*Journal of Internet Services and Applications, 2025, 16:1,* doi: 10.5753/jisa.2025.5218


# Time-Weighted Correlation Approach to Identify High Delay Links in Internet Service Providers

**Francisco V. J. Nobre** [ **State University of Ceará** | ***valderlan.nobre@aluno.uece.br*** ]
**Danielle dos S. Silva** [ **State University of Ceará** | ***danielle.santos@aluno.uece.br*** ]
**Maria C. M. M. Ferreira** [ **State University of Ceará** | ***clara.mesquita@aluno.uece.br*** ]
**Maria L. M. L. Brito** [ **State University of Ceará** | ***malu.linhares@aluno.uece.br*** ]
**Thelmo P. de Araújo** [ **State University of Ceará** | ***thelmo.araujo@uece.br*** ]
**Rafael L. Gomes** [ **State University of Ceará** | ***rafa.lopes@uece.br*** ]

*Center of Science and Technology, State University of Ceará, Fortaleza, CE, 60714-903, Brazil.*

**Received:** 27 November 2024 • **Accepted:** 12 May 2025 • **Published:** 10 July 2025

**Abstract** Companies and Internet Service Providers (ISPs) apply monitoring tools over network infrastructure, encompassing regular performance evaluations, with a primary focus on delivering crucial information about the current state of the network infrastructure and, consequently, the services running on it. However, these monitoring tools require ongoing development to handle more complex tasks, such as detecting performance issues. Within this context, this article proposes a mechanism for identifying high delays and communication links in the network that may cause these performance issues, using a temporally formulated Impact Score. This Score is based on data correlation techniques applied to information collected by monitoring tools. Experiments conducted with real data from the RNP Network indicate the efficiency of the proposal in identifying links impacting data communication, resulting in high end-to-end delays.

**Keywords:** Correlation, Data Analysis, Network Performance.

## 1 Introduction

In today's world, organizations like Internet Service Providers (ISPs), businesses, and institutions need to implement network monitoring services to collect data on the performance and behavior of their network infrastructure [Costa *et al.*, 2024]. This process involves examining various metrics, including latency, packet loss, throughput, and other key indicators [Portela *et al.*, 2024a]. Delay, representing the end-to-end time required to transmit bits across a network, serves as a crucial metric for informed decision-making in network and service management. [Mok *et al.*, 2021]. These management actions encompass network infrastructure expansion, establishing Quality of Service (QoS) expectations, evaluating Service Level Agreements (SLA), and analyzing demand for network resources [Souza *et al.*, 2024].

In terms of QoS, end-to-end delay plays a crucial role in evaluating the efficiency of network services that are in operation [Stafecka *et al.*, 2024]. High delay cases can result in a reduced bit transmission rate, network congestion, and possible failures in service provision, negatively impacting the user experience and decreasing the productivity of both the network and its clients [Portela *et al.*, 2024b]. On the other hand, reducing delays helps enhance the user experience, leading to increased productivity and better overall performance of the network infrastructure [Alam *et al.*, 2024]. Hence, monitoring and optimizing network performance is essential to ensure that SLAs' QoS metrics are met [Ferreira *et al.*, 2024].

Monitoring end-to-end delays between network nodes is essential for maintaining network stability and managing fluctuations in Internet access demand, which are largely shaped by elastic behavior affecting delay performance. This elasticity arises as users access the Internet differently based on location, leading to variable demand for network resources throughout the day. Poor management of these dynamic scenarios can cause service interruptions, slowdowns, and disconnections, risking breaches of SLA agreements [Costa *et al.*, 2024]. To maintain QoS metrics, it is critical to use adaptive strategies that align with this elastic behavior, ensuring smooth system operation. However, despite the importance of network measurement, current monitoring tools need significant improvement, particularly in identifying instances of high delay in network communication links [Ferreira *et al.*, 2024].

Early identification of delays between communication nodes offers multiple benefits to ISPs, as shown in recent studies [Scarpitta *et al.*, 2023; BinSahaq *et al.*, 2022]: (I) Capacity Planning, early detection helps organizations forecast client's demands and reserve resources effectively, preventing network congestion; (II) QoS Monitoring, continuous detection allows ISPs to monitor network behavior over time, ensuring SLA compliance; and, (III) Resource Allocation, efficient detection enables organizations to allocate resources strategically to essential links, reducing costs associated with underutilized links, such as energy and maintenance expenses.

Within this context, this article presents a mechanism for identifying the causes of high-delay situations in network infrastructure. The proposed method introduces an Impact Score, which is applied to data from network monitoring tools. This Impact Score is based on two principles: (i) Data correlation techniques to quantify and analyze relationships between multiple variables in network measurements, uncovering patterns, dependencies, and insights in the data; and, (ii) Period evaluation to expand the analysis of the situation throughout time. The solution used delay data from Ping measurements and the set of links used by the measurements through the traceroute (IPs/Equipment that form the transmission path). This data enables mapping of the paths used in various measurements and facilitates the correlation of infrastructure links with cases of high and low delay, allowing for the identification of potential causes.

To validate the proposal and evaluate its effectiveness in a real scenario, experiments were performed using real data from the *Ipê* Network Monitoring Service (Monipê)[1] of the National Education and Research Network (RNP). The results demonstrate that the proposed method successfully identifies network links that experience significant end-to-end delays.

In our previous research [Silva *et al*., 2024], we introduced a method for detecting high delays in communication links across the network. However, this initial work did not focus on a complete analysis of the network or the impact of the period of the measurements. In this way, the current proposal has the following innovation issues in front of the previous work: (A) Inclusion of a weighted approach based on periods for correlation definition, while the first paper uses a singular time; (B) Expansion of correlation profile considering both positive (low delay) and negative (high delay) situations, differently than previous paper which considers only negative situation; (C) New recent related work was included to improve the analysis of the proposal in front of the existing solutions; and, (D) Execution of broader set of experiments, considering more data and communication points for evaluation.

Regarding the contribution and industrial impact, the proposal offers practical value for service providers by enabling data-driven identification of links likely responsible for network performance degradation. In operational environments, high Impact Scores, especially those recurring over time, can guide network teams in prioritizing diagnostics, reconfiguration, or capacity adjustments. Thus, providers can define context-specific criteria and incorporate the scores into existing monitoring dashboards, where the proposed Impact Score enhances situational awareness and supports more proactive, efficient network management.

The remainder of this paper is organized as follows: Section 2 presents existing solutions related to network management and performance analysis, while Section 3 describes the proposed solution. Section 4 discusses the results of the experiments conducted, and Section 6 concludes the article.

## 2 Related Work

This section describes the main works related to network management and performance analysis published recently, considering performance and quality issues. Table 1 presents several existing proposals (column *Ref.*), where the *Description* column represents the strategy of the corresponding proposal, while column *Contribution* highlights positive points of it.

Scarpitta *et al*. [2023] propose a solution for Software-Defined Wide Area Networks (SD-WANs) that utilizes the Simple Two-way Active Measurement Protocol (STAMP) to monitor the delay of a Segment Routing over IPv6 (SRv6) path between two nodes. This monitoring tool provides valuable data for network management solutions that leverage measurement data to support network administrators. Similarly, Guo et al. [Guo *et al*., 2025] present a post-routing statistical delay prediction framework. To address the limitation of traditional GNNs, where restricted receptive fields result in incomplete cell embeddings and reduced prediction accuracy, the authors enhance the embedding update process to incorporate both local and global information. The improved cell embeddings, combined with PVT (Process-Voltage-Temperature) data, are used to predict the mean delay and squared deviation of signal paths under current operating conditions, enabling more accurate post-routing delay estimation.

Sawabe *et al*. [2024] explore the relationship between packet rate and network delay components by categorizing end-to-end delays into four types: transmission, propagation, queueing, and processing delays. It formulates how these delay components vary with the packet transmission interval, revealing that their relative impact changes depending on the interval duration. Notably, when packets are sent at short intervals, deterministic jitter from processing delay becomes more dominant than stochastic queueing delay, highlighting the significance of processing behavior in high-rate communication scenarios.

Hekmati *et al*. [2024] introduce a correlation-based solution that considers the correlation of IoT traffic through the network infrastructure, evaluating distinct AI models on data coming from real-world IoT networks to detect DDoS attacks. Similarly, Gajewski *et al*. [2022] present an anomaly detection strategy that enables shared responsibility between a service client and the network provider in the context of a Home Area Network (HAN). In both proposals, the authors use a machine learning approach to classify monitoring and correlation data to detect suspicious behaviors in network resources. In this context, correlation techniques are applied to network flow characteristics to group similar anomalies/attacks. This differs from the approach in this article, where correlation occurs between distinct monitoring data sources with an analysis over a time interval.

Kim *et al*. [2024] present an anomaly detection solution based on long short-term memory (LTSM) and gated graph neural network (GGNN) that correlate each protocol of the network flows with defined tasks to create a singular pattern coming from network flows. Nevertheless, it directly depends on the pre-defined set of tasks to allow the correlation.

Arachchige *et al*. [2023] conducted a study using data correlation techniques to examine temperatures reached by devices processing blockchain algorithms and the energy consumption of three commonly used blockchain algorithms executed on low-power microcontrollers within wireless sensor networks. This study focused solely on the temperature data of the devices performing blockchain operations and their en-

[1] monipe-central.rnp.br

**Table 1.** Related Work

| Ref. | Description | Contribution |
|---|---|---|
| [Scarpitta *et al.*, 2023] | Simple Two-way Active Measurement Protocol (STAMP) to monitor the delay of a Segment Routing over IPv6 (SRv6) path between two nodes | It provides valuable data for network management solutions that leverage measurement data to support network administrators. |
| [Guo *et al.*, 2025] | Post-routing statistical delay prediction framework. | Prediction of path mean delay and squared deviation based on enriched cell embeddings and PVT data. |
| [Hekmati *et al.*, 2024] | It analyzes how different packet transmission intervals influence the composition of end-to-end communication delays. | The study reveals that at shorter transmission intervals, deterministic processing delay jitter becomes more influential than stochastic queueing delay. |
| [Hekmati *et al.*, 2024] | IoT traffic correlation-based approach for detecting DDoS attacks using AI models. | Focuses on grouping anomalies based on network flow characteristics. |
| [Gajewski *et al.*, 2022] | Anomaly detection strategy for HAN, with shared responsibility between client and provider. | Uses machine learning to classify monitoring data, enhancing anomaly detection. |
| [Kim *et al.*, 2024] | Solution using LSTM and graph neural networks to correlate network protocols with specific tasks. | Relies on predefined tasks to create unique patterns for network flows. |
| [Arachchige *et al.*, 2023] | Correlation study between IoT device temperatures and energy consumption in blockchain processing. | Highlights the use of correlation techniques for decision-making in sensor networks. |
| [Imran *et al.*, 2023] | Anomaly detection in IoT with MQTT protocol, with correlation analysis to improve feature selection. | Reduces machine learning training and testing times by selecting relevant features. |
| [Wang *et al.*, 2024] | Optimization of RNN using correlation analysis to determine reservoir size. | Applies probability theory to remove low-correlation neurons, improving training efficiency. |
| Our Proposal | Correlation-based method with time-weighted Impact Score to identify delay-causing links in ISPs. | Uses temporal analysis of delay/traceroute data to pinpoint problematic links, improving network diagnostics. |

ergy consumption. Although it applies data correlation techniques, the study context differs from that of this article, but it highlights the importance of data correlation in identifying events and supporting decision-making.

Imran *et al*. [2023] investigate detection times for attacks in IoT networks based on the Message Queuing Telemetry Transport (MQTT) protocol. The authors employ machine learning approaches and propose a method that applies correlation analysis to reduce the training and testing times of these algorithms. They conclude that correlation analysis is highly beneficial for feature engineering, particularly for identifying the most relevant features within the MQTT dataset. Similarly, Wang *et al*. [2024] apply a variation of a Recurrent Neural Network (RNN) based on data correlation to determine the reservoir size that corresponds to a given training task. Their approach uses probability theory and information theory to measure correlation among RNN neurons, dynamically removing neurons with low correlation to create a more concise feature matrix. This proposal focuses on optimizing training and the use of neural network variations without a specific application context.

Based on the literature review conducted, it is noted that no article in the literature has focused on developing a mechanism for identifying the causes of network performance issues using network monitoring data, which is the focus of this article. Our proposal performs statistical actions to improve the detection process, creating a Score that facilitates the network performance management process.

# 3 Proposal

The proposed mechanism for identifying causes of high delays in network infrastructure relies on two key inputs: end-to-end delay between measurement points and a set of links used in end-to-end communication (i.e., the traceroute output that tracks the route that data packets follow when traveling from one computer/device to another). Generally, this proposed mechanism operates by performing the following key steps: Data Collection, Separation of Traceroute and Delay data, Identification of High and Low Delay cases, Generation of the Correlation Matrix, Calculation of the Impact Score (both for a single time and per period) and finally makes a comparison between the data to discover the links with the highest probability of performance problems. An overview of the solution is presented in Figure 1 and the notation of the formulations used throughout the paper is summarized in Table 2.

Initially, the method inputs network infrastructure data, which is collected at one-minute intervals. This dataset includes comprehensive end-to-end delay measurements (denoted as $M$) alongside traceroute data that reveals the specific communication paths between network nodes, i.e., the constituent links forming set $N$. These data points, obtained through specialized network monitoring and measurement

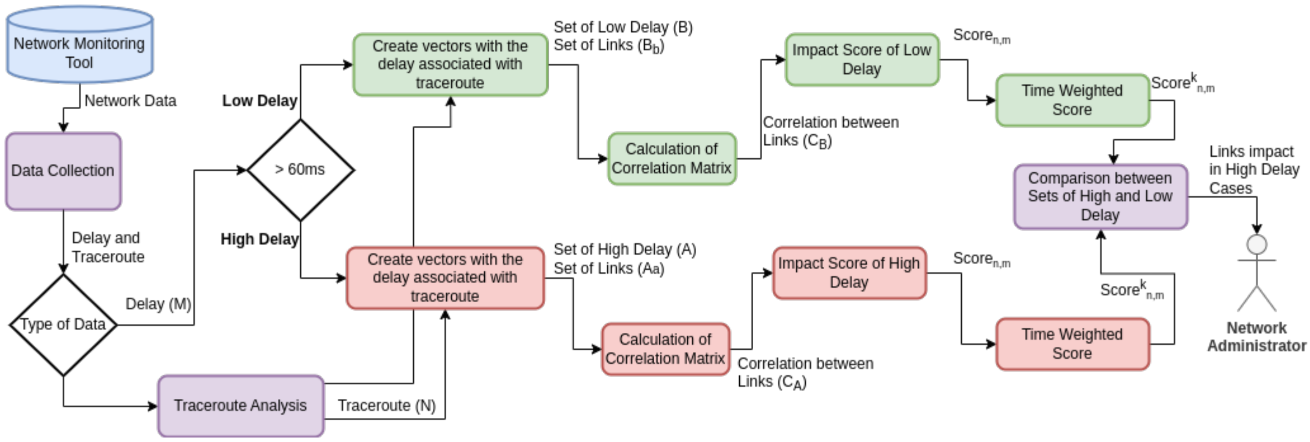



**Figure 1.** Solution overview.

**Table 2.** Notation

| Symbol | Description |
|---|---|
| $M$ | Delay measurements data. |
| $N$ | Communication links set used. |
| $A$ | High Delay Cases. |
| $A_a$ | Communication links in High Delay case $a$. |
| $B$ | Low Delay Cases . |
| $B_b$ | Communication links in Low Delay case $b$. |
| $C_A$ | High Delay Case Correlation Matrix. |
| $C_B$ | Low Delay Case Correlation Matrix. |
| $C_{i,j}$ | Pearson Correlation Coefficient between $i$ and $j$. |
| $Score^k_{n,m}$ | Impact Score of link $n$ as cause of case $m$ on time $k$. |

tools, are crucial for diagnosing and addressing performance degradation. Once collected, the method identifies instances of high delay, defined as end-to-end delays exceeding 60 milliseconds. These cases, considered above the acceptable threshold for ISP environments, are grouped into set $A$ [Scarpitta *et al.*, 2023; BinSahaq *et al.*, 2022]. Cases with delays below this threshold are classified as low delay and grouped into set $B$. Based on the processed traceroute data, after applying link-level filtering, the method identifies and selects the specific links that appear in the communication paths of both high and low delay cases. This step establishes a structured mapping between each delay case (from sets $A$ and $B$) and the set of links traversed in its corresponding network path.

Once the delay cases and their corresponding links are organized, we proceed to calculate the correlation matrices for these two types of cases: one correlation matrix, $C_A$, is defined for High Delay Cases, and another, $C_B$, for Low Delay Cases. Note that the dimensions of each matrix will correspond to the number of links involved in their respective cases, i.e., the union of links ($\bigcup_a A_a$ and $\bigcup_b B_b$). From these correlation matrices, we calculate the Impact Index of a link, denoted as $Score^k_{n,m}$, which represents the *Impact Score* of link $n$ as a cause of case $m$ at time $k$. Initially, only data from the generated matrix is used; subsequently, a time-weighted score is calculated based on the period. These steps are detailed in the following subsections.

Finally, it is important to highlight that this link impact analysis, across both high and low delay cases, enables the solution to determine whether a specific link consistently exhibits a strong influence in high-delay scenarios while having minimal impact in low-delay cases. Such a pattern provides strong evidence that the link is a likely contributor to performance degradation at a given point in time.

## 3.1 Data correlation

Data correlation techniques focus on detecting and measuring relationships, dependencies, or associations between various variables or datasets [Li *et al.*, 2021; Silveira *et al.*, 2023]. The correlation value increases as the relationship between variables strengthens.

One of the most common techniques is the Pearson Correlation Coefficient, typically represented as $r$, which is calculated using Equation 1. In this equation, $x_i$ and $y_i$ represent the data points of the two variables being correlated, while $\overline{x}$ and $\overline{y}$ represent the mean (average) values of each variable. Afterward, a correlation matrix $C$ is generated, where$C_{i,j} = r(x, y)$. It measures the strength and direction of the linear relationship between two continuous variables.

$$C_{i,j} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}} \quad (1)$$

Regarding the properties of Pearson correlation, the coefficient falls between -1 and 1, where: (i) $C_{i,j} = 1$ denotes a perfect positive linear correlation; (ii) $C_{i,j} = -1$ denotes a perfect negative linear correlation; and (iii) $C_{i,j} = 0$ signifies no linear correlation. Thus, $C_{i,j} > 0$ signifies a positive correlation, meaning that as one variable increases, the other tends to increase as well. Conversely, $C_{i,j} < 0$ indicates a negative correlation, meaning that as one variable increases, the other tends to decrease. Meanwhile, $C_{i,j} = 0$ signifies no linear correlation, indicating no linear relationship between the variables. Additional properties include symmetry, meaning $C_{i,j} = C_{j,i}$, and linearity, which evaluates the linear relationship between variables. However, Pearson correlation is sensitive to outliers, which can significantly impact the coefficient [Kim *et al.*, 2015]. In this study, a correlation matrix is generated for the communication links within the network infrastructure involved in performance issues. Each element of the matrix is subsequently used in calculating the Impact Score, as described below.

### 3.2 Impact Score

Once the correlation matrix has been calculated from the data, it's important to extract insights regarding the network infrastructure's current state. To identify which communication links contribute to end-to-end delays, we introduced the Impact Score. This metric quantifies the degree to which a communication link contributes to scenarios of both high and low delay within the network infrastructure, thereby providing a clear indicator of its overall impact on network performance. The calculation of the Score is performed in two phases: initially, the Score for a specific time is calculated and, subsequently, the Score is weighted according to time, where the closer the Score is to the analysis period, the greater the weight. The Impact Score at a given moment is defined by the Equation 2, where $[C_{i,j} > 0]$ is a function returning 1 if $C_{i,j}$ is positive (greater than 0) and 0 otherwise. This allows the Impact Score to aggregate correlation values for a specific link involved in high-delay paths, reflecting its level of responsibility for the current network state.

$$Score_{n,m} = \sum_{i=1}^{m} \sum_{j=1}^{n} (C_{ij} \cdot [C_{ij} > 0]) \tag{2}$$

The strength of the relationship between communication links is reflected in the correlation values, with higher values indicating stronger associations. Accordingly, the proposed Impact Score is designed to identify the set of links that exhibit the highest correlation with other links during instances of high or low delay, enabling a targeted analysis of their influence on network performance. Thus, a higher Impact Score indicates a greater likelihood that the link contributes to the performance issue.

To perform a temporal analysis of the Impact Score behavior over a continuous measurement, a weighting system was implemented for each measurement time. This solution was applied to make it more intuitive to see the influence that a measurement has on the route's final delay. In this system, the weight is applied in such a way that, as a result, the last recorded measurement will have the greatest influence on the result (weight $w = 1$). The oldest records, nevertheless, will follow the formula that represents the Score of a specified link made at time $t_i$, as described in Equation 3, considering $t_0$ as the beginning of the delay measurement and $t_n$ as the end. Consequently, the earliest measurement exerts the least influence on the final route delay, while the latest measurement exerts the greatest influence on the result.

$$Score_{n,m}^{k} = \sum_{i=1}^{k} \frac{1}{2(t_k - t_i) + 1} * Score_{n,m} \tag{3}$$

Equation 3 aims to deal with the dynamic nature of network conditions and the need for a more robust analysis that accounts for temporal variations in performance metrics. In real-world scenarios, network delays and performance issues are rarely static; they evolve due to varying traffic patterns, resource utilization, and external factors such as hardware failures or transient congestion. A single measurement, while informative, may capture only a momentary state of the network, potentially leading to inaccurate conclusions. For example, a link might exhibit high delay in one instance due to a temporary surge in traffic but perform normally in subsequent measurements. Relying solely on this single measurement could result in false positives or negatives in identifying problematic links.

The weighting approach in Equation 3 addresses this challenge by emphasizing more recent measurements while still considering historical data. By assigning a decaying weight to older measurements based on the temporal distance $(t_k - t_i)$, the approach ensures that the most current data has the greatest influence on the calculated impact score. This design reflects the principle that recent performance metrics are more indicative of the current network state, while older data provides context without disproportionately affecting the analysis.

This temporal weighting also allows the model to smooth out fluctuations and identify persistent trends in link performance. Links that consistently contribute to delays over time will naturally accumulate higher scores, providing stronger evidence of their impact on network issues. Conversely, links with transient performance problems will exhibit lower cumulative scores, reducing the likelihood of misclassification.

In summary, the approach applied in Equation 3 ensures a balanced and context-aware analysis of network performance, aligning with the need for proactive and informed decision-making in dynamic network environments. By incorporating temporal factors into the scoring mechanism, this method enhances the reliability and accuracy of identifying links that significantly impact network performance. Finally, it's essential to note that correlation does not imply causation; in other words, the presence of a correlation between two variables does not mean one causes the other. Therefore, the Impact Score is valuable in pinpointing performance issues and enhancing the management of Internet service providers.

## 4 Experiments

This section will describe how the experiments performed were configured (Subsection 4.1) and discuss the results obtained within this set of experiments performed, using real network measurement data (Subsection 4.2). It is worth mentioning that the code developed, as well as the data used in the experiments, are available in the project repository[2] and with the necessary instructions for reproducibility.

### 4.1 Experiments Configurations

As mentioned before, to conduct realistic experiments, we utilized data from RNP's (MonIPÊ)[3] monitoring service, which follows the international perfSONAR monitoring standard[4]. The choice of MonIPÊ was strategic due to its

[2]https://github.com/LarcesUece/Correlation-Analysis-of-Network-Monitoring-Data/
[3]www.rnp.br/en/ipe-network
[4]www.perfsonar.net

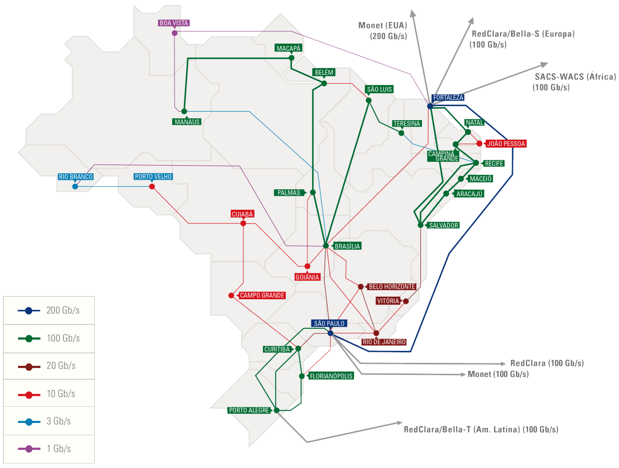



**Figure 2.** *Ipê* Network.

comprehensive coverage and reliable data collection, while perfSONAR provides standardized, high-precision measurements that ensure data consistency, crucial for our correlation analysis. In this setup, throughput measurements are collected every four hours, delay measurements every minute, and traceroute measurements every ten minutes throughout the day. RNP's network spans all 26 Brazilian states and the Federal District, with each point exhibiting unique behavior influenced by local conditions. This geographical diversity strengthens the external validity of our approach. Figure 2 provides an overview of the network infrastructure, detailing characteristics such as nodes, link capacity, geographic distribution, and more.

Initially, all delay and traceroute data were collected over six months. Subsequently, all delays for each route were selected and filtered to select only high delays (greater than $60ms$), classifying the results as good and bad, labeled as 0 and 1, respectively. After this filtering, the traceroute closest to the time of occurrence of each high delay was selected. The next step involved selecting fixed times for all high delays to create a correction matrix using traceroute data of the links traversed, compared against the complete end-to-end route. Later, only the low-delay data was selected. The above method was repeated to generate the same data obtained with high delays for comparison purposes and to extract new information.

An important issue is the frequency of data collection. Increasing the frequency of data collection can enhance the accuracy of the proposed method by capturing more detailed variations in delay and improving the precision of correlation analysis. This allows for quicker detection of emerging performance issues and more responsive Impact Score calculations. However, it also introduces scalability challenges, as the higher volume of traceroute and delay data increases the computational and storage demands. To maintain efficiency, future implementations may require optimization strategies such as incremental matrix updates, data aggregation techniques, or distributed processing frameworks. Balancing granularity and scalability is essential to ensure the method remains both accurate and practical in larger or more dynamic network environments.

## 4.2 Results

In this section, three results from the experiments performed will be presented: Defined Correlation Matrices, Link Impact Scores considering only one Time, and the Impact Score considering a time interval, which is discussed in Sections 4.2.1, 4.2.2 and 4.2.3, respectively.

### 4.2.1 Correlation Matrix Definition

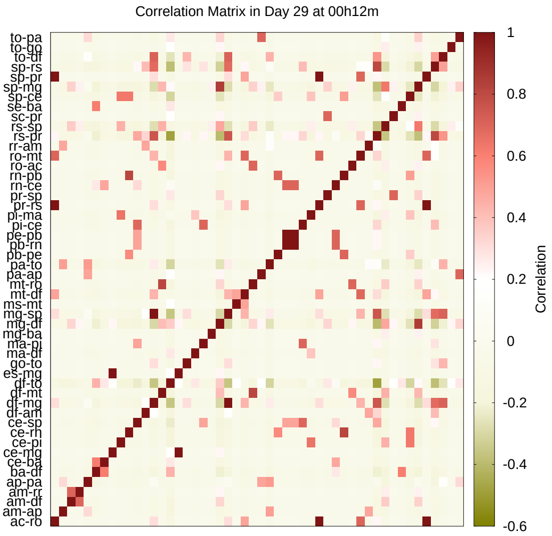
At first, data were collected at two specific times on May 29, 2023, at 00:12 and 00:13. These times were chosen based on the schedule that showed a higher frequency of high delay measurements. Consequently, the resulting matrices differ in size due to the varying quantities of delays classified as high or low at specific times. Figure 3 presents the correlation matrices for these scenarios. In the matrices, the values $C_{i,j}$ represent the correlation coefficient between links utilized in a route and the direct links where high delay measurements occurred within a given timeframe. The *Ipê* Network includes Points of Presence (PoPs) located in the capitals of all Brazilian states, as well as the Federal District, with links labeled by the abbreviations of their origin and destination states.

From the correlation matrices, it can be observed in the values in red that some communication links have a high correlation with several cases of high delay in both measurement times (such as the links $CE-SP$, $AP-PA$, $RS-SP$, $RS-PR$, $DF-MG$, and $SP-MG$). In contrast, other links have a high correlation in one measurement, but a low correlation in another (such as in the cases of the links $TO-GO$, $DF-TO$, and $MG-BA$).

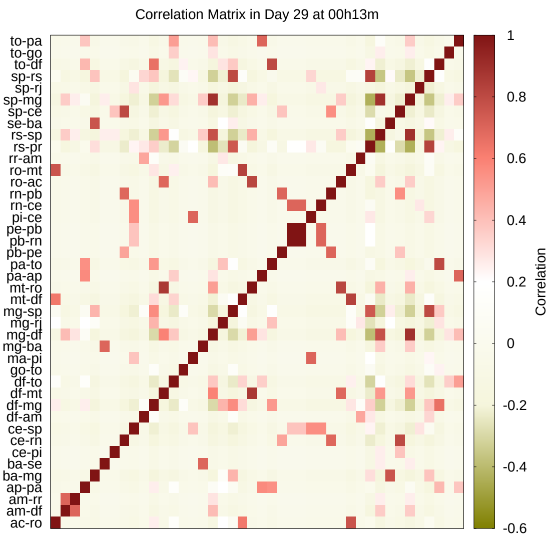
High correlations between links like $CE-SP$ and $RS-SP$ with several high-delay cases suggest their potential significance in understanding network bottlenecks. These correlations can direct attention to specific segments of the infrastructure that consistently exhibit problematic behavior. Conversely, links such as $TO-GO$, which demonstrate variability in correlation across different measurement intervals, underscore the dynamic nature of network performance and suggest that delay issues might be transient or context-dependent.

Despite its utility, the correlation matrix has inherent limitations that restrict its ability to serve as a standalone diagnostic tool. Correlation does not imply causation; thus, while the matrix identifies links associated with high delay, it cannot confirm whether these links are the root causes of the performance issues. For instance, a link might show a high correlation with delays due to its proximity to the actual bottleneck, rather than being the direct source of the problem. Additionally, correlations can be influenced by transient events or outliers, which may not reflect consistent network behavior.

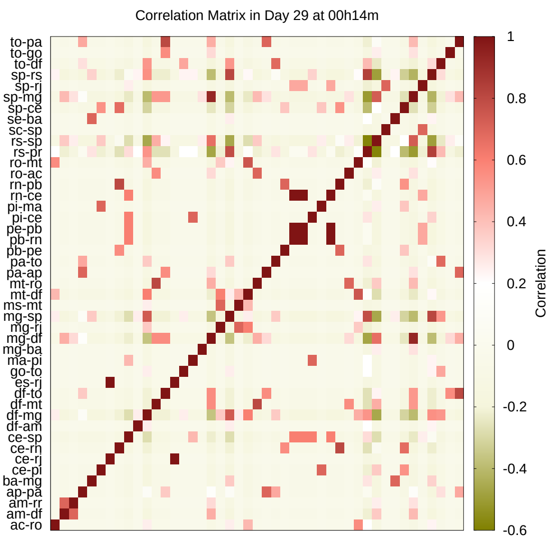
To enhance its effectiveness, the correlation matrix should be used in conjunction with other analytical tools and metrics. Combining the correlation matrix with the proposed temporal analyses (time-weighted impact scores) can help discern patterns over extended periods, reducing the risk of misinterpreting short-lived anomalies.




(a) Correlation Matrix on Period 00:12h.




(b) Correlation Matrix on Period 00:13h.




(c) Correlation Matrix on Period 00:14h.

**Figure 3.** Correlation Matrices Results

#### 4.2.2 Impact Score on Single Period

While the correlation matrix offers useful insights into potential causes of network issues, relying on it alone may not provide a complete understanding of link performance or the root causes of high delay. Using information from the correlation matrix, Figure 4 presents the Scores for links involved during periods of high and low delay. A higher score indicates an increased likelihood that a communication link contributes significantly to delay issues within the network. The addition of the Impact Score refines the identification process, highlighting specific links that may be responsible for increased delays. In Figure 4, we observe the same links with the highest Impact Scores, helping to pinpoint possible problem areas and enabling a more focused resolution for these cases.

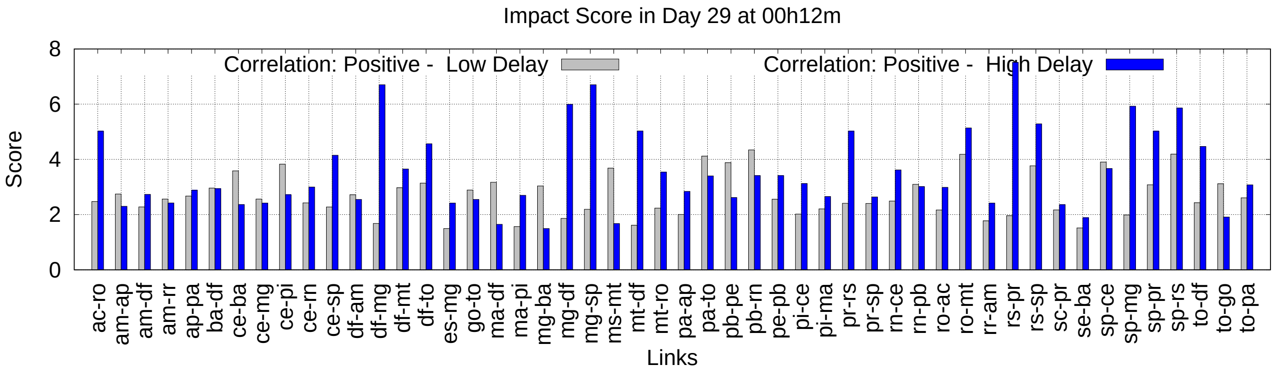
Additionally, when a link's Score for High Delay is greater than its Score for Low Delay, this strongly suggests that the link contributes to performance issues, as its impact is more pronounced in low-performance scenarios. An example of this situation occurs in the links $AC-RO$, $DF-MG$, $MG-SP$, and $RS-PR$, where there's a vast difference between the Score of High Delay and Low Delay. A similar pattern is observed in the links $TO-GO$, $PA-TO$, $PB-PE$, $CE-BA$, and others, that the Score of Low Delay is greater than that of High Delay, indicating that these links are not responsible for the cases of high delay.

However, when analyzing the data in Figure 4, cases of substantial variation in Score between consecutive measurements (i.e., within a one-minute interval) can be observed, as occurs in links like $AC-RO$, $CE-SP$, $MA-PI$, $RR-AM$. This finding suggests that relying on a single measurement point may lead to premature conclusions, as a measurement could capture only a transient instability in the network infrastructure. Therefore, an analysis considering a longer period tends to provide greater efficiency in identifying the causes of performance problems. By extending the evaluation across multiple measurement intervals, network administrators can better differentiate between persistent issues and temporary fluctuations, enabling targeted interventions that enhance overall network performance and user experience.

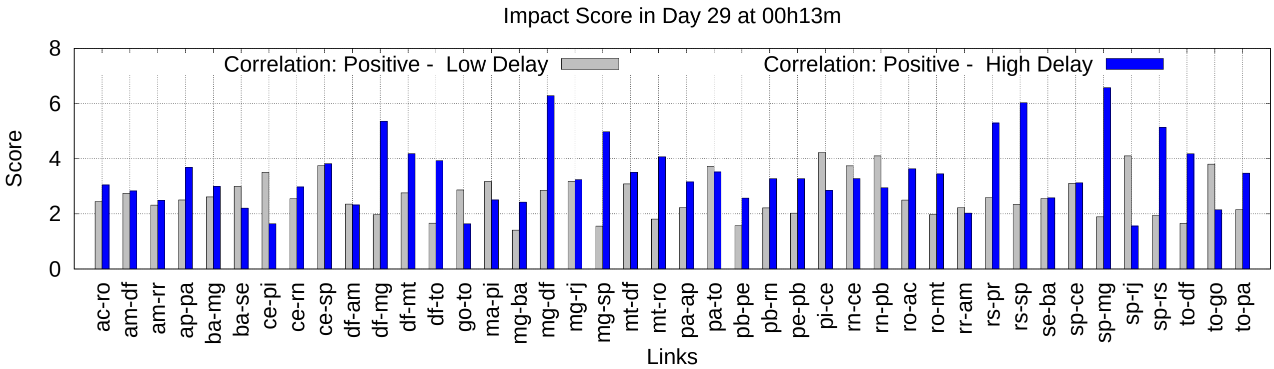
#### 4.2.3 Impact Score with Time Period

In order to enable an analysis of links from a temporal perspective, this article proposed an Impact Score considering a period (as described in Section 3). In the experiments conducted, a 10-minute analysis window was applied, meaning 10 measurements were taken into account (covering data from May 29th between $00h12$ and $00h21$) to determine the Score of a link in both High Delay and Low Delay scenarios. These results are illustrated in Figure5, showing the behavior of the Score of the links at three different times (Figures 5(a) and 5(b)): one measurement ($00h12$), after five measurements ($00h16$) and at the end of ten measurements ($00h21$). Additionally, four specific cases are shown (in Figures 5(c) and 5(d)) to discuss the conclusions arising from the application of the solution proposed in this article.

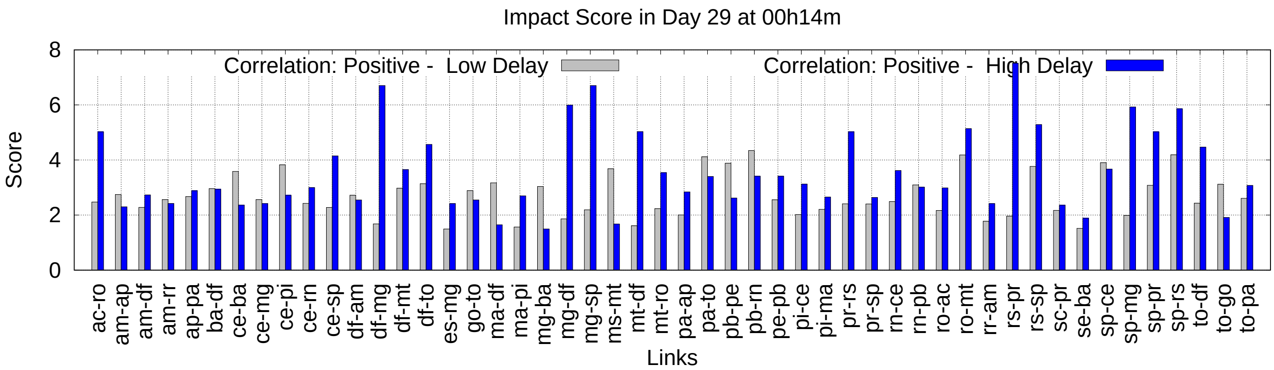
Figures 5(a) and 5(b) show that, when considering a period




(a) Impact Score at Time 00:12h.




(b) Impact Score at Time 00:13h.

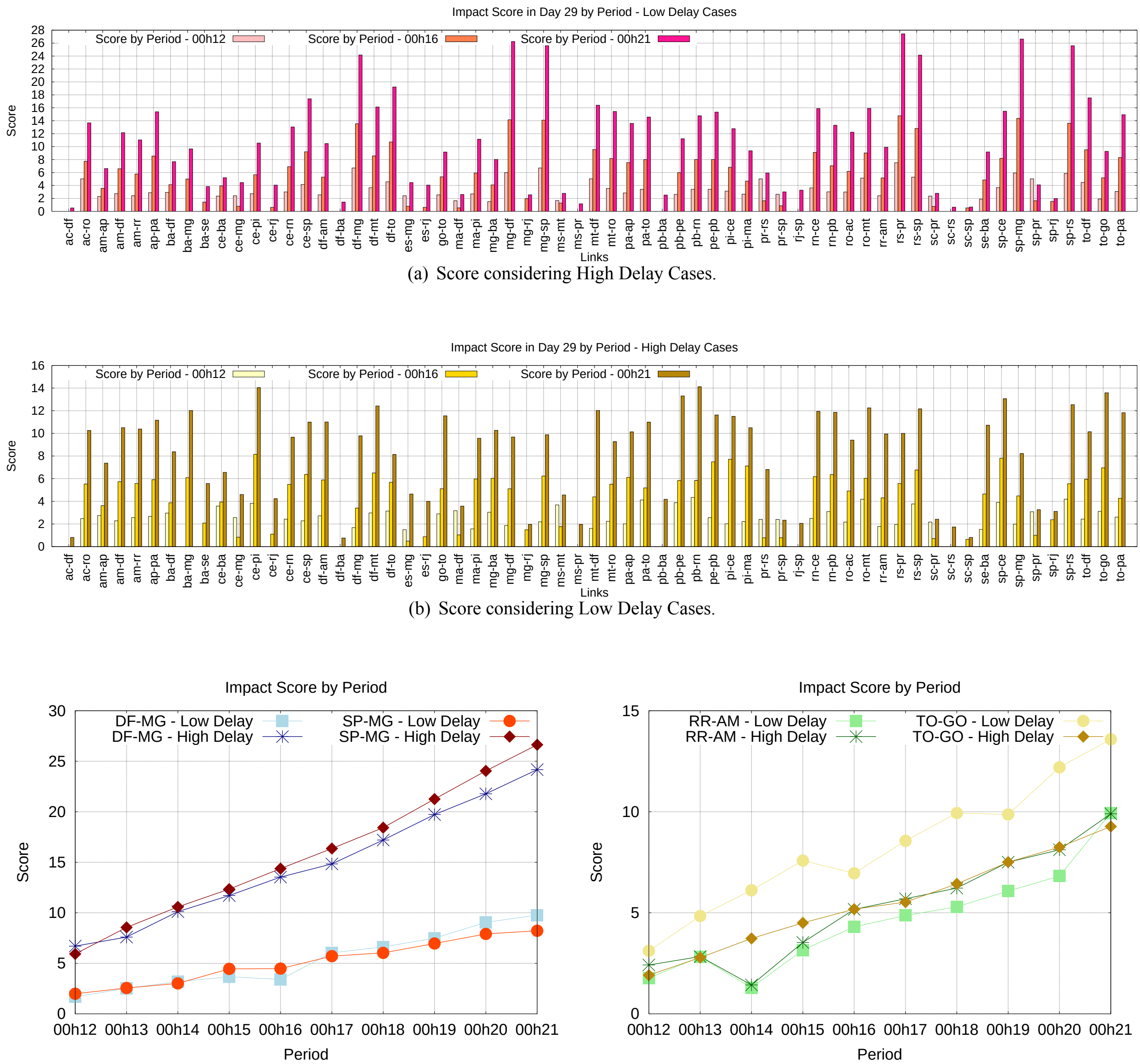



(c) Impact Score at Time 00:14h.

**Figure 4.** Impact Score Results

of time, links with similar Scores diverge over time. This can be observed with links $AP - PA$ and $CE - BA$, which had similar Scores at the beginning of the period ($00h12$) but by the end of the period ($00h21$), the Score for $AP - PA$ was three times higher than that of $CE - BA$ in the High Delay case. Conversely, the analysis of the period also reveals cases of link instability (such as $CE - MG$, $MS - MT$, and $PR - SP$), where the Scores fluctuate positively and negatively. It is worth noting that a Score reduces from one period to the next when the link in question was not used in any communication route at the time of the respective measurement, suggesting a reduced impact of that link on performance.

Among the identified situations, the most critical involves links whose Impact Score for High Delay increases significantly over time, while the Score for Low Delay shows only minimal growth. This pattern is particularly evident in links $DF - MG$ and $SP - MG$, which are highlighted in Figure 5(c). The difference between the Scores (for high and low delay) for these links expands exponentially, indicating a high level of involvement in performance issues and suggesting they may be a (partial or total) cause. In contrast, Figure 5(d) presents two cases where the likelihood of the links being related to performance issues is very low. This conclusion arises from the fact that the Scores for high and low delay remain similar over time (as seen with link $RR - AM$) or the Score for Low Delay exceeds that of High Delay (as




(a) Score considering High Delay Cases.

(b) Score considering Low Delay Cases.

(c) Example of a Problem Indication.

(d) Performance Stability Example.

**Figure 5.** Period of time-based Score

in the case of link $TO - GO$).

Based on the experimental analysis conducted in this study, we conclude that the proposed mechanism offers a significant advantage for ISP management and network performance. The precise identification of potential causes of network underperformance enables managers to optimize performance, reduce costs, and improve user satisfaction. Consequently, our mechanism enables network administrators to make better-informed decisions, allowing them to proactively resolve issues before they affect the network and its users.

## 5 Final discussion

The proposed Time-Weighted Correlation Approach has proven effective in identifying high-delay network links within ISP infrastructures. By employing correlation techniques to create an Impact Score, the methodology enhances traditional monitoring capabilities, moving beyond simple delay detection to provide a more detailed view of delay sources within the network. This improvement can support ISPs in making proactive decisions regarding network infrastructure adjustments and managing high-demand periods with greater precision.

The approach presents several tangible operational benefits. It allows for early identification of critical links that contribute disproportionately to network delays, enabling timely intervention. This feature is especially beneficial for ISP management as it aids in capacity planning and QoS assurance. By identifying problematic links, network administrators can implement targeted actions, such as re-routing or resource allocation, ultimately improving overall network reliability and user satisfaction.

While the Impact Score provides valuable insights into high-delay and low-delay periods, it is important to acknowledge its limitations. The score does not establish causation but rather highlights correlation, which transient conditions or outliers could influence in network behavior. For this reason, further validation steps, such as additional testing under varied network conditions, could enhance the robustness of the Impact Score's reliability and the conclusions drawn from it.

An interesting point is the impact of the organization and size of the network infrastructure on the proposal. In smaller networks or environments with lower traffic volumes, the proposal may offer benefits such as improved interpretability and lower computational overhead. With fewer nodes and links, correlation matrices are simpler and Impact Scores are easier to analyze, potentially enabling faster and more accurate identification of problematic links. Additionally, reduced network complexity enhances the traceability of delay sources, making the method highly suitable for targeted diagnostics in small-scale scenarios. However, limitations arise due to reduced data diversity and less frequent network variation. Fewer delay cases can weaken the statistical reliability of correlation results, and sparse matrices may limit the method's ability to highlight meaningful patterns. To address these issues, adaptations such as longer data collection windows or the incorporation of additional performance metrics (e.g., jitter or packet loss) could improve robustness.

Another important point is the network dynamics, where several challenges may hinder the effectiveness of the proposed solution. One key issue is the frequent variation in traceroute data, which can make it difficult to establish stable correlation patterns over time. As routes change rapidly, the set of links involved in communication paths may vary significantly between measurements, reducing the consistency needed for meaningful statistical analysis. Additionally, the appearance and disappearance of links from communication routes introduce discontinuities in the Impact Score calculation. As previously discussed, the score for a given link decreases when it is not present in any end-to-end route during a measurement period. In highly dynamic environments, such behavior is more frequent, potentially leading to underestimation of a link's impact or misinterpretation of its relevance. This instability can impact the identification of persistent performance issues and may require adaptive mechanisms to maintain the accuracy and reliability of the analysis.

# 6 Conclusion and Future Work

This paper presents an approach for identifying high delays in network connections, introducing the Time-Based Impact Score, which relies on data correlation techniques and information from network monitoring tools. In general, the developed solution allows network administrators to identify whether the network communication links have active and impactful participation in cases of performance problems, considering that they are part of the network communication routes. This approach allows administrators to determine whether heavily utilized links are impacting overall network performance and highlights potential areas within the network infrastructure that may require expansion or increased data transmission capacity.

Experiments using real data from RNP's *Ipê* Network confirm the effectiveness of the proposed mechanism in identifying specific network links that contribute to significant end-to-end delays, thereby helping to mitigate performance issues within the network and its services. These results underscore the utility of the proposed approach, highlighting its potential impact on enhancing performance and optimizing network management.

In future work, we will explore additional features for the Impact Score, such as incorporating machine learning models to predict delay trends based on historical data. Additionally, implementing real-time alerting systems when specific Impact Score thresholds are met could enhance its utility as a network management tool. By integrating these refinements, the solution could further optimize network performance management in ISP environments and sectors such as smart cities and industrial IoT, where network efficiency is paramount. Additionally, another future work is the inclusion of additional network performance metrics (such as bandwidth usage, jitter, or packet loss) could be explored to enhance the robustness and precision of the proposed Impact Score methodology. By integrating these complementary data sources into the correlation framework, the analysis could capture a more holistic view of network behavior, enabling the detection of performance issues caused not only by delay but also by congestion or quality fluctuations. This multi-metric approach would allow for the development of a more nuanced impact assessment model, potentially leading to better prioritization of critical links and more accurate root-cause analysis.

# Declarations

## Acknowledgements

The authors would like to thank the National Council for Scientific and Technological Development (CNPq) of Brazil (Numbers 303877/2021-9 and 405940/2022-0) and CAPES (Number 88887.954253/2024-00) of Brazil for the financial support.

## Funding

This research was funded by the National Council for Scientific and Technological Development (CNPq) of Brazil (Numbers 303877/2021-9 and 405940/2022-0) and CAPES (Number 88887.954253/2024-00) of Brazil.

## Authors' Contributions

All authors contributed to the development of the proposal and the writing of this manuscript.

## Competing interests

The authors declare no conflicts of interest.

## Availability of data and materials

The code developed, as well as the data used in the experiments, are available in the project repository (https://github.com/LarcesUece/Correlation-Analysis-of-Network-Monitoring-Data/) and with the necessary instructions for reproducibility.

# References


Alam, M. I. I., Anshelevich, E., Kar, K., and Yuksel, M. (2024). Pricing for efficient traffic exchange at ixps. *IEEE/ACM Transactions on Networking*, 32(3):2053–2068. DOI: 10.1109/TNET.2023.3336352.

Arachchige, K. G., Branch, P., and But, J. (2023). Evaluation of correlation between temperature of iot microcontroller devices and blockchain energy consumption in wireless sensor networks. *Sensors*, 23(14). DOI: 10.3390/s23146265.

BinSahaq, A., Sheltami, T., Mahmoud, A., and Nasser, N. (2022). Fast and efficient algorithm for delay-sensitive qos provisioning in sdn networks. *Wireless Networks*, pages 1–22. DOI: 10.1007/s11276-022-03028-3.

Costa, M. A., Costa, Y. M., Almeida, Y. O., Cardoso, F. J., and Gomes, R. L. (2024). Connection management using automated firewall based on threat intelligence. In *Proceedings of the 2024 Latin America Networking Conference*, LANC '24, page 32–37, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3685323.3685331.

Ferreira, M. C., Ribeiro, S. E., Nobre, F. V., Linhares, M. L., Araújo, T. P., and Gomes, R. L. (2024). Mitigating measurement failures in throughput performance forecasting. In *2024 20th International Conference on Network and Service Management (CNSM)*, pages 1–7. DOI: 10.23919/CNSM62983.2024.10814394.

Gajewski, M., Mongay Batalla, J., Mastorakis, G., and Mavromoustakis, C. X. (2022). Anomaly traffic detection and correlation in smart home automation iot systems. *Transactions on Emerging Telecommunications Technologies*, 33(6):e4053. DOI: 10.1002/ett.4053.

Guo, J., Ning, X., Hua, C., Yang, J., and Cai, Z. (2025). A path statistical delay prediction framework based on global graph neural network. *IEEE Transactions on Circuits and Systems I: Regular Papers*, pages 1–13. DOI: 10.1109/TCSI.2025.3542438.

Hekmati, A., Zhang, J., Sarkar, T., Jethwa, N., Grippo, E., and Krishnamachari, B. (2024). Correlation-aware neural networks for ddos attack detection in iot systems. *IEEE/ACM Transactions on Networking*, 32(5):3929–3944. DOI: 10.1109/TNET.2024.3408675.

Imran, Zuhairi, M. F. A., Ali, S. M., Shahid, Z., Alam, M. M., and Su'ud, M. M. (2023). Improving reliability for detecting anomalies in the mqtt network by applying correlation analysis for feature selection using machine learning techniques. *Applied Sciences*, 13(11). DOI: 10.3390/app13116753.

Kim, J., Moon, Y., and Ko, H. (2024). Correlation-based advanced feature analysis for wireless sensor networks. *The Journal of Supercomputing*, 80(7):9812–9828. DOI: 10.1007/s11227-023-05739-6.

Kim, Y., Kim, T.-H., and Ergün, T. (2015). The instability of the pearson correlation coefficient in the presence of coincidental outliers. *Finance Research Letters*, 13:243–257. DOI: 10.1016/j.frl.2014.12.005.

Li, W., Wang, X., Zhang, Y., and Wu, Q. (2021). Traffic flow prediction over muti-sensor data correlation with graph convolution network. *Neurocomputing*, 427:50–63. DOI: 10.1016/j.neucom.2020.11.032.

Mok, R. K. P., Zou, H., Yang, R., Koch, T., Katz-Bassett, E., and Claffy, K. C. (2021). Measuring the network performance of google cloud platform. In *Proceedings of the 21st ACM Internet Measurement Conference*, IMC '21, page 54–61, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3487552.3487862.

Portela, A., Linhares, M. M., Nobre, F. V. J., Menezes, R., Mesquita, M., and Gomes, R. L. (2024a). The role of tcp congestion control in the throughput forecasting. In *Proceedings of the 13th Latin-American Symposium on Dependable and Secure Computing*, LADC '24, page 196–199, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3697090.3699869.

Portela, A. L. C., Ribeiro, S. E. S. B., Menezes, R. A., de Araujo, T., and Gomes, R. L. (2024b). T-for: An adaptable forecasting model for throughput performance. *IEEE Transactions on Network and Service Management*, pages 1–1. DOI: 10.1109/TNSM.2024.3349701.

Sawabe, A., Shinohara, Y., and Iwai, T. (2024). Rethinking delay behavior in mobile networks as a lifeline of industrial applications. In *GLOBECOM 2024 - 2024 IEEE Global Communications Conference*, pages 1803–1808. DOI: 10.1109/GLOBECOM52923.2024.10901521.

Scarpitta, C., Sidoretti, G., Mayer, A., Salsano, S., Abdelsalam, A., and Filsfils, C. (2023). High performance delay monitoring for srv6 based sd-wans. *IEEE Transactions on Network and Service Management*, pages 1–1. DOI: 10.1109/TNSM.2023.3300151.

Silva, D., Nobre, F., Ferreira, M., Portela, A., Araújo, T., and Gomes, R. (2024). Identificação das causas de situações de alto atraso em provedores de internet. In *Anais do XVI Simpósio Brasileiro de Computação Ubíqua e Pervasiva*, pages 111–120, Porto Alegre, RS, Brasil. SBC. DOI: 10.5753/sbcup.2024.2881.

Silveira, M., Santos, D., Souza, M., Silva, D., Mesquita, M., Neto, J., and Gome, R. L. (2023). An anonymization service for privacy in data mining. In *Proceedings of the 12th Latin-American Symposium on Dependable and Secure Computing*, LADC '23, page 214–219, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3615366.3625074.

Souza, M. S., Ribeiro, S. E. S. B., Lima, V. C., Cardoso, F. J., and Gomes, R. L. (2024). Combining regular expressions and machine learning for sql injection detection in urban computing. *Journal of Internet Services and Applications*, 15(1):103–111. DOI: 10.5753/jisa.2024.3799.

Stafecka, A., Lizunovs, A., Olins, A., Rjumsins, M., Spolitis, S., and Bobrovs, V. (2024). Defining internet access service qos measurement server placement criteria in national internet network. In *2024 Photonics Electromagnetics Research Symposium (PIERS)*, pages 1–6. DOI: 10.1109/PIERS62282.2024.10617842.

Wang, B., Lun, S., Li, M., and Lu, X. (2024). Echo state network structure optimization algorithm based on correlation analysis. *Applied Soft Computing*, 152:111214. DOI: 10.1016/j.asoc.2023.111214.