network operators, researchers and the technical community.

In order to operate such a vast number of probes, RIPE Atlas relies on volunteers in the Internet community to act as hosts. This group includes individuals, corporations and academic institutions. In exchange for hosting, these volunteers earn credits to run their own customized measurements. This creates a mutually beneficial ecosystem that promotes the growth of the RIPE Atlas network.

RIPE Atlas offers a variety of measurement types. These include DNS lookups to investigate domain name resolutions or ping tests to measure the latency and availability of hosts. Most significantly, the platform also provides traceroute measurements to identify paths the packets take across the Internet and the RTT (Round-Trip-Time) it takes for those packets. These traceroute measurements will form the essential foundation for the data in our comprehensive analysis.

Since the project was launched in 2010, RIPE Atlas has grown to become the largest measurement platform of its kind, providing a comprehensive view of the Internet connection state [5]. As of August 2025, RIPE Atlas was composed of almost 55.000 probes and over 1.000 active anchors, primarily located in Europe [1].

Therefore, RIPE Atlas represents an ideal foundation for the data we will be using in our analysis, which focuses on finding network anomalies based on traceroute measurements from Europe.

3.4 ClickHouse

Our analysis involves handling massive amounts of data. Consequently, a high-performance storage and query backend is essential. Traditional row-oriented databases designed for Online Transaction Processing (OLTP) are often inefficient for the types of large-scale aggregations and scans as required for our research. Therefore, a columnar Database Management System (DBMS) is a more suitable choice.

ClickHouse is an open-source, column-oriented DBMS. It was explicitly designed for Online Analytical Processing (OLAP) and allows the generation of analytical reports using SQL queries in real time. Unlike row-oriented systems that store all values of a record in a row sequentially, the column-oriented architecture in ClickHouse stores each column of a dataset independently. This architecture provides several advantages for analytical queries and heavy data loads. [18][19][20][21]

- High Performance on Aggregations: When calculating a statistic, like the median, the system only needs access to data from a single column instead of reading every row and filtering for that specific column. This approach significantly reduces the amount of data loaded from disk, resulting in considerable performance increases.
- Efficient Data Compression: Data within a single column is often very uniform, allowing for higher compression ratios compared to row-based storage. This leads to reduced disk space usage and improved query execution times due to smaller data read requirements.
- Vectorized Query Execution: The query engine in ClickHouse is designed to process chunks of data from a column, known as vectors, rather than handling individual rows sequentially. This method makes better use of modern CPU caches and SIMD (Single Instruction, Multiple Data) instructions, resulting in significant performance improvements.

ClickHouse is an ideal backend for storing and analyzing the large datasets used in our analysis. Its architecture and features allow us fast and interactive data querying, essential for establishing a data baseline and conducting comprehensive analysis. [20][21]

3.5 Interquartile Range Method

The Interquartile Range (IQR) method is a widely recognized statistical approach commonly used in various data mining and machine learning frameworks to identify data points that significantly differ from the rest of the dataset [22]. Therefore, the IQR

method represents an effective statistical method for outlier detection, to categorize extreme data points [23][24][25].

To find outliers using the IQR method, we first sort the dataset from smallest to largest and divide it into four equal parts, or quartiles. Next, we determine the first quartile, denoted as Q_1 and the third quartile, denoted as Q_3. The first quartile represents the 25^{th} percentile, while the third quartile represents the 75^{th} percentile of the data. Subsequently, we calculate the Interquartile Range by subtracting the first quartile from the third quartile.

$$\text{IQR} = Q_3 - Q_1 \tag{3.1}$$

The IQR represents the spread of the middle 50 % of the values in our dataset. Finally, we use the IQR value to specify an upper fence UF, using the common threshold of 1.5.

$$\text{UF} = Q_3 + (1.5 * \text{IQR}) \tag{3.2}$$

Any value in our data that falls above this upper fence is determined as an outlier.

3.6 Kolmogorov-Smirnov Test

The Kolmogorov-Smirnov (K-S) test was first introduced by Andrey Kolmogorov in 1933 and later extended by Nikolai Smirnov in 1948 [26][27]. It represents one of the most fundamental nonparametric statistical tests to compare probability distributions. The K-S test provides a method to determine whether two independent data samples come from the same underlying distribution or whether a sample follows a specified theoretical distribution. This makes the test particularly useful in scenarios where the underlying distribution of the data is unknown. The test is based on the empirical distribution function and provides a robust method for comparing distributions without making assumptions about their underlying shape. [28]

The fundamental principle of the K-S test is the analysis of the empirical cumulative distribution functions (ECDFs) of the two samples being compared. The test measures the maximum absolute difference between these empirical distribution functions across all possible values in the combined sample space. This approach differs significantly from parametric tests, which focus on specific characteristics of the distribution, such as means or variances. Instead, the K-S test analyzes the entire shape of the distribution. [28]

The two-sample K-S test is beneficial for comparing a data sample against an established base sample, providing both a test statistic and a p-value that measures the probability

of such differences appearing [29]. The null hypothesis in this context is that both samples are drawn from the same distribution.

A significant advantage of the K-S test is its distribution-free nature. This means it does not make any assumptions about the form of the underlying distributions of the analyzed data. This makes it especially useful for analyzing network measurement data, where latency distributions often display complex, non-Gaussian characteristics due to factors like network congestion or routing changes [30].

However, the K-S test has limitations that must be considered in practical applications. It is most sensitive to differences in the central portions of the distributions and may be less effective at detecting differences in the tails [31].

3.7 Anderson-Darling Test

The Anderson-Darling (A-D) test, developed by Theodore Anderson and Donald Darling in 1952, is a statistical method used to determine whether a data sample comes from a specified distribution [32]. While sharing conceptual similarities with the Kolmogorov-Smirnov test in a nonparametric approach to distribution comparison, described in Section 3.6, the A-D test utilizes a weighted distance measure. The Anderson-Darling test applies greater significance to observations in the tail regions of the distribution. The k-sample form of the A-D test evaluates whether multiple samples come from the same distribution without specifying the underlying distribution beforehand. [33]

The mathematical foundation of the A-D test involves the integration of squared differences between empirical distribution functions, weighted by the inverse of the variance of the base distribution. This results in a quadratic form that identifies not only the magnitude of distributional differences but also their statistical significance at each point along the distribution. The weighting function ensures that deviations in regions where the reference distribution has low density receive proportionally greater influence on the test statistic. [31]

This characteristic makes the Anderson-Darling test particularly effective at detecting differences in the lower and upper extremes of data. In the context of network latency analysis, this is an advantage. Network performance degradation, such as that caused by congestion or flapping, often appears as an increase in latency. These events populate the upper tail of the RTT distribution [11][34]. Due to the increased sensitivity in tails, the A-D test is very capable at detecting such differences in latency. The k-sample version of the test allows for the comparison of two or more independent samples. It

produces a statistic as well as a corresponding p-value to assess the null hypothesis of a common underlying distribution. [33]

Chapter 4

Methodology

The foundation of our approach is a multistage process designed to analyze large amounts of RIPE Atlas traceroute data. We transform raw measurement information through multiple stages into conclusive insights about network health, allowing us to identify anomalies. Before starting our process, we must collect massive amounts of raw historical traceroute measurements from RIPE Atlas and select key data fields relevant for our analysis. Subsequently, we store these values in a ClickHouse database. Next, we process our data in multiple stages, preparing it for our deep analysis procedure. Since network performance is inherently dynamic, atypical behavior can best be identified relative to an operationally typical network state. Instead of relying on static or absolute thresholds, we first establish a baseline profile of normal behavior. Consequently, we are able to detect deviations from the baseline. Subsequently, we introduce a number of anomaly indicators, triggered when significant deviations from the baseline are detected. Finally, we correlate individual indicators to identify composite events, providing a recognition of network anomalies with high confidence.

4.1 Data Gathering

The foundation of our analysis is based on the extensive and publicly available traceroute measurement data provided by RIPE Atlas. As detailed in Section 3.3, the global distribution and scale of the RIPE Atlas project provide an optimal resource to study Internet routing behavior. For our analysis, we utilize historical data archives from the platform. These dumps contain all globally gathered traceroute measurements from RIPE Atlas, organized into large hourly files. With each file and measurement con-

taining a lot of detail, we need to extract key information and store the selected data appropriately and efficiently.

4.1.1 RIPE Atlas Traceroute Data

Each line in the RIPE Atlas data relates to a JSON object representing a single traceroute measurement. While each record contains numerous fields, we focus on extracting a specific subset of keys and their corresponding values that are essential for our analysis. Parsing these fields is a crucial step in our data collection process, assembling the structured values that are then inserted into our ClickHouse database. The selected fields, along with a short explanation, are detailed in the following.

- `msm_id`: The Measurement Identifier is a unique integer identifying the specific measurement task. It is the primary key for categorizing all traceroutes that are part of the same measurement process.
- `prb_id`: The Probe Identifier is a unique integer identifying the source RIPE Atlas probe that executed the measurement. This is essential for analysis involving probe locations or types.
- `timestamp`: The Unix Timestamp specifies when the measurement was started. This serves as the primary identifier for analysis requiring an exact period of time.
- `endtime`: The Measurement End Time specifies when the measurement finished. It can be used together with the field of the `timestamp` to calculate the total duration of the measurement.
- `src_addr`: The Source Address contains the IP address of the source probe. This is the initial router in the path and is crucial for any analysis of the routes.
- `dst_addr`: The Destination Address contains the IP address of the target destination. This is the final hop in the route and invaluable for path-based analysis.
- `proto`: The Protocol used for the probe packets. It can be `ICMP`, `UDP` or `TCP`. This allows us to filter for specific protocols.
- `af`: The Address Family specifies which version of the Internet Protocol, denoted by `4` for IPv4 or `6` for IPv6, was used in the measurement. This is essential for interpreting and handling addresses correctly.
- `destination_ip_responded`: The Destination Responded indicates whether or not a reply was received from the destination. This key can be used to calculate the success rate over all measurements.

- `result`: The Result contains an array of objects, representing each single hop with individual measurements for all TTL values in the traceroute. This array contains the whole path and latency information. It is detailed further below.

The Result field is the most intricate and important part of the schema. Each element consists of a hop index, correlating to the TTL number, and a nested `result` array. This nested array holds the actual probe responses for that specific hop, typically three objects, one for each probe packet sent. Each response object contains key fields that we extract for our analysis, explained below.

- `from`: The IP address of the router that returned the ICMP Time Exceeded message. This is the core data for our path reconstruction, explained in Section 4.2.3.
- `rtt`: The measured Round-Trip-Time in milliseconds for that specific probe packet.

When a timeout occurred, the response object only contains the field `x` with an asterisk (`*`) as value, indicating that no response was received for that probe packet.

4.1.2 Data Storage

To store the RIPE Atlas data, we need a method capable of handling the massive volume of traceroute measurements we process. Thus, we utilize the ClickHouse database system, detailed in Section 3.4, to store our data using a schema, where core measurement metadata is separated from the detailed hop results. However, querying raw hop data to calculate the metrics for every analysis would be computationally expensive and slow. Therefore, we employ database views to compute values within ClickHouse, outsourcing a significant portion of the data processing to the database system itself.

4.2 Processing

The traceroute data collected from RIPE Atlas requires some preprocessing through multiple stages to be suitable for our analysis. We apply steps to filter, normalize and structure the data to ensure that analyses are based on consistent and high-quality information. Our processing consists of several key stages, including filtering the dataset for a specific scope, standardizing inconsistent hop information, reconstructing deterministic paths and identifying statistical outliers. Subsequently, we determine metrics for each dataset, on which we base our further analysis. Each step is essential to reduce noise and ensure data integrity as well as quality.

4.2.1 Filtering

The vast amount of measurement data gathered from RIPE Atlas contains a significant number of details and records that can be irrelevant to our analysis. To ensure the quality and integrity of our results, we employ multiple filtering strategies and validation checks while processing the datasets.

First, we exclusively filter for traceroute measurements where both the source and the destination are RIPE Atlas anchors. As detailed in Section 3.3, anchors are highly reliable and powerful probes with a verified geolocation and a stable network connectivity. This is important, as it minimizes the variability and noise that the less reliable standard probes can introduce. Therefore, we ensure our results are based on the most trustworthy and consistent data available.

Additionally, we apply specific geographical filters to the source as well as the destination anchors. This way, we are able to define a precise network traffic passage or area for our analysis. We achieve this through two primary mechanisms. For broad studies between different regions, we filter probes by their registered country codes, allowing us to examine traffic within one or between two or more countries. For a more detailed analysis, we define a geographical point using latitude and longitude coordinates and apply a radius to create an area that defines our selected probes. This is particularly useful for examining traffic over specific, known infrastructure, such as studying certain submarine cables.

Finally, we validate the traceroute data before processing. Each measurement is checked to ensure it contains a result array, guaranteeing that only traceroutes with valid hop data are included in our analysis. Our filter discards records that exhibit signs of data corruption, such as a path length of zero, while the raw path string clearly contains hops.

This strategy, combining geographical and probe type constraints with our integrity validation, assures that every analysis is precisely targeted to study specific sections of the Internet. Additionally, it ensures that our dataset is based on reliable, consistent and high-quality traceroute measurements.

4.2.2 Data Standardization

Each traceroute measurement can contain hops with information that lacks significance, consistency or clarity for our purposes. Those hops can add noise to our path-based analysis and present themselves in two different types.

- Timeouts: A timeout, typically represented by an asterisk (*) in the traceroute measurement, occurs when a hop does not respond to a probe packet. This can be due to network congestion, packet filtering policies or temporary issues. When a timeout happens, the only information we have is that the probe packet went unanswered. Consequently, we do not obtain an IP address or Round-Trip-Time value for that specific probe packet. If all probe packets for a TTL stay unreplied, we do not have any information about the hop in the measurement.
- Private IP addresses: A response from an IP address within the private ranges defined by *RFC1918* provides no significant information for a global path analysis [35]. As these addresses are not unique on the public Internet, a private IP address from one traceroute is unrelated to the same private IP address in another. Treating them as identical would incorrectly merge distinct paths during statistical analysis.

Therefore, to reliably analyze traceroutes, we need to ensure the integrity and comparability of the hop data. To address this, we apply two different standardization methods to the data. First, we apply a filter to all IP addresses, replacing every address within any private IP subnet with a single, consistent marker represented as `PRIVATE`. This prevents the high variation of private IP addresses from skewing path frequency calculations. Furthermore, for our dominant path and core path segment analysis, described in Section 4.2.5, we need to ensure a high grade of information value and comparability in our paths. Therefore, we introduced an uninformative threshold of 80 % that any path must not exceed. A threshold of 80 % has shown to be the best balance between filtering paths with little information and obtaining mostly paths that hold substantial value while still preserving a statistically significant number of routes to get meaningful analysis results. Uninformative hops are hops that did not receive a response, as well as hops in a path that have a private IP address. We calculate the ratio of uninformative hops in a path by dividing the number of uninformative hops by the total path length. Any path that exceeds the uninformative threshold is entirely excluded from the dominant path and core path segment analysis.

This approach ensures that all path data is uniform and comparable, while guaranteeing that the core analysis is performed only on data of high quality and information value.

4.2.3 Path Reconstruction

A key challenge in interpreting traceroute data results from per-packet load balancing, where the multiple packets sent to a single hop may be returned by different router interfaces, as explained in Section 3.2. Therefore, we do not have definite paths for our

measurements, but single-hop information for each value of the TTL parameter on the way to the destination for each measurement. Consequently, we need to reconstruct the paths for our traceroute measurements.

To address this, we employ a deterministic strategy to select the lowest RTT values of a measurement to reconstruct the path. We iterate through the list of responses for each hop and select the IP address of the first router with a valid response. Any subsequent reply for the same hop is ignored during the path reconstruction, regardless of whether the responding router has the same or a different IP address. If no probe packet for a hop has a valid response, an asterisk (∗) is used as a placeholder. This process is repeated for every hop in a measurement, using the hop number to create an ordered list. The result is a single, reproducible path for each traceroute measurement, essential for the large-scale statistical analysis of dominant paths and core path segments, further explained in Section 4.2.5.

4.2.4 Outlier Detection

Raw network latency data frequently contains extreme values that can influence statistical aggregates sensitive to variance, such as the mean and standard deviation. Extremes are not necessarily errors, but can represent valid though infrequent network phenomena, such as temporary congestion or brief rerouting events. Therefore, we identify and characterize those extreme values as outliers without discarding them entirely.

We utilize the IQR method, detailed in Section 3.5, for outlier detection. This method has a few advantages for our analysis. First, as a nonparametric approach, it does not make any assumptions about the underlying distribution of our dataset. This is an important factor for network latency data, which often does not follow a normal distribution. Furthermore, the IQR method is computationally efficient, allowing easy use with massive amounts of data.

To determine outliers within our data, we use the values of the destination RTT and determine the Interquartile Range as well as the upper fence of these values as described in Section 3.5. Subsequently, we compare the destination Round-Trip-Time of each measurement in our data against our determined upper fence to identify the outliers.

A crucial design choice in our approach is not removing measurements identified as outliers. Instead, we mark them with an outlier flag. This method ensures that the complete dataset is preserved. We use analytical methods that are robust to the distortion of the statistical values in our analysis. Our system relies on the median, the 50^{th} percentile, rather than the mean, as the median is highly resistant to the influence of extreme values. Additionally, we apply distribution tests, as described in Section 4.4.1,

to evaluate the entire shape of the data distribution. Since outliers are not a distortion of the data distribution, but an essential part, we require these extremes for a valid and accurate result of our distribution tests.

4.2.5 Metrics

To effectively evaluate network behavior and identify irregularities, we utilize a specific set of metrics derived from RIPE Atlas traceroute measurements. These metrics capture both the performance characteristics and the topological properties of the network paths in our dataset. They serve the foundational data for our anomaly detection analysis, providing the raw values from which we derive our anomaly indicators, detailed in Section 4.4.1. We categorize these metrics into two groups. The statistical metrics and the path metrics are detailed in the following subsections.

Statistical Metrics

Statistical metrics provide numeric values of the network performance and reliability. The values are derived from timing and response data within each traceroute measurement. They are essential for capturing temporal changes in network behavior and represent the base for our statistical indicators, described in Section 4.4.1.

Round-Trip-Time (RTT)

The Round-Trip-Time is an essential metric in network performance monitoring and, therefore, in our analysis. It represents the time difference between sending a probe and receiving a response from a hop in the traceroute path. We specifically capture the RTT value for two different hops in each traceroute measurement for our analysis. Most importantly, we collect the Round-Trip-Time to the final hop, as it represents the end-to-end latency to the destination. On top of that, we also record the first-hop RTT. To ensure accurate statistics, we exclude traceroute measurements that lack a valid response from the destination in our Round-Trip-Time collection. We calculate the median for both of these hop positions over all traceroute measurements in our dataset. We chose the median as a robust measurement value as it is less susceptible to outliers than the mean.

Success Rate

The success rate is defined as the percentage of traceroute measurements in our dataset that successfully reached the intended destination and received a response. A high success rate indicates a stable and reachable destination as well as a reliable path. Consequently, a low success rate is a strong indicator of reachability issues, packet loss or routing problems along the route. This metric provides a high-level overview of the end-to-end connection health.

Timeout Rate

Timeouts occur when probe packets fail to receive responses within a predefined waiting period. We calculate the timeout rate as the percentage of traceroute measurements that contain a timeout in at least one hop during their path to the destination or in the destination itself. Timeouts can be a signal of network problems. An increased timeout rate is an indicator of reachability issues and instability in the network. Therefore, it provides an important metric for detecting network congestion or routing issues.

Path Length

The path length measures the number of hops between the source probe and the destination. It is a very efficient metric for detecting changes in the network path. We compute the median path length across all traceroutes within our current dataset. Changes in the median path length can indicate routing changes. In our analysis, we distinguish between two different types of path length. First, we consider the observable hop count, which is defined as the total number of hops recorded, regardless of whether the traceroute measurement successfully reached its destination. Furthermore, we separately record the path length for successful traceroutes. A significant increase in path length can lead to higher latency and may be a symptom of route flapping or rerouting due to a failure in the primary path.

Duration

The duration of a measurement is the total time passed for a single traceroute measurement to complete, as detailed in Section 3.2. While related to the RTT, the duration is a distinct metric that can be influenced by timeouts and the number of hops. Although not a primary metric for our anomaly detection, it can provide additional information. A significant increase in the average duration of traceroutes can be a secondary indicator of network congestion or other issues that cause delays in the measurement process itself.

Path Metrics

Path metrics focus on the topological characteristics of the network paths. They are critical for understanding the routing behavior of the network and detecting changes that might not be immediately apparent from statistical metrics alone. We use a variety of path metrics in our analysis to detect network anomalies. Traceroute measurements that do not receive a successful reply from the destination are excluded from all path metrics.

Total Paths

The Total Paths metric is a simple count of all valid traceroute paths recorded in a given period. This provides a base value for other path-related metrics.

Unique Paths

The unique paths metric is defined as the number of distinct paths with a unique sequence of IP addresses from source to destination, over all the traceroute measurements in our dataset. The relationship between total and unique paths provides an elemental estimation of path diversity in the network.

Dominant Paths

The Dominant Path is the single most frequently observed unique path within a dataset. It represents the primary route taken by traffic between the source and destination. We track the five most recurring paths and their number of appearances. The dominant paths and their occurrences can be used as a measure of path stability.

Core Paths

The core path analysis is a more advanced path metric that allows us to identify stable segments within network paths. Traceroute measurements often have a high diversity in routing, even when considering a dataset purely containing measurements with the same source and destination. This variety is caused by factors such as load balancing. We primarily analyze data filtered by geographical locations, as described in Section 4.2.1, amounting to a large cluster of different probes and their respective networks for source and destination. This diversity in probe locations introduces a significant amount of noise to the traceroute paths. However, with shared geographical areas for source and destination probes, respectively, they are still likely to follow common core path segments within their routes. To prevent a segment from being a single hop or full path, we define a minimum length of two and a maximum length of twelve hops to be classified as a core path. Figure 4.1 depicts a core path segment within a route [2].

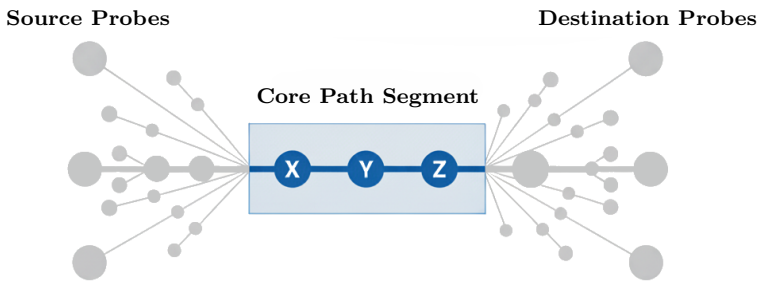


Figure 4.1: Core Path Segment

By identifying and tracking the most frequent continuous sequences of hops throughout all paths in our dataset, we can monitor the most commonly taken routes between two areas. Consequently, we can catch changes in routing by observing the frequency, emergence or disappearance of these core paths. A detected shift can suggest path instability or a decrease in network health.

4.3 Baseline Establishment

The first operational phase of our methodology is the establishment of a robust performance baseline. We explicitly reject the use of static or global thresholds for performance metrics, as such an approach fails to account for the natural and predictable fluctuations inherent in network behavior. Instead, we determine a dynamic baseline from historical data that takes temporal patterns into account. This allows us to define a network state as stable, enabling us to utilize it for comparing further measurements against.

The selection of the baseline dataset is critical, since the statistical properties of that data serve as the benchmark against which we compare future data. Typically, the traceroute measurements chosen should span a sufficiently long period of at least two weeks. We need to do this to capture statistically significant patterns, but still have to avoid known outages, network failures or periods of instability.

We establish our baseline by analyzing the large set of selected traceroute measurement data from RIPE Atlas. During this phase, we process the filtered and standardized dataset to compute the full range of statistical and path-based metrics detailed in Section 4.2.5. This includes calculating the central tendency and distribution for performance metrics like the Round-Trip-Time and path length, as well as the dominant path and core path segments. The aggregated results, including median, standard deviations and path frequencies, are returned as a baseline result.

Additionally, we group data into day of the week, as well as hour of the day brackets. Since the Internet is used in different amounts depending on the weekday as well as the time of day, traffic also behaves differently, demonstrating a change in routing and latency. Therefore, we determine key metrics for each of these groups as well and attach them to our results. We count the number of measurements in each bracket and calculate their median for the Round-Trip-Time as well as observed and successful path length. Furthermore, we compute their RTT standard deviation, success and timeout rate. Likewise, we also provide these same metrics for each single hour in our traceroute dataset.

4.4 Anomaly Detection

Once we have established a stable baseline, we can run the primary analysis to detect anomalies in new datasets. Our detection process is not a single check but a structured procedure that builds simple observations or indicators to composite events with high confidence. The core principle is to compare the metrics of our current dataset against the corresponding values from the baseline to identify statistically significant deviations.

The first layer of our anomaly detection involves the evaluation of a comprehensive set of individual anomaly indicators. Each indicator corresponds to a specific metric and is triggered when a considerable decline from the baseline is detected. Our indicators are designed to be sensitive to specific types of changes, such as a rise in the median latency or a drop in path stability.

While the individual indicators are valuable to identify potential issues, they can be triggered by temporal network changes. Thus, a single indicator is no definitive evidence of the appearance of a network anomaly. Therefore, to increase the robustness of our detection and reduce false positives, we introduce a second layer of analysis that correlates the individual indicators to identify composite events. These events are predefined signatures of network problems. The simultaneous occurrence of several related indicators provides a much higher degree of confidence that a genuine and significant anomaly has occurred.

This structured approach allows us to combine statistical deviations into a robust assessment of network health. The subsequent sections detail the specific indicators and the logic for combining them into the composite events that form the final output of our analysis.

4.4.1 Indicators

Building upon the metrics described in Section 4.2.5, we develop a set of indicators to cover and detect a wide range of potential network anomalies. The indicators are derived by comparing the metrics from a set of data to an established baseline. An indicator is triggered when a statistically significant deviation from the baseline is detected.

We divide our indicators into statistical and path-related groups, mirroring the structure of our metrics. Furthermore, we apply statistical distribution tests, which serve as additional anomaly indicators when triggered.

Statistical Indicators

We design our statistical indicators to capture significant deviations in the performance characteristics of traceroute paths. To do so, we utilize fundamental statistical methods, primarily the median and standard deviation, to provide a robust assessment of performance decrease that is resilient to outliers.

High Median

A *High Median* indicator is triggered for the RTT, path length of successful traceroutes and the first hop Round-Trip-Time when the median of one of those key metrics substantially exceeds the median of that metric in the baseline. To determine when an increase in median is significant enough, we introduce a threshold T for each individual metric, which is calculated as the sum of the baseline median M_{base} and a sensitivity factor k_i multiplied by the standard deviation σ_{base} of the baseline. Each variable corresponds to the values of this metric.

$$T = M_{base} - k_i \cdot \sigma_{base} \tag{4.1}$$

The sensitivity factor k_i has a different value for each metric. We defined the factor for our Round-Trip-Time measurements to be $k_{\mathrm{RTT}} = 1.2$, the multiplier for the successful path length values to be $k_{\mathrm{PL}} = 1.4$ and for the first hop RTT to be $k_{\mathrm{FH}} = 2.0$. A higher k value makes the indicator less sensitive, requiring a more substantial deviation to trigger. We then compare the calculated threshold against the median of a metric from the current dataset.

$$M_{curr} < T \tag{4.2}$$

When this comparison holds true, the indicator is triggered.

High RTT Volatility

The indicator of *High RTT Volatility*, commonly referred to as jitter, focuses on the stability of the Round-Trip-Time values. We measure jitter based on the standard deviation of the RTT from successful measurements. An increase in the standard deviation of the Round-Trip-Times suggests that the latency of the traceroute packets is becoming less predictable and more inconsistent. We multiply a sensitivity factor k with the baseline RTT standard deviation σ_{base} to establish a threshold T.

$$T = k \cdot \sigma_{base} \tag{4.3}$$

We define the sensitivity factor to be $k = 1.4$. Subsequently, we compare the standard deviation of the Round-Trip-Time in our current dataset, σ_{curr}, to the

determined threshold to specify a trigger for our indicator.

$$\sigma_{curr} > T \tag{4.4}$$

Detecting a significant increase in jitter is a particularly effective indicator for detecting emerging network congestion before it necessarily impacts the median RTT.

Decreased Success Rate

A drop in success rate is a direct and critical indicator of a connectivity problem. The indicator is triggered when the success rate of the current dataset, S_{curr}, falls below the baseline rate, S_{base}, by more than a fixed absolute threshold of ten percent.

$$S_{curr} < S_{base} - 0.1 \tag{4.5}$$

Using a fixed threshold provides a clear and definite sign of degradation, ensuring that any significant drop in reachability is detected. The *Decreased Success Rate* indicator is particularly sensitive to emerging reachability issues, often providing early warning of developing problems before other metrics show significant changes.

Increased Timeout Rate

The timeout rate measures the proportion of traceroute measurements that contain at least one timeout, as detailed in Section 4.2.5. The indicator becomes active when the current timeout rate, denoted as t_{curr}, surpasses the baseline rate, represented as t_{base}, by more than a specified absolute threshold of ten percent.

$$t_{curr} > t_{base} + 0.1 \tag{4.6}$$

An *Increased Timeout Rate* points to unresponsive routers or packet loss along the network path. It serves as a complementary indicator to the success rate, providing an early warning of path degradation.

Decreased Path Length

This indicator detects a significant decrease in the median of the observed path length. While a shorter path might seem advantageous, it can be a critical symptom of a problem. This is because, even when traceroutes fail, the measurement data still contains a path with hops, although it may not be complete. Each failed traceroute provides an observable path, but this path will be shorter than the one from a successful traceroute with the same source and destination. Therefore, a sudden decrease in the median length of observed traceroute paths may suggest

that a significant number of measurements are failing to reach their destination. To determine when a decrease in path length is significant, we introduce a threshold T, calculated with baseline values about the observable path length median, M_{base}, and standard deviation, σ_{base}.

$$T = M_{base} - k \cdot \sigma_{base} \tag{4.7}$$

Factor k describes the sensitivity of the threshold. In our analysis, we chose a sensitivity factor of $k = 1.2$. The *Decreased Path Length* indicator is triggered when the median of our current dataset is lower than the threshold T.

$$M_{curr} < T \tag{4.8}$$

Path Indicators

We design our path indicators to detect routing changes within the network behavior. They are essential for identifying rerouting events, path instability and other topological changes in the network.

Increased Path Diversity

While some path diversity is normal, a sudden and severe increase indicates rapidly changing routes and consequently path instability. This can signal potential routing issues or load balancing changes. Path diversity is defined as the measure of how many different routes we can observe in our traceroute measurement data from source to destination. It is calculated as the ratio of unique paths P^U to the total number P^Σ of valid successful measurements in our dataset.

$$D = \frac{P^U}{P^\Sigma} \tag{4.9}$$

We categorize an increase in path diversity as sufficient for our analysis when we observe a substantial rise, both in relative and in absolute terms, simultaneously, in our comparison of the baseline and the current data. To do so, we check if the path diversity of our current dataset, described as D_{curr}, exceeds the diversity of the baseline, denoted as D_{base}, by more than our sensitivity factor of 80 %.

$$D_{curr} > D_{base} \cdot 1.8 \tag{4.10}$$

Additionally, we review whether the path diversity of our current data also surpasses the baseline by an absolute value of five percent.

$$D_{curr} > D_{base} + 0.05 \tag{4.11}$$

The indicator is triggered only when both these conditions are met at the same time. This dual condition approach prevents false positives in cases where baseline diversity is very low, ensuring that only significant and substantial increases in diversity are flagged.

Decreased Dominant Path Frequency

As detailed in Section 4.2.5, the dominant path is the single most frequently occurring route within a dataset of traceroute measurements. We observe and monitor the frequency of the number one dominant path over all valid and successful routes. The *Decreased Dominant Path Frequency* indicator is triggered when the frequency of the dominant path in the current dataset, declared as F_{curr}, significantly drops below the baseline frequency F_{base}.

$$F_{curr} < F_{base} \cdot k \tag{4.12}$$

Our sensitivity factor $k = 0.7$ denotes a decrease in dominant path frequency of 30 %. This relative comparison effectively measures path stability, regardless of the absolute frequency of the dominant path. The indicator suggests that traffic is being routed away from the primary, previously stable, route and to alternative paths. Consequently, a *Decreased Dominant Path Frequency* shows a loss of stability and could be an early indicator of a rerouting event of increased instability.

Dominant Path Shift

A *Dominant Path Shift* occurs when the sequence of hops in the dominant path differs between the baseline and the current dataset. Therefore, the indicator is triggered when a different path emerges as dominant, regardless of whether the overall performance metrics show any sign of deterioration. This can be an explicit sign of an emerging rerouting event.

Core Path Shift

While the dominant path considers the full route from source to destination, a core path represents a stable segment of the network, as detailed in Section 4.2.5. We identify and monitor the ten most significant and recurring core paths and their frequencies in each of our datasets. The *Core Path Shift* indicator triggers when there are major differences in the set of core path segments between the baseline and the current dataset. To ensure statistical significance, we employ a minimum support threshold of five percent for a segment to be considered a major core path. Only core path segments with an appearance rate of at least our support threshold are considered by our indicator. A *Core Path Shift* can manifest

in two ways. Either as a major core path segment that was present in the baseline disappears or a new one emerges in the current data. The indicator is crucial for detecting significant changes in routing within core backbone networks, which might not be visible by analyzing full end-to-end paths alone. This is especially important for traceroute measurements with complex networks and routing with a high diversity.

Statistical Distribution Tests

A fundamental element of our anomaly detection is the statistical comparison of Round-Trip-Time distributions over time. While simple, yet robust, metrics like the median RTT can indicate performance shifts, they may fail to find more subtle changes in network behavior, such as increased jitter. To better detect anomalies, we apply non-parametric statistical tests to compare the entire distribution of RTT values from the traceroute measurement we are currently analyzing against our established baseline. This approach allows us to detect changes in the shape, spread and tails of the entire latency distribution. These differences can often be direct indicators of network anomalies.

We specifically employ two distinct and separate distribution tests for our analysis to address all parts of the entire distribution. The two-sample Kolmogorov-Smirnov test, as outlined in Section 3.6, serves as a robust and general-purpose tool. It is sensitive to any significant deviation between the current and baseline RTT distributions, whether in central tendency or in the overall shape. The k-sample Anderson-Darling test, detailed in Section 3.7, is used in parallel to identify significant changes in the tails of the distribution between the RTT values of baseline and current data. This sensitivity to changes in the tails of the distribution is essential for detecting emerging performance issues. A shift in high RTT values towards the tails can be a valuable indication of network anomalies such as traffic congestion or suboptimal routing. Using both tests, we can flag a broader, more subtle range of performance profile shifts.

In our analysis, we establish a baseline by collecting a representative and stable set of traceroute measurements during a period of consistent network traffic behavior, as detailed in Section 4.3. For each subsequent dataset analysis, the newly collected RTT values are used as the second sample for each distribution test. The two-sample K-S and the k-sample A-D tests are fed the baseline and new data. The null hypothesis for both tests is that the current sample is drawn from the same distribution as the baseline. A p-value below our defined significance level of 0.05 signals a statistically significant change.

Although statistical significance is critical, it does not always reflect operational significance. Network measurements can be inherently noisy, leading to minor, short-term fluctuations that may yield statistically significant results without indicating a genuine network problem. We employ a secondary check based on the degree of impact to reduce the risk of false positives. A change in distribution is only considered a valid anomaly indicator if two conditions are satisfied. First, the already mentioned p-value of the K-S test or the A-D test, respectively, must be below the significance threshold. Secondly, the absolute difference between the median of the RTT of the current sample and the baseline sample must exceed a minimum increase of two milliseconds. This approach ensures that our application of distribution tests only considers changes that are both statistically significant and practically relevant in terms of perceived performance. Consequently, keeping the false-positive rate as low as possible provides a much higher degree of confidence that a detected distribution shift represents a genuine and meaningful network event.

Finally, an identified distribution shift alone is not treated as a definitive anomaly yet, but as an indicator of such. As further explained in Section 4.4.2, it is a piece of evidence combined in a correlation method, combining multiple indicators to create correlated events, highlighting possible anomalies.

4.4.2 Composite Events

While the individual indicators described in Section 4.4.1 are effective at discovering and recognizing specific deviations, a single indicator cannot definitively detect an anomaly. Therefore, we correlate multiple indicators to identify and classify complex composite anomaly events.

These events characterize well-defined network behaviors that arise from the simultaneous occurrence of specific sets of indicators. This approach allows us to implement a sophisticated diagnosis of network problems and significantly increases the confidence in the certainty of our anomaly detection. Consequently, our method also reduces the likelihood of false positives in our analysis.

The following sections detail the specific composite events we design to detect anomalies. This includes an explanation of the various combinations of indicators for each event, their significance and the resulting impact on the network. An overview of all events and their respective indicators can be seen in Table 4.1.

Indicator Condition	Composite Events				
	Major Rerouting Event	**Perf-Impacting Reroute**	**Path Instability (Flapping)**	**Perf. Profile Shift**	**Early Path Failure**
Performance Decrease	✅	✅			
High Median	☑	☑			
High RTT Volatility	☑	☑	✅		
Distribution Test Alert	☑	☑		✅	
Path Shift		✅			
Dominant Path Shift	✅	☑		❌	
Core Path Shift	✅	☑		❌	
Increased Path Diversity			✅		
Decreased Dom. Path Freq.			✅		
Decreased Success Rate					✅
Decreased Path Length					✅

Table 4.1: Composite Events

Major Rerouting Event

The *Major Rerouting Event* represents the most significant type of topological change, confirmed by a significant performance impact. It is defined as the combination of at least three separate specific indicators being triggered.

Firstly, at least one statistical performance degradation indicator must be present. This can be a *High Median* in any of the key metrics of the Round-Trip-Time, the successful path length or the first hop RTT, as detailed in Section 4.4.1. Once one of those metrics displays a *High Median*, the indicator is triggered, and therefore, a performance decrease is present. Furthermore, a *High RTT Volatility* or a significant change in the Round-Trip-Time distribution, identified when at least one *Statistical Distribution Test* is positive, as described in Section 4.4.1, also displays a decline in statistical performance. This ensures that we only flag reroutes that negatively impact the operational performance.

Additionally, a *Dominant Path Shift* must exist to set off this event. Consequently, we know that the most frequently observed path in the current analysis data is different from the dominant path identified in the baseline, as described in Section 4.4.1. This indicator is a clear sign that a reroute has occurred. However, a change in the dominant path alone is not significant enough to confirm a *Major Rerouting Event*.

Finally, a *Core Path Shift* has to be present as well. As described in Section 4.4.1, this indicator is triggered if a previously stable core segment disappears or if a new one

emerges with significant support. Consequently, we identified a change in the backbone network.

The requirement for both path shift indicators, a *Dominant Path Shift* alongside a *Core Path Shift*, ensures that we are observing a substantial rerouting event rather than a localized last-mile change. The simultaneous performance degradation confirms the operational significance of the event, filtering out reroutes with benign or neutral effects on the performance. This combination of topological and performance indicators provides us with a very high confidence that a significant reroute, with a performance impact, is present in our traceroute measurement data.

The *Major Rerouting Event* is particularly valuable for detecting considerable infrastructure changes, such as the deactivation or failure of network links, changes in routing policies or large-scale network outages that force traffic through alternative paths.

Performance-Impacting Reroute

The *Performance-Impacting Reroute* is a more general rerouting event. It describes any topological change in combination with some kind of negative effect on performance.

To identify this event, we require any statistically significant degradation in performance to be detected within our current dataset. As described in more detail in the *Major Rerouting Event* description, a performance decrease is present when at least one of the indicators for *High Median*, a *High RTT Volatility* or a *Statistical Distribution Test* is triggered.

Furthermore, any path shift must be occurring in our traceroute measurements. This means that the activation of either the *Dominant Path Shift* or *Core Path Shift* indicator, detailed in Section 4.4.1, is already a substantial change in topology for a *Performance-Impacting Reroute* to be possible.

The requirement for some kind of topological change in combination with any statistical performance degradation allows us to identify relatively minor routing changes that still result in a measurable performance impact. This combination gives us high confidence that an anomaly causing a reroute with a negative operational effect exists in our current dataset.

The *Performance-Impacting Reroute* is crucial for detecting subtle but impactful routing changes, such as small-scale network failures or ones affecting only a subset of traffic.

Path Instability

The *Path Instability*, commonly referred to as flapping, characterizes a network state where routing shows high temporal variability, with frequent and unpredictable changes between different paths. It represents an event where no single dominant path can establish itself, but the paths taken by packets to a destination change rapidly and repeatedly. This does not arise from a stable reroute but rather from a chaotic instability between paths. We detect *Path Instability* with three specific indicators.

First of all, an *Increased Path Diversity* serves as the initial indicator, signaling that significantly more different paths are being used than in the stable baseline dataset, as detailed in Section 4.4.1. This suggests that the traceroute packets are being forced to use multiple alternative paths. However, an increase in diversity alone is insufficient, as it could simply represent the addition of multiple new and stable paths.

Therefore, we also require a *High RTT Volatility*, which identifies a significant increase in the standard deviation of the Round-Trip-Time, as detailed in Section 4.4.1. This is a good indication of path flapping. As network packets choose differentiating routes with varying latencies, the resulting RTT values become highly inconsistent.

Finally, a *Decreased Dominant Path Frequency* must be present. In a stable network, one path typically carries a large majority of the traffic. When this frequency drops significantly, it implies that the primary route is less reliable, forcing traffic onto alternative, less optimal paths.

The combination of these indicators provides a robust composite event for path flapping. Therefore, the simultaneous emergence of all three indicators provides us with a high degree of confidence that we detected an anomaly causing *Path Instability*.

This composite event is essential for identifying potential routing or connectivity issues and performance inconsistencies that arise from a highly dynamic and unstable state rather than a persistent one.

Performance Profile Shift

We design the composite event of the *Performance Profile Shift* to detect subtle but substantial changes in network performance that cannot be attributed to a change in routing. It is detected when a significant deviation in the RTT distribution is present while any topological shift is absent.

Therefore, the key indicator for this event is the *Statistical Distribution Test*, which is positive when either the Kolmogorov-Smirnov Test or the Anderson-Darling Test yields a p-value of below our significance threshold, as described in Section 4.4.1. These

tests are sensitive to changes in the variance of the entire distribution as well as in the tails. This indicator can detect shifts like the emergence of a second peak in the RTT distribution, implying that packets are being delayed on the path, suggesting the development of congestion.

Furthermore, a *Performance Profile Shift* is only triggered if the network topology is stable. Thus, there must not be a *Dominant Path Shift* or a *Core Path Shift*, which are described in Section 4.4.1. This condition distinguishes the event as one that is exclusively based on performance. The stability of the path implies the cause of the performance change is on the existing route.

The specification of a performance change along the Round-Trip-Time distribution, combined with the absence of a path shift, allows us to detect subtle performance degradations on the frequently used paths. Thus, when this event occurs, we can be reasonably confident in detecting an anomaly.

The *Performance Profile Shift* serves as an early warning for developing issues like network congestion or hardware problems along a stable route.

Early Path Failure

An *Early Path Failure* suggests that traceroutes consistently fail to reach their destination and terminate at a hop along the route closer to the source than in the baseline. It is identified by two distinct indicators.

Firstly, a *Decreased Path Length* shows that the hop count of all routes is significantly lower than the baseline. This includes successful measurements and those that did not reach their destination, as described in Section 4.4.1. This decrease alone would imply an improvement in network health. Therefore, we must combine it with a second, correlated indicator.

We use the *Decreased Success Rate*, detailed in Section 4.4.1, to confirm that the *Decreased Path Length* is due to packets not reaching their destination, rather than discovering shorter routes.

A low observed path length might suggest that the destination is now topologically closer. However, if there is also a significant decrease in the percentage of traceroutes that receive a response from the destination, it strongly indicates that the shorter paths are due to failure rather than increased efficiency.

This composite event is particularly valuable for identifying connectivity problems that are likely affecting a large number of measurements and might otherwise be lost in the noise of the overall Round-Trip-Time statistics. Therefore, an *Early Path Failure* gives

us a high confidence that a substantial reachability issue has emerged in our traceroute data. It can indicate a routing black hole, a faultily reconfigured firewall or a complete outage of a device early in the network path.

Chapter 5

Analysis and Results

In this chapter, we apply the methodology detailed in Chapter 4 to real-world network events. First, we describe a validation approach that we apply to each of our analyses. Subsequently, each section presents a case study of a known or suspected network anomaly. We detail the specific setup, including the time periods and geographical filters used for the analysis. We then present the findings, demonstrating the efficacy of our system in detecting and classifying different types of network anomalies.

5.1 Validation Methodology

To ensure the validity of our findings and the stability of our baseline, we employ a strict validation methodology using a control dataset. We choose the check data to be comparable to the baseline, with both periods starting on the same weekday, stretching over the exact same timespan and ending on the same day of the week. Furthermore, we apply the exact same filters to both sets of traceroute data.

Most importantly, we perform a comparison by analyzing data of the reference period against the baseline. Subsequently, we assess the baseline against the values of our control. This confirms that no significant deviation is occurring in the baseline and check traceroute datasets. This validates the integrity and stability of our baseline. Furthermore, we choose our baseline period to be before a suspected incident, while selecting a timespan after the restoration of the event for our control data. This allows us to verify that the connection state is back to normal and there are no lasting effects on network health.

Additionally, we use the control period as an alternative baseline for the main analysis of the outage itself. Therefore, we compare the data during an incident against the baseline, set before the event, and the check baseline, determined from data after the network restoration. Consequently, we can confidently attribute any detected deviations to the outage event itself, rather than to unrelated or short-term shifts in network behavior.

5.2 The 2025 Iberian Peninsula Outage

This case study examines a widely reported power outage that primarily affected the Iberian Peninsula on the 28$^{\text{th}}$ of April 2025. The incident itself was active for around 15 hours from 10:00 UTC. The grid was fully restored at around 02:00 UTC of the following day. The purpose of this analysis is to examine the impacts on network infrastructure during the recovery from a large-scale power outage, as machines begin to restore connectivity while the network still experiences partial blackouts. We aim to demonstrate the efficacy of our approach in detecting and characterizing the impact of a significant, real-world network disruption across multiple geographical scopes.

5.2.1 Setup

To conduct a thorough and robust analysis of the event, we define three distinct time periods.

Baseline

We choose a period of two weeks prior to the event, from the 10$^{\text{th}}$ of April 2025, to the 26$^{\text{th}}$ of April 2025, to establish a stable and representative performance baseline.

Outage

We analyze the time from the start of power recovery at around 13:00 UTC until the end of the 28$^{\text{th}}$ of April 2025.

Check

We select a span from the 1$^{\text{st}}$ of May 2025 until the 17$^{\text{th}}$ of May 2025 as a control dataset after the incident had been solved. We choose the check data similar to the baseline, with both time periods starting on a Thursday, stretching 17 days and ending on a Saturday.

Additionally, we define a comprehensive set of geographical traffic corridors. Our analysis only selects measurements with either the source or the destination address located in the Iberian Peninsula. To determine the associated records, we filter our traceroute

data for probes in either Spain or Portugal. To understand the scope and impact of the outage, we investigate network traffic between this region and several key regions in Europe.

- Iberia Internal: Traffic within and between Spain and Portugal, analyzing the traffic within the outage area.
- British Isles: Traffic between the Iberian Peninsula and the British Isles, Great Britain and Ireland. This traffic routes through a number of submarine cables to the north of the Iberian Peninsula.
- France: Traffic between Spain and Portugal with France, investigating the traceroute data with a direct neighboring country.
- Central Europe: Traffic between the Iberian Peninsula and a group of European countries, namely Germany, Belgium, Switzerland as well as Austria. These packets have a number of options to reach their target.
- Italy: Traffic between the Iberian Peninsula and Italy, analyzing measurements through undersea cables to the southwest of the affected area.

For each location, we conduct analysis in both directions to obtain a complete view of the impact on both inbound and outbound traffic. As a foundational step, all measurements are filtered to be exclusively between RIPE Atlas anchors to ensure the highest degree of data quality and reliability.

5.2.2 Results

Composite Event	Iberia Internal	British Isles		France		Central Europe		Italy	
		To	From	To	From	To	From	To	From
Major Rerouting	✅	✅							
Perf-Impacting Reroute	✅	✅		✅		✅		✅	✅
Path Instability									✅
Perf. Profile Shift									
Early Path Failure									

Table 5.1: Iberian Peninsula Outage 2025: Composite Events

Based on our results, we discuss the measurements within Spain and Portugal, between the Iberian Peninsula and the British Isles as well as between Iberia and Italy separately. In our investigation of traffic to and from France and Central Europe, we present our

findings jointly. An overview of all anomalies we identified is shown in Table 5.1, categorized in our composite events.

Iberia Internal

Anomaly Indicator	Iberia Internal
Performance Decrease	✅
High Median	
High RTT Volatility	✅
Distribution Test Alert	
Path Shift	✅
Dominant Path Shift	✅
Core Path Shift	✅
Increased Path Diversity	✅
Decreased Dom. Path Freq.	
Decreased Success Rate	✅
Decreased Path Length	

Table 5.2: Iberian Peninsula Outage 2025: Indicators within Iberia

We gathered traceroute records of 23 anchors within Spain and Portugal for our analysis of the internal traffic during the outage timeframe. For our baseline, we collected the data of 1 555 150 measurements, comparing them with the 25 712 traceroute results from the much shorter time period during the outage. A detailed list of all values from our analysis can be seen in Appendix A.1.1.

Our analysis reveals a significant impact of the incident on the traffic within the Iberian Peninsula. The triggered indicators, shown in Table 5.2, demonstrate a *Major Rerouting Event* with a substantial performance degradation. The metrics of our control period further strengthen our results, demonstrating the reliability of our baseline.

The most noticeable observation is the considerable drop in the success rate of our measurements, decreasing from 78.33 % in the baseline period to 56.73 % during the investigated incident. This shows that a large part of the traffic did not reach its destination, exceeding the threshold and therefore confirming our corresponding anomaly indicator.

Additionally, the median length of all routes increased from 9 to 11 hops, with the successful path length increasing from 8 to 9 hops. This suggests that the traffic was

trying to find alternative, longer routes. This is also supported by the triggered *Increased Path Diversity* indicator, resulting from the spike in the relative number of unique to total valid paths. While the baseline maintains a low path diversity of 1.03 %, our outage period demonstrates a remarkably high value of 15.77 %. Figure 5.1 illustrates the path length in hourly aggregates, comparing the baseline to our incident data.

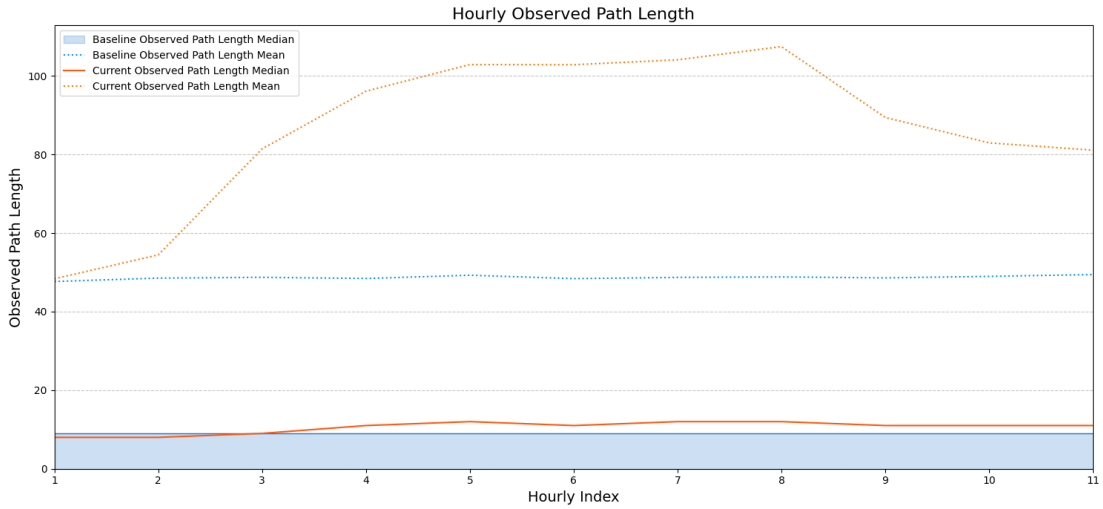


Figure 5.1: Iberian Peninsula Outage 2025: Hourly Observed Path Length within Iberia
Data from 2025-04-28 13:00 to 2025-04-28 23:59 UTC (Analyzed Outage Period)

The increased variety in routing also has a measurable impact on the Round-Trip-Time, as can clearly be seen from the growth in standard deviation (std dev) from 21.16 ms to 29.82 ms. This significant difference between the RTT std dev values leads to a *High RTT Volatility*, displaying a decrease in performance.

Reviewing the dominant and core paths verifies our deduction from the statistical values, triggering a *Dominant Path Shift* as well as a *Core Path Shift*.

The combination of both path shifts, along with the increased path diversity, results in not only a *Performance-Impacting Reroute* but also the more substantial *Major Rerouting Event*. This confirmed anomaly is accompanied by additional indicators, presenting a *Decreased Success Rate* and an *Increased Path Diversity*.

British Isles

Anomaly Indicator	Iberia to British Isles	British Isles to Iberia
Performance Decrease	✅	
High Median		
High RTT Volatility	✅	
Distribution Test Alert		
Path Shift	✅	✅
Dominant Path Shift	✅	✅
Core Path Shift	✅	✅
Increased Path Diversity	✅	✅
Decreased Dom. Path Freq.		
Decreased Success Rate		✅
Decreased Path Length		

Table 5.3: Iberian Peninsula Outage 2025: Indicators between Iberia and the British Isles

Our analysis of traffic between the Iberian Peninsula and the British Isles compared the values from around 2.5 million traceroute measurements in our baseline per direction to almost 40 thousand records to Great Britain and Ireland as well as more than 61 thousand outbound traceroutes from Spain and Portugal during the outage period, respectively. For our analysis, we gathered the data from the traffic between 23 anchors in Iberia and 74 anchors in the British Isles in our individual timeframes. The detailed metrics are available in Appendix A.1.2. The values of our check duration also validate the stability of the baseline and confirm the deviations after a full recovery of the connectivity state.

Our investigation of the datasets reveals an asymmetrical impact on network health during the incident. Table 5.3 provides an overview of our identified anomaly indicators for each direction. While outbound traffic from the Iberian Peninsula to the British Isles shows a *Major Rerouting Event* with an impact on performance, inbound measurements seem to explore alternate routes but eventually fail to reach their destination in many cases.

In either direction, we observe an *Increased Path Diversity*, increasing from 2.25 % to 22.62 % for outgoing and from 1.27 % to 19.17 % for incoming traffic. Hence, greatly surpasses our sensitivity factor and triggers our corresponding anomaly indicator.

Furthermore, we detect a shift in topology for both outbound as well as inbound traceroute measurements. This change manifests itself as a *Dominant Path Shift* in conjunction with a *Core Path Shift* in each analysis. Consequently, we can clearly identify a rerouting event occurring in either direction.

For traffic from Spain and Portugal to Great Britain and Ireland, we observe a significant gain in the standard deviation of our Round-Trip-Time values. With an increase from 20.89 ms in the baseline to 33.13 ms during the incident, we have a definite performance decrease in the form of *High RTT Volatility*. Figure 5.2 outlines the RTT standard deviation over the single hours of the day.

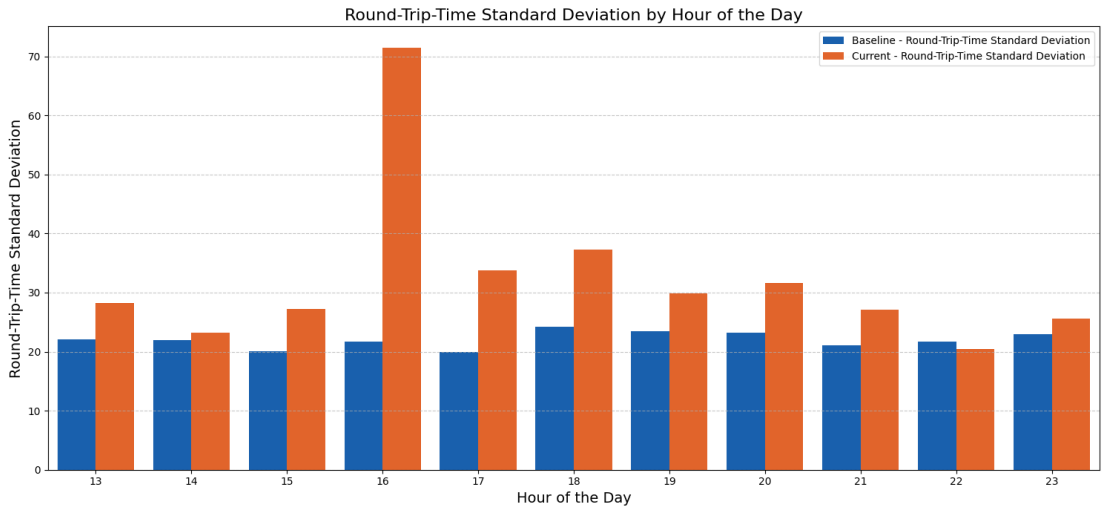


Figure 5.2: Iberian Peninsula Outage 2025: RTT standard deviation - Iberia to British Isles

The spike at 16:00 UTC aligns with the hour the first major power line was re-energized [36].

The recognized jitter in combination with the significant path shift illustrates an anomaly categorized as a *Major Rerouting Event* and consequently also results in the less severe *Performance-Impacting Reroute*.

In the opposite direction, the effect manifests itself in a different way. While the impact on the success rate for outgoing traffic was relatively moderate, it dropped significantly from 76.13 % to 53.61 % for traffic from the British Isles to Iberia. Figure 5.3 illustrates the success rates for baseline and the outage in hourly aggregates.

Figure 5.3: Iberian Peninsula Outage 2025: Hourly Success Rate - British Isles to Iberia
Data from 2025-04-28 13:00 to 2025-04-28 23:59 UTC (Analyzed Outage Period)

This degradation in success to reach the destination, in combination with the increased path diversity and change in topology, suggests that traffic chooses alternative, less reliable routes.

France and Central Europe

Anomaly Indicator	Iberia to France	France to Iberia	Iberia to Central Europe	Central Europe to Iberia
Performance Decrease	✅		✅	
High Median				
High RTT Volatility	✅		✅	
Distribution Test Alert				
Path Shift	✅	✅	✅	✅
Dominant Path Shift	✅	✅	✅	✅
Core Path Shift		✅		✅
Increased Path Diversity	✅	✅	✅	✅
Decreased Dom. Path Freq.				
Decreased Success Rate		✅		✅
Decreased Path Length				

Table 5.4: Iberian Peninsula Outage 2025: Indicators between Iberia and France & Central Europe

Due to the similarities in metric deviations and topological observations in our analysis of traceroutes from the Iberian Peninsula to France and Central Europe, we present the results together.

For our analysis, we utilized measurements from 23 anchors in Iberia, 80 anchors in France and 272 anchors across our investigated countries in Central Europe, respectively. This resulted in massive amounts of records, counting almost 3 million records per direction for the baseline with France and approximately 10 million traceroutes each with Central Europe. The datasets for our incident period consist of almost 50 thousand measurements from Iberia to France and close to 80 thousand traceroutes in the opposite direction. From the Iberian Peninsula to Central Europe, we analyzed 171 516 entries of outbound and 272 977 measurements of inbound traffic in the timeframe of our outage. The detailed values of our analyses are presented in Tables A.1.3 and A.1.4 in the Appendix.

Our results reveal similar asymmetrical patterns for both France and Central Europe, identifying the same anomaly indicators per traffic direction in each geographical location, as shown in Table 5.4.

While the outbound traffic from Iberia to either region experiences a *Performance-Impacting Reroute*, inbound traceroutes appear to be routed along unstable alternative paths, displaying difficulties in reaching their destination.

Each of our analyses between the regions finds a substantial spike in path diversity, rising to around ten times the baseline value in each individual direction and destination. Thus, triggering the *Increased Path Diversity* indicator.

For both France and Central Europe, we identify a path shift in either direction. However, the topological change in traffic towards the Iberian Peninsula is notably more significant, demonstrating a *Core Path Shift* on top of the more common *Dominant Path Shift*.

Furthermore, outbound traceroutes demonstrate a considerable performance decline, manifested as *High RTT Volatility*. This indicator results from an increase in the standard deviation of the Round-Trip-Time by approximately 57 % for each of the two destinations.

Meanwhile, inbound traffic to Iberia from both source areas reveals a significantly *Decreased Success Rate*. Either dataset shows a degradation of around 23 %, surpassing the threshold value for our indicator.

The combination of a performance decrease and a moderate rerouting event in our data of outbound measurements identifies an anomaly in the form of a *Performance-Impacting Reroute*. Although we did not identify a specific composite event in our analysis of inbound traffic, the anomaly indicators suggest that packets are trying to find alternative paths to their destinations but, in many cases, fail to reach them.

Italy

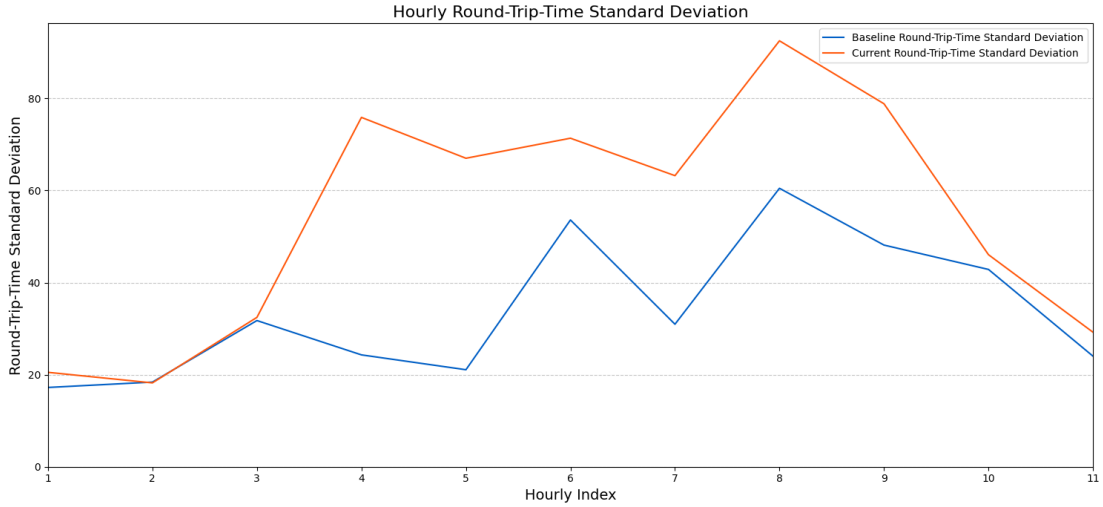
Anomaly Indicator	Iberia to Italy	Italy to Iberia
Performance Decrease	✅	✅
High Median		
High RTT Volatility	✅	✅
Distribution Test Alert		
Path Shift	✅	✅
Dominant Path Shift	✅	
Core Path Shift		✅
Increased Path Diversity	✅	✅
Decreased Dom. Path Freq.		✅
Decreased Success Rate		✅
Decreased Path Length		

Table 5.5: Iberian Peninsula Outage 2025: Indicators between Iberia and Italy

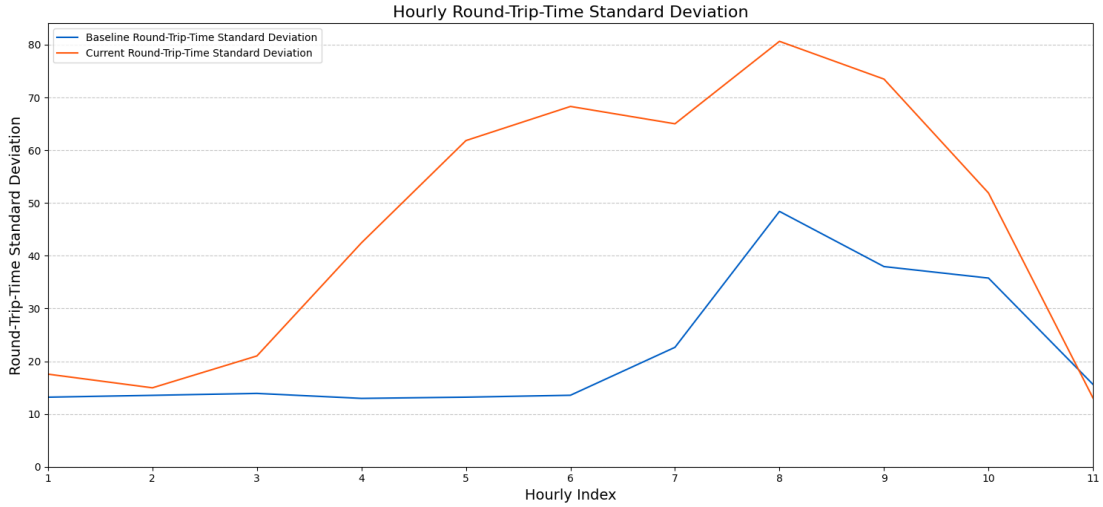
We gathered data from traceroutes between 23 anchors within the Iberian Peninsula and 32 anchors in Italy. The outbound traffic from Iberia to Italy resulted in over 1.3 million measurements in our baseline and around 23 thousand records in our outage period. The inbound dataset amounts to more than 1.4 million traceroutes in the baseline and approximately 38 thousand records in the incident data. The metrics of our analysis are detailed in Appendix A.1.5.

Our results reveal that the analyzed connection experienced a *Performance-Impacting Reroute* in both directions. The traffic routing from Italy to the Iberian Peninsula additionally exhibits *Path Instability*. A comprehensive view of all anomaly indicators is shown in Table 5.5. Our findings are further reinforced by the values of our control period after the network has fully recovered from the event.

In both directions, we observe a significant performance decrease, with the standard deviation of the Round-Trip-Time displaying more than double the baseline value. Consequently, this triggers the anomaly indicator for jitter, *High RTT Volatility*. Figure 5.4 presents the RTT standard deviation for the duration of our analysis, compared to the baseline in hourly aggregates.



(a) The Iberian Peninsula to Italy



(b) Italy to the Iberian Peninsula

Figure 5.4: Iberian Peninsula Outage 2025: Hourly RTT std dev between Iberia and Italy
Data from 2025-04-28 13:00 to 2025-04-28 23:59 UTC (Analyzed Outage Period)

Furthermore, we notice a decrease in the success rate, showing a minor decrease of around 4 % to 76.86 % in outbound traffic. More significant is the drop in the success of traceroutes reaching their destination for inbound traffic, dropping from 78.07 % to only 54.16 %, thus triggering the *Decreased Success Rate* anomaly indicator.

Additionally, we observe a topological change in each direction, accompanied by an *Increased Path Diversity*, indicating the shift onto many alternative paths. The route from the Iberian Peninsula to Italy shows a *Dominant Path Shift* with a path diversity of 21.77 % from the previously 1.95 % in the baseline. The opposite direction demonstrates a *Core Path Shift* with an increase in route variety from 1.08 % to 18.84 % during our investigated incident.

Finally, the inbound traffic to Iberia shows a decrease in dominant path frequency to around one-third of the stable value in the baseline, triggering the *Decreased Dominant Path Frequency* indicator. This further suggests traffic frequently changing paths, not stabilizing on a primary and stable dominant route.

The detected change in routing, in combination with the performance degradation, confirms a bidirectional *Performance-Impacting Reroute* during our analyzed timeframe. Our analysis of the measurements from Italy to the Iberian Peninsula additionally finds a *Path Instability*. This composite event results from the *High RTT Volatility* in conjunction with an *Increased Path Diversity* and *Decreased Dominant Path Frequency*. The control period after the network completely recovered to full health verifies our observations with stable values, similar to the baseline.

5.3 C-Lion1 Incident of November 2024

In November of 2024, a fault was reported in the C-Lion1 submarine communication cable connecting Finland and Germany. Our analysis aims to identify and characterize the performance degradation on this specific Baltic Sea route. A depiction of the submarine cable C-Lion1 and the location of the damage can be seen in Figure 5.6 [3].

Figure 5.6: C-Lion1 November 2024 Incident

5.3.1 Setup

The analysis is structured around three distinct time periods to isolate the incident from normal network behavior.

Baseline

We define a period of ten days from the 4^{th} of November 2024, to the 14^{th} of November 2024, to establish our performance baseline.

Outage

The period of the reported incident was from the 18^{th} of November 2024 to the 28^{th} of November 2024.

Check

We choose the time from the 2nd of December 2024 until the 12th of December 2024 for our control dataset to validate the baseline and confirm the outage.

To specifically target traffic utilizing the C-Lion1 cable, we add constraints for both ends of our traceroute dataset. For the first one, we restrict the anchors to be in a circle of 375 km around Helsinki and additionally filter for addresses located in Finland only. The second end is restrained to anchors in Germany in an area of 660 km around Rostock. We perform the analysis bidirectionally, examining the traffic from Finland to Germany and vice versa to ensure a comprehensive assessment.

5.3.2 Results

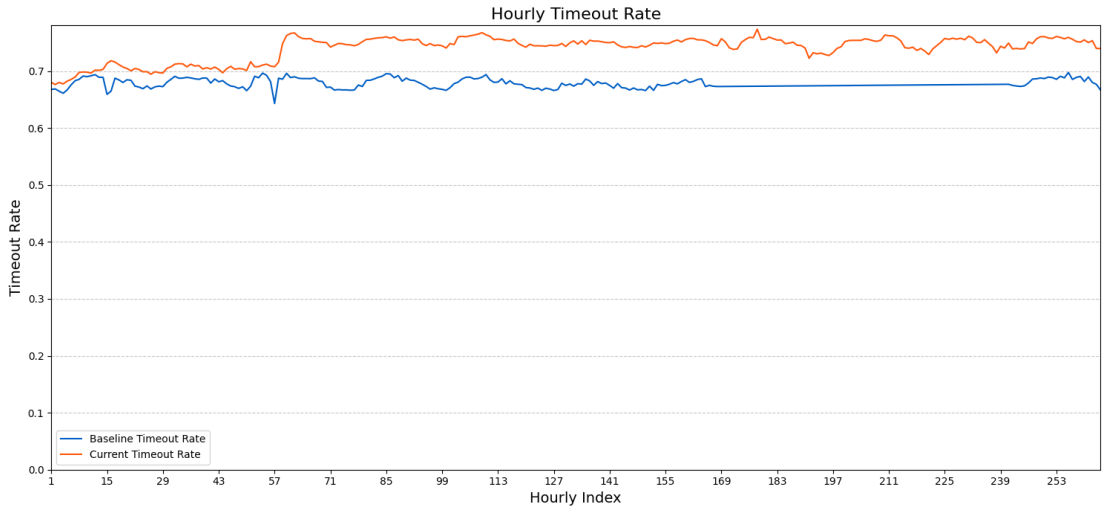
Anomaly Indicator	Finland to Germany	Germany to Finland
Performance Decrease		
High Median		
High RTT Volatility		
Distribution Test Alert		
Path Shift	✅	✅
Dominant Path Shift	✅	✅
Core Path Shift	✅	✅
Increased Path Diversity		
Decreased Dom. Path Freq.		
Decreased Success Rate		✅
Decreased Path Length		

Table 5.6: C-Lion1 November 2024: Indicators between Finland and Germany

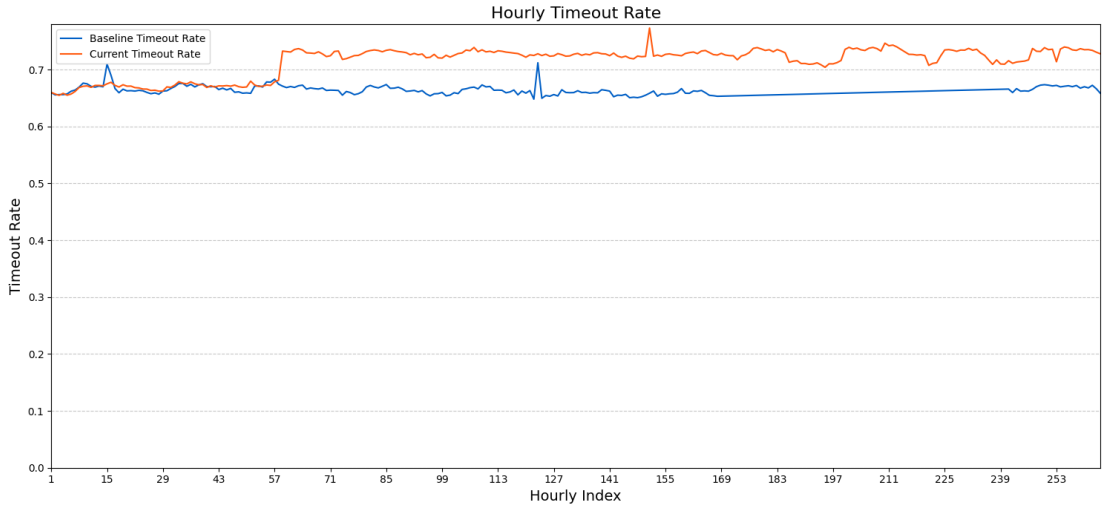
In our analysis, we identified 18 anchors in Finland and 177 anchors in Germany that fit our criteria. From these probes, we performed analyses between those two countries. In each direction, our selected time periods contained around two million traceroute measurements in our individual datasets for baseline, outage and control. A comprehensive view of the metrics can be found in Appendix A.2.

The results reveal an asymmetric impact on network health and topology between Finland and Germany. While the degradation was not severe enough to trigger one of our composite events, the indicators point to a rerouting event with a measurable effect on network health in the traffic from Germany to Finland. Table 5.6 shows a summary of the triggered indicators.

Our investigation shows an increase in the median observed path length from 11 to 12 hops and a corresponding rise in the timeout rate by more than 5 % in both directions. Figure 5.7 displays a comparison of the timeout rates between the baseline and the outage aggregated in hourly buckets for the whole timeframe of our analysis.



(a) Finland to Germany



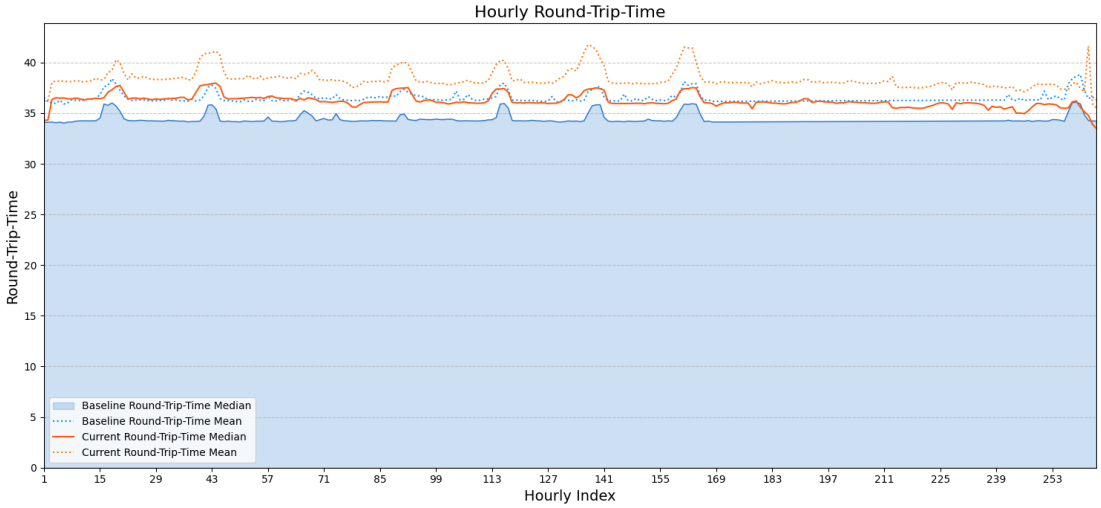
(b) Germany to Finland

Figure 5.7: C-Lion1 November 2024: Hourly Timeout Rates

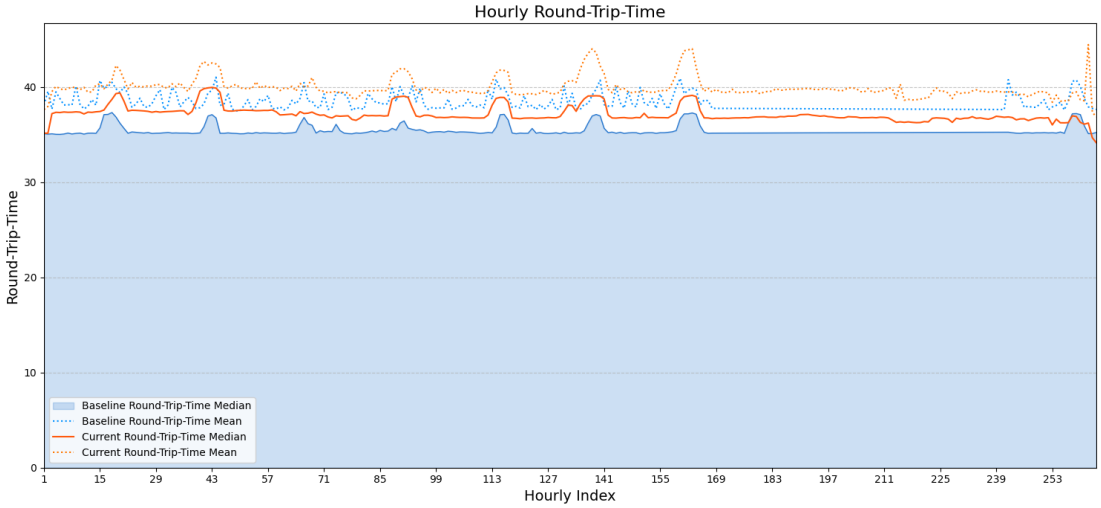
Data from 2024-11-18 00:00 to 2024-11-28 23:59 UTC (Analyzed Outage Period)

This increase suggests that the traffic was being routed through alternative paths that were less reliable and marginally longer. Additionally, we identify a topological change

in the dominant and core paths, triggering the corresponding indicators and confirming a rerouting event.



(a) Finland to Germany



(b) Germany to Finland

Figure 5.9: C-Lion1 November 2024: Hourly Round-Trip-Times
Data from 2024-11-18 00:00 to 2024-11-28 23:59 UTC (Analyzed Outage Period)

Furthermore, we can observe a mild rise in the median Round-Trip-Time in both directions of the analysis, depicted in Figure 5.9. While the increase of around 2 ms respectively might not be enough to cross the sensitivity threshold required to flag a *High Median*, it displays a consistent change in performance. This strengthens our assumption of a reroute onto a slightly longer path.

Additionally, we notice a decrease in success rate, dropping from 83.89 % to 77.75 % in our route from Finland to Germany. Similarly, in the other direction, we detect a change from 92.98 % to 79.07 %, displaying a significant degradation in successful traceroutes and triggering the *Decreased Success Rate* indicator. Figure 5.11 shows the success rate of the baseline and incident for traffic from Germany to Finland in hourly aggregates.

Figure 5.11: C-Lion1 November 2024: Hourly Success Rate - Germany to Finland
Data from 2024-11-18 00:00 to 2024-11-28 23:59 UTC (Analyzed Outage Period)

This provides clear evidence that the traffic was being routed over more unstable and less efficient alternative paths.

While our control period clearly manifests the stability and reliability of our baseline, we can observe the timeout rate as well as the success rate, maintaining similar values to the period of the incident. At the same time, the median Round-Trip-Time of the check data returned to the level of the baseline after the incident was reported to be resolved. This is also apparent from Figures 5.7 and 5.11 holding their values, while Figure 5.9 showcases a recovery at the end of the outage. While we observe a significant shift in the timeout and success rate in our graphs at 10:00 UTC on the 20$^{\text{th}}$ of November, we could not link this deviation to any real-world factors during the reported incident.

Our comprehensive analysis of the C-Lion1 incident from November 2024 highlights the robustness of Internet infrastructure. During the time of the damage to this crucial submarine cable, we were able to observe a rerouting event that successfully handled the disruption. Our results show no significant impact on the performance of network traffic. This demonstrates the ability of modern Internet architecture to reliably maintain connectivity during critical infrastructure failures.

5.4 BCS East-West Interlink Incident

Concurrent with the C-Lion1 incident, there was a disruption in the BCS East-West Interlink. This submarine cable in the Baltic Sea connects Lithuania and the Swedish island of Gotland. On the 18th of November 2024, a cut in the communication cable was announced. A figure with the BCS East-West Interlink can be seen in Figure 5.12 [3].

Figure 5.12: BCS East-West Interlink

5.4.1 Setup

To maintain consistency and allow for potential correlation with the November C-Lion1 event, we use the exact same time periods for this analysis.

Baseline

We establish a baseline with the data from the 4th of November 2024 to the 14th of November 2024.

Outage

Parallel to the C-Lion1 incident, the outage occurred on the 18th of November 2024 and lasted until the 28th of November 2024.

Check

We use a control period from the 2nd of December 2024 to the 12th of December 2024.

We define the geographical scope of this analysis to capture traffic between Lithuania and Sweden. We filter all measurements to include only records between RIPE Atlas anchors with a source in one of these countries and the destination in the other. Finally, we perform our analysis in both directions.

5.4.2 Results

Anomaly Indicator	Lithuania to Sweden	Sweden to Lithuania
Performance Decrease	✅	✅
High Median		
High RTT Volatility		
Distribution Test Alert	✅	✅
Path Shift	✅	
Dominant Path Shift		
Core Path Shift	✅	
Increased Path Diversity		
Decreased Dom. Path Freq.		
Decreased Success Rate		
Decreased Path Length		

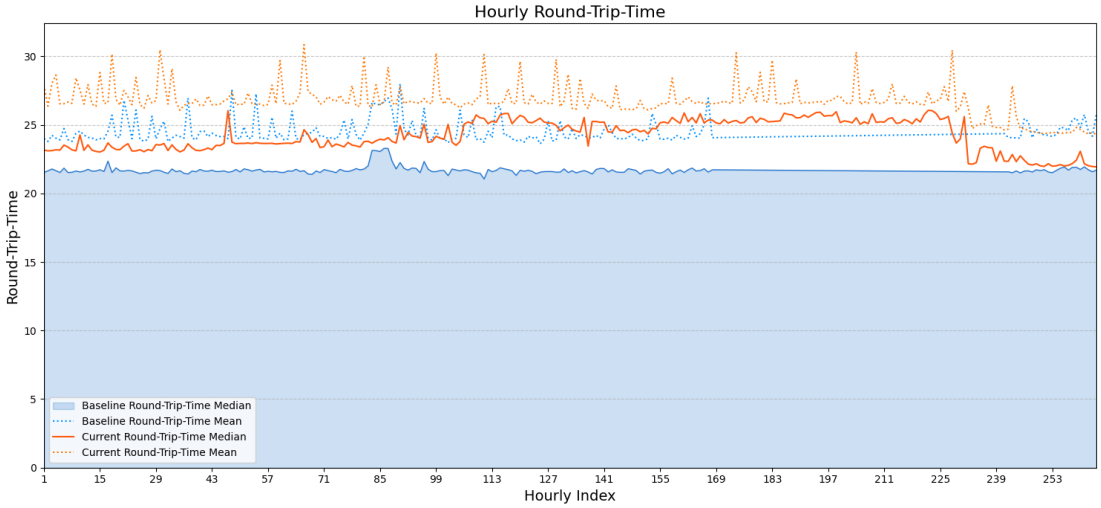
Table 5.7: BCS East-West Interlink: Indicators between Lithuania and Sweden

Our analysis is based on datasets of traffic between 5 anchors in Lithuania and 24 anchors in Sweden. For the baseline period, we collected 125 146 traceroutes from Lithuania to Sweden and 132 743 measurements in the opposite direction. During the outage, we gathered 170 103 data records of outbound traffic and 186 348 traceroutes of packets from Sweden to Lithuania. A detailed table of the complete metrics is shown in Appendix A.3.

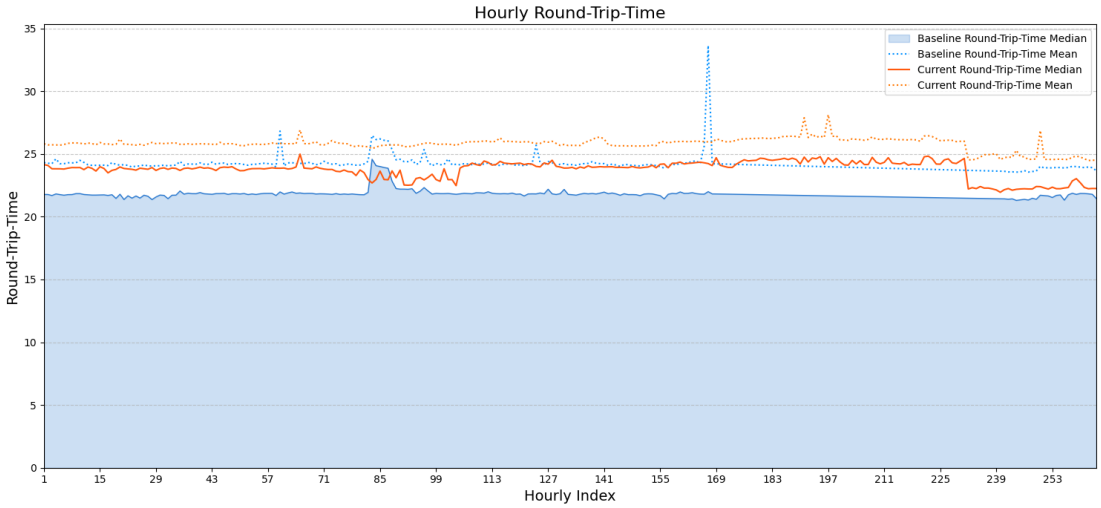
Our identified anomaly indicators, summarized in Table 5.7, demonstrate an asymmetric impact of the incident on the traffic between the two countries. While both directions show a decrease in performance, only the data records from Lithuania to Sweden display a topological change.

In our analysis, we find both our distribution tests reporting a p-value below the threshold, triggering the indicator. In each direction, the Kolmogorov-Smirnov Test, detailed in Section 3.6, reports a p-value of 0.0. Additionally, the Anderson-Darling Test, described in Section 3.7, reports a p-value of 0.001. This signifies a statistically significant

change in the entire distribution of the Round-Trip-Time values, surpassing mere alterations in median or volatility. The median RTT itself sees a slight increase of around 2 ms in either direction. Figure 5.13 provides a comparison of the Round-Trip-Time between the baseline and the time of the outage, aggregated in groups by the hour.



(a) Lithuania to Sweden



(b) Sweden to Lithuania

Figure 5.13: BCS East-West Interlink: Hourly Round-Trip-Times

Data from 2024-11-18 00:00 to 2024-11-28 23:59 UTC (Analyzed Outage Period)

In the traffic from Lithuania to Sweden, we detect a change in topology in the form of a *Core Path Shift*. This rerouting event is likely the origin of the performance degradation, also supported by the increase in RTT standard deviation from 20.41 ms to 25.31 ms. However, the success rate in these measurements improved from 86.93 % to 91.26 % during the incident. This suggests that the alternative routes were slightly slower but more reliable. The combination of the path shift and the decrease in performance confirms a *Performance-Impacting Reroute*. Table 5.8 displays the composite events found in our analysis.

Composite Event	Lithuania to Sweden	Sweden to Lithuania
Major Rerouting		
Perf-Impacting Reroute	✅	
Path Instability		
Perf. Profile Shift		✅
Early Path Failure		

Table 5.8: BCS East-West Interlink: Composite Events between Lithuania and Sweden

In the reverse direction, from Sweden to Lithuania, we observe a different effect of the incident. Because of the absence of a significant change in routing but a present decrease in performance, our analysis reveals a *Performance Profile Shift*. Unlike in the opposite way, the RTT volatility decreased, with the standard deviation dropping from 12.61 ms to 10.49 ms and a stable success rate at over 97 %. Therefore, this shows a change in the performance characteristics of the existing paths rather than a rerouting event.

The data from our control period confirms that the network health improved back to higher levels after the incident, validating the stability of the baseline. Furthermore, this confirms that our observed changes are related to the cable disruption.

5.5 C-Lion1 Incident of Late 2024

In this study, we investigate a second, distinct incident affecting the C-Lion1 submarine cable, reported towards the end of 2024. In this case, the damage did not affect all of the traffic between Finland and Germany, though. The defect was located close to Helsinki, as depicted in Figure 5.15 [3].

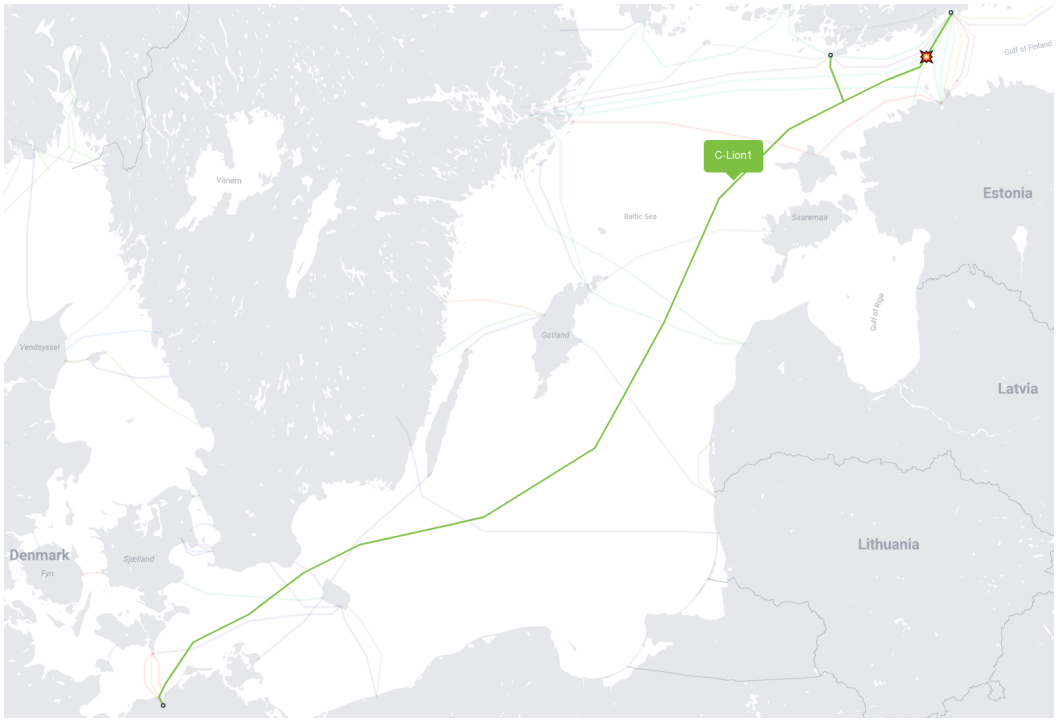


Figure 5.15: C-Lion1 Late 2024 Incident

5.5.1 Setup

We define the three distinct time periods to isolate anomalies for our analysis.

Baseline

We choose the period for our baseline establishment to be from data collected before the damage between the 16$^{\text{th}}$ of December 2024 and the 24$^{\text{th}}$ of December 2024.

Outage

The incident appears to be in the span between the 25$^{\text{th}}$ of December 2024 and the 5$^{\text{th}}$ of January 2025, also covering the Christmas and New Year holidays.

Check

We select our control period after the incident from traceroute measurements

between the 7th of January 2025 to the 15th of January 2025, to serve as a stable reference for comparison.

Consistent with the previous analysis of this cable, we filter for traffic from Finland in a range of 375 km around Helsinki for one end. And probes in a circle of 660 km around Rostock, choosing only anchors from Germany for the other end. Subsequently, we conduct our analysis for traffic in both directions.

5.5.2 Results

Anomaly Indicator	Finland to Germany	Germany to Finland
Performance Decrease	✅	✅
High Median		
High RTT Volatility		
Distribution Test Alert	✅	✅
Path Shift	✅	✅
Dominant Path Shift	✅	✅
Core Path Shift	✅	✅
Increased Path Diversity		
Decreased Dom. Path Freq.		
Decreased Success Rate		
Decreased Path Length		

Table 5.9: C-Lion1 Late 2024: Indicators between Finland and Germany

In our analysis, we utilize traceroutes between 18 anchors in Finland and 177 anchors in Germany. In the baseline dataset, we gathered over two million measurements in each direction. Our data during the time of the incident consists of almost 2.4 million records from Finland to Germany and close to 2.5 million in the reverse direction. The complete metrics are available in Appendix A.4.

Our findings reveal a symmetric reaction of network traffic between the two countries to the outage. We observe a major path shift in combination with a change in the underlying distribution of Round-Trip-Time values in each direction. The triggered anomaly indicators are shown in Table 5.9.

Composite Event	Finland to Germany	Germany to Finland
Major Rerouting	✅	✅
Perf-Impacting Reroute		
Path Instability		
Perf. Profile Shift		
Early Path Failure		

Table 5.10: C-Lion1 Late 2024: Composite Events between Finland and Germany

Table 5.10 shows the composite events we discovered in our analysis. In either direction, we can see a *Major Rerouting Event* occurring. This results from the combination of the mentioned significant changes between the baseline and the outage data.

The considerable change in topology observed in our analysis manifests as both a *Dominant Path Shift* as well as a *Core Path Shift*. At the same time, our distribution tests report p-values that describe a significant shift in the distribution of Round-Trip-Time values, representing a degradation in performance.

While the baseline contains data of Christmas Eve and the outage period includes the Christmas holidays and New Year, we did not observe a measurable deviation in metrics within their respective timeframes.

Despite the major rerouting, the direct impact on connectivity was moderate. The damage to the C-Lion1 submarine cable affected only one of the two connections to Finland, which likely explains the minimal impact. The median RTT increased by only around 2 ms and the success rate remained stable in both directions. We can observe a decrease in path diversity, especially for traffic from Finland to Germany, suggesting that network packets were routed onto fewer, slightly slower paths.

The control period data show that the network performance and topology returned to the baseline state after the incident, confirming that the observed changes were temporary and related to the cable fault. Furthermore, this verifies the reliability of our baseline.

Chapter 6

Evaluation

Our analysis of the case studies in Chapter 5 demonstrates the effectiveness of our approach in detecting and classifying real-world network anomalies. The combination of statistical and path-based metrics, coupled with our methodology to establish a stable baseline, provides a robust and reliable procedure for identifying significant deviations in traceroute behavior. In this chapter, we evaluate the effectiveness of our methodology and discuss the results of our case studies.

6.1 Effectiveness of the Methodology

The core strength of our methodology is the hierarchical approach to detecting anomalies. By building from basic metrics to indicators and finally to composite events, we can distinguish between minor fluctuations and significant, impactful incidents. This structured approach provides a high degree of confidence in our findings and minimizes the risk of false positives.

Our use of a dynamic and comparable baseline is another key advantage. By analyzing data against a stable baseline established from a similar timeframe, such as starting on the same day of the week, we can account for normal variations in network traffic and performance. This is particularly important to prevent false positives that may result from predictable daily and weekly patterns in Internet usage.

By combining statistical and path-based metrics, we gain a comprehensive understanding of network health. The vast number and variety of metrics we analyze is a significant advantage in itself, as they provide detailed information that allows us to capture a broad spectrum of network behaviors that may be invisible to analyses with a more limited

scope. Even without the subsequent step of relating our indicators to composite events, the extensive set of individual metrics alone offers a thorough and insightful view into the state of the network. While statistical metrics are essential to measure impacts on performance, our path-based analysis, including the investigation of dominant and core paths, is crucial to understand the underlying topology and detect changes in routing. Our case studies demonstrate that the most valuable insights are often derived from the combination of these two types of metrics.

6.2 Case Studies

The outage in the Iberian Peninsula highlights the effectiveness and ability of our approach to detect large-scale network events. We successfully identified routing changes and significant performance impacts on different routes between geographical locations. The ability to analyze the event in various topological perspectives and in both directions provided us with a comprehensive understanding of the effect on network health.

Furthermore, the C-Lion1 incident of November 2024 demonstrates the ability of our methodology to detect subtle and asymmetric effects. The impact of the damage was not a complete outage but a measurable performance degradation and a substantial topological shift. We successfully identified the rerouting event within our path-based analysis, even though the decrease in performance was not significant enough to trigger a composite event. This highlights the value of our individual indicators, which can show network changes that might be missed by approaches that only focus on major failures.

The concurrent incident of the BCS East-West Interlink further showcases the effectiveness of our method. We successfully identified an asymmetric impact, detecting an anomaly in the form of a *Performance-Impacting Reroute* in one direction and characterized as a *Performance Profile Shift* in the other. This displays our capability to distinguish between an anomaly caused by a topological change and one that affects the performance of existing paths. Our use of distribution tests was particularly effective during this incident, allowing us to detect the statistically significant shift in the Round-Trip-Time distribution.

Finally, the late 2024 C-Lion1 damage emphasizes the ability of our methodology to detect major routing changes with a low impact on performance. The identification of a *Major Rerouting Event* in both directions, despite only a slight increase in median Round-Trip-Time, highlights the value of the broad coverage of metrics provided by our indicators.

Chapter 7

Conclusion

In this thesis, we designed and implemented a comprehensive methodology for the automated detection and classification of network anomalies using large-scale RIPE Atlas traceroute data. Our approach employs the principle of dynamic baselines from historical data rather than relying on static thresholds. This allows us to account for the inherent variability in traffic patterns across different times and days and establish a reliable and stable profile of normal network behavior.

The core of our methodology is a multistage analysis that combines statistical performance metrics, such as the median Round-Trip-Time and success rate, with topological metrics derived from dominant and core path analyses. We developed a comprehensive set of ten distinct anomaly indicators, which cover various statistical performance metrics and topological path characteristics. These individual indicators combine to form five composite events that identify and classify anomalies with a high degree of confidence. Therefore, our approach is able to distinguish true anomalies from normal network fluctuations.

We demonstrated the effectiveness and ability of our methodology through an analysis of various real-world case studies. Our approach reliably detected network anomalies, such as those caused by large-scale incidents like the Iberian Peninsula outage, by identifying significant rerouting events and their widespread performance impacts. Furthermore, our method has shown its capability to detect more subtle and asymmetric shifts in analyzing the C-Lion1 and BCS East-West Interlink incidents from November 2024. It successfully distinguished between reroutes with performance impacts and more nuanced shifts in performance profiles, highlighting the capabilities of combining topological and statistical analyses.

Our work moves beyond the limitations of conventional network monitoring, which often relies on static thresholds of single metrics, like the mean RTT. Instead, we introduce a method that systematically combines statistical and topological analysis to thoroughly diagnose the network state. Additionally, we employ nonparametric statistical tests, specifically the Kolmogorov-Smirnov and Anderson-Darling tests, to identify shifts in the overall distribution of latency values. This combination of various analyses enables a definitive identification and classification of abnormal network events. Furthermore, our approach offers a deep and detailed understanding of the effects anomalies have on network performance and topology. Therefore, our method presents a highly effective and robust procedure for using regular traceroute measurements to detect and classify network latency anomalies on a large scale.

Chapter 8

Future Work

Building on the foundation we established in this thesis, there are several possibilities for future research and extensions to further enhance the capabilities of anomaly detection.

8.1 Real-Time Anomaly Detection

As our current methodology relies on historical data dumps, a valuable and effective extension would be to adapt the procedure for real-time analysis. This would involve creating a data pipeline that subscribes to the RIPE Atlas streaming API, allowing for the continuous analysis of traceroute data as it is generated. Consequently, this addition in functionality would provide immediate insights into developing network issues and allow for a faster detection of network anomalies. This extension presents considerable challenges in statistical significance and evaluation when a degradation in network health presents a true anomaly.

8.2 Automated Threshold and Sensitivity Adjustment

An area for improvement would be the automation of the sensitivity thresholds for our indicators. Currently, these thresholds are manually defined and based on empirical analysis. A more advanced process could employ machine learning to dynamically adjust these values. By training a model on a large set of traceroute data and known anomalies, it could learn to automatically change the sensitivity based on the specific network environment, time of day, holidays or other factors.

8.3 Path Analysis with per-hop Geolocation

To gain a deeper and more detailed understanding of the topological changes detected by our path-based analysis, future work could integrate IP geolocation data. By mapping the IP address of each hop along a traceroute path to its real-world location, we could visualize the physical route network packets take. This would allow us to more easily understand the physical change in pathing during rerouting events and see how traffic routes around areas affected by an incident.

8.4 Root Cause Analysis

Our current procedure is highly effective at detecting and classifying anomalies, but does not determine their root cause. A significant extension of our process would be to develop strategies that can automatically discover the likely cause of a detected anomaly. This could allow us to quickly identify and address issues at their source, leading to improved network reliability, performance and overall health.

Chapter A

Appendix

A.1 Iberia Results

A.1.1 Iberia Internal

Metric	Baseline	Outage	Check
Measurement Count	1555150	25712	1460261
IQR Upper Fence (ms)	43.14	52.48	41.63
Outliers	41915	539	42070
Median RTT (ms)	11.74	13.39	12.10
RTT standard deviation (ms)	21.16	29.82	22.25
Median first Hop RTT (ms)	0.68	0.53	0.69
Median Duration (s)	16000	25000	22000
Success Rate	78.33%	56.73%	72.08%
Timeout Rate	68.49%	67.99%	71.34%
Median Timeouts	1	2	2
Total Paths	1218128	14586	1052534
Unique Paths	12502	2300	8898
Median Observed Path Length	9	11	10
Median Successful Path Length	8	9	8
Dominant Path Frequency	0.63%	0.90%	0.64%

Table A.1: Metrics of the Iberia Outage: Traffic within Iberia

A.1.2 British Isles

Metric	Baseline	Outage	Check
Measurement Count	2299635	39522	2113502
IQR Upper Fence (ms)	59.55	62.42	61.61
Outliers	63801	1231	57487
Median RTT (ms)	35.42	35.46	35.58
RTT standard deviation (ms)	20.89	33.16	23.14
Median first Hop RTT (ms)	0.70	0.53	0.73
Median Duration (s)	25000.0	36000.0	26000.0
Success Rate	78.80%	71.51%	77.91%
Timeout Rate	78.46%	75.79%	78.95%
Median Timeouts	2	2	2
Total Paths	1812189	28261	1646624
Unique Paths	40839	6393	37870
Median Observed Path Length	12	13	12
Median Successful Path Length	11	11	11
Dominant Path Frequency	0.35%	0.47%	0.35%

Table A.2: Metrics of the Iberia Outage: Iberia to British Isles

Metric	Baseline	Outage	Check
Measurement Count	2307441	61461	2173608
IQR Upper Fence (ms)	55.53	59.48	56.88
Outliers	68986	1398	67491
Median RTT (ms)	34.05	34.67	33.83
RTT standard deviation (ms)	23.15	21.15	20.77
Median first Hop RTT (ms)	0.64	0.57	0.64
Median Duration (s)	25000.0	41000.0	33000.0
Success Rate	76.13%	53.61%	69.80%
Timeout Rate	76.35%	77.60%	79.07%
Median Timeouts	2	3	3
Total Paths	1756567	32947	1517259
Unique Paths	22221	6317	19732
Median Observed Path Length	12	13	12
Median Successful Path Length	11	11	11
Dominant Path Frequency	0.28%	0.41%	0.32%

Table A.3: Metrics of the Iberia Outage: British Isles to Iberia

A.1.3 France

Metric	Baseline	Outage	Check
Measurement Count	2791765	46994	2537772
IQR Upper Fence (ms)	64.52	65.72	65.74
Outliers	73512	1736	56405
Median RTT (ms)	30.84	31.24	31.44
RTT standard deviation (ms)	23.65	37.15	23.82
Median first Hop RTT (ms)	0.73	0.56	0.74
Median Duration (s)	27000.0	30000.0	25000.0
Success Rate	82.36%	76.34%	83.55%
Timeout Rate	82.28%	79.08%	81.61%
Median Timeouts	2	2	2
Total Paths	2299194	35875	2120270
Unique Paths	52908	8648	45176
Median Observed Path Length	11	11	11
Median Successful Path Length	11	11	11
Dominant Path Frequency	0.39%	0.34%	0.22%

Table A.4: Metrics of the Iberia Outage: Iberia to France

Metric	Baseline	Outage	Check
Measurement Count	2904663	78666	2819507
IQR Upper Fence (ms)	60.32	63.60	61.74
Outliers	77877	2128	57436
Median RTT (ms)	29.87	31.13	30.31
RTT standard deviation (ms)	21.33	25.63	20.24
Median first Hop RTT (ms)	0.51	0.43	0.51
Median Duration (s)	29000.0	41000.0	33000.0
Success Rate	80.40%	57.01%	75.37%
Timeout Rate	81.27%	80.05%	82.34%
Median Timeouts	2	3	3
Total Paths	2335445	44846	2124968
Unique Paths	38015	9636	34055
Median Observed Path Length	11	13	12
Median Successful Path Length	11	11	11
Dominant Path Frequency	0.21%	0.30%	0.25%

Table A.5: Metrics of the Iberia Outage: France to Iberia

A.1.4 Central Europe

Metric	Baseline	Outage	Check
Measurement Count	9958664	171516	9096015
IQR Upper Fence (ms)	68.86	68.32	68.70
Outliers	221716	6787	201706
Median RTT (ms)	39.93	39.65	40.10
RTT standard deviation (ms)	25.24	39.77	23.41
Median first Hop RTT (ms)	0.75	0.55	0.76
Median Duration (s)	28000.0	33000.0	26000.0
Success Rate	81.81%	75.23%	82.01%
Timeout Rate	77.87%	74.80%	78.25%
Median Timeouts	2	2	2
Total Paths	8147195	129040	7459676
Unique Paths	207389	32437	187471
Median Observed Path Length	12	13	12
Median Successful Path Length	11	11	11
Dominant Path Frequency	0.08%	0.06%	0.09%

Table A.6: Metrics of the Iberia Outage: Iberia to Central Europe

Metric	Baseline	Outage	Check
Measurement Count	10025454	272977	9770945
IQR Upper Fence (ms)	63.34	66.13	63.72
Outliers	262820	7519	212999
Median RTT (ms)	38.00	38.79	38.28
RTT standard deviation (ms)	24.70	27.06	20.41
Median first Hop RTT (ms)	0.57	0.46	0.57
Median Duration (s)	25000.0	41000.0	32000.0
Success Rate	78.37%	54.93%	72.32%
Timeout Rate	76.60%	77.94%	79.79%
Median Timeouts	2	3	2
Total Paths	7856719	149946	7066308
Unique Paths	140131	29451	120515
Median Observed Path Length	12	13	12
Median Successful Path Length	11	12	11
Dominant Path Frequency	0.06%	0.10%	0.07%

Table A.7: Metrics of the Iberia Outage: Central Europe to Iberia

A.1.5 Italy

Metric	Baseline	Outage	Check
Measurement Count	1310291	23343	1175793
IQR Upper Fence (ms)	83.46	86.48	83.63
Outliers	18128	924	13211
Median RTT (ms)	40.91	41.28	40.67
RTT standard deviation (ms)	24.03	57.19	25.55
Median first Hop RTT (ms)	0.67	0.51	0.69
Median Duration (s)	25000.0	26000.0	25000.0
Success Rate	81.24%	76.86%	82.09%
Timeout Rate	77.66%	74.36%	78.04%
Median Timeouts	2	2	2
Total Paths	1064537	17941	965150
Unique Paths	20745	3906	18206
Median Observed Path Length	11	11	11
Median Successful Path Length	10	10	10
Dominant Path Frequency	0.46%	0.49%	0.49%

Table A.8: Metrics of the Iberia Outage: Iberia to Italy

Metric	Baseline	Outage	Check
Measurement Count	1415949	38399	1382850
IQR Upper Fence (ms)	73.24	79.41	73.21
Outliers	17715	1015	13844
Median RTT (ms)	37.97	39.74	38.19
RTT standard deviation (ms)	17.61	49.17	18.98
Median first Hop RTT (ms)	0.51	0.43	0.50
Median Duration (s)	24000.0	36000.0	25000.0
Success Rate	78.07%	54.15%	71.59%
Timeout Rate	74.69%	74.47%	77.72%
Median Timeouts	2	2	2
Total Paths	1105488	20792	990002
Unique Paths	11910	3917	12283
Median Observed Path Length	11	12	12
Median Successful Path Length	10	11	11
Dominant Path Frequency	0.43%	0.29%	0.46%

Table A.9: Metrics of the Iberia Outage: Italy to Iberia

A.2 C-Lion1 November 2024 Results

Metric	Baseline	Outage	Check
Measurement Count	1973828	2289084	2368043
IQR Upper Fence (ms)	61.28	61.67	60.21
Outliers	36041	39203	43439
Median RTT (ms)	34.40	36.24	34.24
RTT standard deviation (ms)	12.22	11.71	15.43
Median first Hop RTT (ms)	0.60	0.63	0.60
Median Duration (s)	24000.0	25000.0	25000.0
Success Rate	83.89%	77.75%	78.47%
Timeout Rate	67.92%	73.81%	73.25%
Median Timeouts	1	2	2
Total Paths	1655783	1779785	1858102
Unique Paths	18333	23650	22300
Median Observed Path Length	11	12	12
Median Successful Path Length	11	11	11
Dominant Path Frequency	0.09%	0.12%	0.11%

Table A.10: Metrics of the C-Lion1 incident of November 2024: Finland to Germany

Metric	Baseline	Outage	Check
Measurement Count	2008208	2531969	2576513
IQR Upper Fence (ms)	63.71	66.17	63.81
Outliers	47307	49968	54694
Median RTT (ms)	35.50	37.13	35.44
RTT standard deviation (ms)	40.02	20.43	18.79
Median first Hop RTT (ms)	0.47	0.46	0.45
Median Duration (s)	17000.0	23000.0	21000.0
Success Rate	92.98%	79.07%	82.23%
Timeout Rate	66.44%	71.48%	71.76%
Median Timeouts	1	1	1
Total Paths	1867289	2002078	2118695
Unique Paths	21040	29374	24486
Median Observed Path Length	11	12	12
Median Successful Path Length	11	12	11
Dominant Path Frequency	0.09%	0.10%	0.10%

Table A.11: Metrics of the C-Lion1 incident of November 2024: Germany to Finland

A.3 BCS East-West Interlink Results

Metric	Baseline	Outage	Check
Measurement Count	125146	170103	150691
IQR Upper Fence (ms)	48.24	53.61	46.04
Outliers	2001	6032	4526
Median RTT (ms)	21.69	23.70	21.64
RTT standard deviation (ms)	20.41	25.31	11.44
Median first Hop RTT (ms)	0.46	0.47	0.45
Median Duration (s)	27000.0	25000.0	25000.0
Success Rate	86.93%	91.26%	88.87%
Timeout Rate	67.66%	67.65%	67.62%
Median Timeouts	2	2	1
Total Paths	108784	155233	133924
Unique Paths	1566	2587	2151
Median Observed Path Length	10	10	11
Median Successful Path Length	10	10	11
Dominant Path Frequency	1.43%	1.36%	1.45%

Table A.12: Metrics of the BCS East-West Interlink incident 2024: Lithuania to Sweden

Metric	Baseline	Outage	Check
Measurement Count	132743	186348	202105
IQR Upper Fence (ms)	48.63	52.84	46.24
Outliers	216	2279	4311
Median RTT (ms)	21.82	23.88	21.39
RTT standard deviation (ms)	12.61	10.49	11.54
Median first Hop RTT (ms)	0.74	0.71	0.66
Median Duration (s)	34000.0	36000.0	37000.0
Success Rate	97.66%	97.71%	80.67%
Timeout Rate	74.91%	72.81%	75.30%
Median Timeouts	2	2	3
Total Paths	129635	182076	163031
Unique Paths	1515	2024	2095
Median Observed Path Length	11	11	12
Median Successful Path Length	11	11	11
Dominant Path Frequency	1.20%	1.16%	1.29%

Table A.13: Metrics of the BCS East-West Interlink incident 2024: Sweden to Lithuania

A.4 C-Lion1 Late 2024 Results

Metric	Baseline	Outage	Check
Measurement Count	2047355	2392274	2044964
IQR Upper Fence (ms)	61.09	60.99	60.85
Outliers	33825	38737	35155
Median RTT (ms)	34.40	36.44	34.51
RTT standard deviation (ms)	11.08	11.34	10.41
Median first Hop RTT (ms)	0.59	0.59	0.58
Median Duration (s)	25000.0	25000.0	25000.0
Success Rate	77.67%	78.63%	78.75%
Timeout Rate	72.94%	71.83%	71.74%
Median Timeouts	2	2	2
Total Paths	1590139	1881126	1610421
Unique Paths	17664	15909	17590
Median Observed Path Length	12	12	12
Median Successful Path Length	11	11	11
Dominant Path Frequency	0.11%	0.11%	0.11%

Table A.14: Metrics of the C-Lion1 incident of Late 2024: Finland to Germany

Metric	Baseline	Outage	Check
Measurement Count	2104478	2490258	2119478
IQR Upper Fence (ms)	63.67	65.70	64.07
Outliers	44374	40578	43766
Median RTT (ms)	35.44	37.90	35.58
RTT standard deviation (ms)	15.65	17.21	17.56
Median first Hop RTT (ms)	0.46	0.46	0.46
Median Duration (s)	21000.0	19000.0	18000.0
Success Rate	85.40%	84.86%	85.45%
Timeout Rate	70.98%	69.87%	69.82%
Median Timeouts	1	1	1
Total Paths	1797130	2113241	1810991
Unique Paths	19980	21363	19389
Median Observed Path Length	12	12	12
Median Successful Path Length	11	11	11
Dominant Path Frequency	0.10%	0.10%	0.10%

Table A.15: Metrics of the C-Lion1 incident of Late 2024: Germany to Finland

A.5 List of Acronyms

A-D	Anderson-Darling
API	Application Programming Interface
AS	Autonomous System
BGP	Border Gateway Protocol
CPU	Central Processing Unit
DBMS	Database Management System
DNS	Domain Name Server
ECDF	empirical cumulative distribution function
ICMP	Internet Control Message Protocol
IP	Internet Protocol
IPv4	Internet Protocol Version 4
IPv6	Internet Protocol Version 6
IQR	Interquartile Range
ISP	Internet Service Provider
JSON	JavaScript Object Notation
K-S	Kolmogorov-Smirnov
OLAP	Online Analytical Processing
OLTP	Online Transaction Processing
ping	Packet Internet Groper
RIPE NCC	Réseaux IP Européens Network Coordination Centre
RTT	Round-Trip-Time
SIMD	Single Instruction, Multiple Data
SQL	Structured Query Language
std dev	standard deviation
TCP	Transmission Control Protocol
TTL	Time-to-Live
UDP	User Datagram Protocol
UTC	Coordinated Universal Time

Bibliography

[1] RIPE NCC. "RIPE Atlas". https://atlas.ripe.net/. (2010).

[2] Google, *Google Gemini*, 2025.

[3] TeleGeography. "Submarine Cable Map". https://www.submarinecablemap.com/. (2025).

[4] E. Aben, J. Cowie, and A. Davies, *A Deep Dive Into the Baltic Sea Cable Cuts*, Dec. 2024. [Online]. Available: `https://labs.ripe.net/author/emileaben/a-deep-dive-into-the-baltic-sea-cable-cuts/`.

[5] RIPE NCC. "RIPE NCC". https://www.ripe.net/. (2010).

[6] E. Aben and A. Davies, *Does the Internet Route Around Damage? - Baltic Sea Cable Cuts*, Nov. 2024. [Online]. Available: `https://labs.ripe.net/author/emileaben/does-the-internet-route-around-damage-baltic-sea-cable-cuts/`.

[7] C. Yang and W. Jia, "Bgp anomaly detection - a path-based apporach", in *2023 3rd Asia-Pacific Conference on Communications Technology and Computer Science (ACCTCS)*, 2023, pp. 408–414. DOI: `10.1109/ACCTCS58815.2023.00100`.

[8] Y. Liu, X. Luo, R. K. C. Chang, and J. Su, "Characterizing inter-domain rerouting by betweenness centrality after disruptive events", *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 6, pp. 1147–1157, 2013. DOI: `10.1109/JSAC.2013.130616`.

[9] B. Al-Musawi, P. Branch, and G. Armitage, "Bgp anomaly detection techniques: A survey", *IEEE Communications Surveys & Tutorials*, vol. PP, pp. 1–1, Oct. 2016. DOI: `10.1109/COMST.2016.2622240`.

[10] B. Hou, C. Hou, T. Zhou, Z. Cai, and F. Liu, "Detection and characterization of network anomalies in large-scale rtt time series", *IEEE Transactions on Network and Service Management*, vol. 18, no. 1, pp. 793–806, 2021. DOI: `10.1109/TNSM.2021.3050495`.

[11] R. Fontugne, C. Pelsser, E. Aben, and R. Bush, "Pinpointing delay and forwarding anomalies using large-scale traceroute measurements", in *Proceedings of the*

2017 Internet Measurement Conference, ser. IMC '17, London, United Kingdom: Association for Computing Machinery, 2017, pp. 15–28, ISBN: 9781450351188. DOI: 10.1145/3131365.3131384. [Online]. Available: https://doi.org/10.1145/3131365.3131384.

[12] J. Hawkinson and T. Bates, *Rfc1930: Guidelines for creation, selection, and registration of an autonomous system (as)*, USA, 1996.

[13] Y. Rekhter, T. Li, and S. Hares, *Rfc 4271: A border gateway protocol 4 (bgp-4)*, USA, 2006.

[14] J. L. Sobrinho, "Network routing with path vector protocols: Theory and applications", in *Proceedings of the 2003 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, ser. SIGCOMM '03, Karlsruhe, Germany: Association for Computing Machinery, 2003, pp. 49–60, ISBN: 1581137354. DOI: 10.1145/863955.863963. [Online]. Available: https://doi.org/10.1145/863955.863963.

[15] Orixcom, *What is bgp protocol?* [Online]. Available: https://www.orixcom.com/resources/bgp-protocol (visited on 09/07/2025).

[16] *Traceroute man page*, https://linux.die.net/man/8/traceroute6. (visited on 08/25/2025).

[17] F. Viger, B. Augustin, X. Cuvellier, *et al.*, "Detection, understanding, and prevention of traceroute measurement artifacts", *Computer Networks*, vol. 52, no. 5, pp. 998–1018, 2008, ISSN: 1389-1286. DOI: https://doi.org/10.1016/j.comnet.2007.11.017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1389128607003428.

[18] ClickHouse Inc., *Clickhouse*. [Online]. Available: https://clickhouse.com/ (visited on 08/25/2025).

[19] ClickHouse Inc., *Clickhouse docs*. [Online]. Available: https://clickhouse.com/docs (visited on 08/25/2025).

[20] N. S. T. Thallam, "Columnar storage vs. row-based storage: Performance considerations for data warehousing", *Journal of Scientific and Engineering Research*, 2022, ISSN: 2394-2630. [Online]. Available: https://jsaer.com/download/vol-9-iss-4-2022/JSAER2022-9-4-238-249.pdf.

[21] KnowTechie. "Clickhouse explained: Fast queries and real-time analytics". (2024), [Online]. Available: https://knowtechie.com/clickhouse-high-performance-olap-database/.

[22] S. K. S. Durai and M. D. Shamili, "Smart farming using machine learning and deep learning techniques", *Decision Analytics Journal*, vol. 3, p. 100 041, 2022, ISSN: 2772-6622. DOI: https://doi.org/10.1016/j.dajour.2022.100041. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S277266222200011X.

[23] R. Pu, J. Xu, L. Yang, *et al.*, "Coulomb's law-inspired parameter-free outlier detection algorithm", *Applied Soft Computing*, vol. 167, p. 112 348, 2024, ISSN: 1568-4946. DOI: https://doi.org/10.1016/j.asoc.2024.112348. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1568494624011220.

[24] C. S. K. Dash, A. K. Behera, S. Dehuri, and A. Ghosh, "An outliers detection and elimination framework in classification task of data mining", *Decision Analytics Journal*, vol. 6, p. 100 164, 2023, ISSN: 2772-6622. DOI: https://doi.org/10.1016/j.dajour.2023.100164. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2772662223000048.

[25] A. Smiti, "A critical overview of outlier detection methods", *Computer Science Review*, vol. 38, p. 100 306, 2020, ISSN: 1574-0137. DOI: https://doi.org/10.1016/j.cosrev.2020.100306. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1574013720304068.

[26] A. N. Kolmogorov, "Sulla determinazione empírica di una legge di distribuzione", vol. 4, 1933, pp. 83–91.

[27] N. Smirnov, "Table for estimating the goodness of fit of empirical distributions", *The Annals of Mathematical Statistics*, vol. 19, no. 2, pp. 279–281, 1948, ISSN: 00034851. [Online]. Available: http://www.jstor.org/stable/2236278 (visited on 08/22/2025).

[28] V. Berger and Y. Zhou, "Kolmogorov–smirnov test: Overview", in Sep. 2014, ISBN: 9781118445112. DOI: 10.1002/9781118445112.stat06558.

[29] D. A. Darling, "The kolmogorov-smirnov, cramér-von mises tests", *The Annals of Mathematical Statistics*, vol. 28, no. 4, pp. 823–838, 1957, ISSN: 00034851, 21688990. [Online]. Available: http://www.jstor.org/stable/2237048 (visited on 08/22/2025).

[30] V. Paxson, "End-to-end routing behavior in the internet", *IEEE/ACM Trans. Netw.*, vol. 5, no. 5, pp. 601–615, Oct. 1997, ISSN: 1063-6692.

[31] M. A. Stephens, "Edf statistics for goodness of fit and some comparisons", *Journal of the American Statistical Association*, vol. 69, no. 347, pp. 730–737, 1974, ISSN: 01621459, 1537274X. [Online]. Available: http://www.jstor.org/stable/2286009 (visited on 08/22/2025).

[32] T. W. Anderson and D. A. Darling, "Asymptotic Theory of Certain "Goodness of Fit" Criteria Based on Stochastic Processes", *The Annals of Mathematical Statistics*, vol. 23, no. 2, pp. 193–212, 1952. DOI: 10.1214/aoms/1177729437. [Online]. Available: https://doi.org/10.1214/aoms/1177729437.

[33] F.-W. Scholz and M. A. Stephens, "K-sample anderson–darling tests", *Journal of the American Statistical Association*, vol. 82, pp. 918–924, 1987. [Online]. Available: https://api.semanticscholar.org/CorpusID:38906795.

[34] L. Dall'Asta, I. Alvarez-Hamelin, A. Barrat, A. Vázquez, and A. Vespignani, "Exploring networks with traceroute-like probes: Theory and simulations", *Theoretical Computer Science*, vol. 355, no. 1, pp. 6–24, 2006, Complex Networks, ISSN: 0304-3975. DOI: https://doi.org/10.1016/j.tcs.2005.12.009. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0304397505009126.

[35] R. Moskowitz, D. Karrenberg, Y. Rekhter, E. Lear, and G. J. de Groot, *Address Allocation for Private Internets*, USA, 1996.

[36] ENTSO-E, *Entso-e expert panel initiates the investigation into the causes of iberian blackout*, May 2025. [Online]. Available: https://www.entsoe.eu/news/2025/05/09/entso-e-expert-panel-initiates-the-investigation-into-the-causes-of-iberian-blackout/.