International Journal of Innovative Technology and Interdisciplinary Sciences
https://journals.tultech.eu/index.php/ijitis
ISSN: 2613-7305
Volume 8, Issue 3
DOI: https://doi.org/10.15157/IJITIS.2025.8.3.374-387
Received: 30.07.2025; Revised: 22.08.2025; Accepted: 27.08.2025

Research Article

Mapping ISP Malware Trends in Albania: Clustering for Smarter Cyber Defences

Klorenta Pashaj[1*], Eralda Gjika[2]

[1] Directorate of Monitoring and Incidence Response, Operations Center SOC C-SIRT, National Cyber Security Authority of Albania, Tirana, Albania
[2] Faculty of Computer Engineering and IT, Metropolitan University Tirana, Tirana, Albania
***klorenta.pashaj@gmail.com**

Abstract

Cybersecurity plays a vital role in protecting digital infrastructure, with Internet Service Providers (ISPs) standing at the core of this ecosystem. This research takes an exploratory perspective, given the limitations of both the dataset and the number of available features. The analysis draws on malware detection data from Albania, reported through the Shadow Server platform, covering a 15-month period across seven ISPs. By applying time-series clustering alongside statistical methods, the study groups ISPs according to their security patterns. The time-series analysis points to three distinct periods of heightened malware activity, while the characteristics-based approach identifies three groups of ISPs that differ in their vulnerability profiles. Taken together, these results underline the need for customized cybersecurity strategies and stronger cooperation among ISPs. Despite the constraints of a relatively small dataset, clustering techniques prove useful for optimizing resources, supporting regulatory compliance, and informing strategic decisions aimed at more effective threat prevention and mitigation.

Keywords: Cybersecurity; ISPs; Clustering; Trend Analysis; Vulnerability Patterns; Threat Mitigation

INTRODUCTION

The growing complexity and frequency of cyber-attacks call for increasingly sophisticated defence strategies, particularly for Internet Service Providers (ISPs), which serve as the backbone of digital infrastructure. Among the approaches available, clustering analysis of cyber-attack data offers a valuable means of uncovering patterns and recurring vulnerabilities. By identifying these trends, ISPs can better anticipate threats and strengthen their resilience. This paper examines the role of clustering methods in enhancing ISP security and highlights their broader significance for improving cybersecurity practices.

Although global cybersecurity has received increasing attention, research that focuses specifically on Albania remains scarce. Among the limited studies, authors at [1] underscore the urgent need to safeguard the country's information infrastructure. In a related work, authors in [2] examine how cybersecurity measures in Albania have evolved

in response to emerging threats. Furthermore, authors in [3] emphasize the importance of proactive strategies for mitigating network-level threats and malware, with a particular focus on systematic monitoring within the Albanian context. Building on this foundational understanding of Albania's cybersecurity landscape, the present study extends the discussion by analysing malware activity traced to Albania through detections associated with the country's Internet Service Providers.

Clustering ISPs according to the frequency and characteristics of malware detections in Albania makes it possible to uncover common patterns of vulnerability and recurring trends in malicious activity. Through the application of clustering techniques, ISPs can strengthen their overall security posture, allocate resources more effectively, and gain insight into broader threat dynamics. This approach also facilitates the design of tailored cybersecurity measures that address the specific risks and needs of each identified cluster.

Prior research highlights the usefulness of clustering in cybersecurity, particularly for uncovering attack patterns and system vulnerabilities. For instance, authors in [4] demonstrated how clustering methods can be applied to network traffic to detect anomalies and identify potential cyber threats. Similarly, authors in [5] examined the role of machine learning, including clustering techniques, in enhancing the detection and mitigation of distributed denial-of-service (DDoS) attacks within ISP networks [6-10]. While such studies illustrate the promise of clustering, research that focuses specifically on clustering Internet Service Providers remains scarce [11-15]. The present study seeks to address this gap by applying clustering techniques to ISP data, offering a more detailed perspective on their value for strengthening cybersecurity strategies and practices.

The dataset for this study covers a 15-month period and includes malware detections reported by the Shadowserver platform from seven Internet Service Providers in Albania. By applying time-series clustering and statistical analysis, ISPs were grouped according to similarities in their security profiles. The results point to the necessity of adopting tailored cybersecurity measures while also reinforcing the value of joint defence initiatives across providers. In practice, clustering proves useful not only for improving the allocation of resources but also for supporting regulatory compliance and guiding strategic choices that strengthen threat management and mitigation.

The paper is organized into five main parts. First, it reviews the existing literature and sets out the motivation behind this study. Next, it explains the methodology, describing the process of data collection, the clustering approach, the statistical techniques applied, and the visualization tools employed. This is followed by the presentation of results, including descriptive statistics, clustering outcomes, identified trends, and proposed tailored solutions. Finally, the paper concludes by summarizing the main findings and suggesting avenues for future research.

LITERATURE REVIEW AND STATE OF THE ART

Recent research has shown the value of clustering techniques in detecting patterns within cyberattack data and strengthening threat detection. For example, authors in [14]

applied Gaussian Mixture Models and Spectral Clustering to more than 9,000 cyberattack incidents. Their dataset, derived from a selective filter of over one million news articles, examined the relationships between 95 cyberattacks and 29 different industry sectors. Similarly, authors in [12] assessed both traditional and emerging clustering algorithms for User and Entity Behavior Analytics (UEBA) within SIEM platforms, drawing on data from existing literature and testing fifteen clustering methods. Their findings indicated that HDBSCAN and DenMune achieved strong performance, particularly when applied to the CERT behaviour-related dataset. In another study, authors in [13] proposed quantum clustering frameworks that outperformed traditional methods by delivering greater accuracy in vulnerability clustering.

While these studies underscore the importance of clustering techniques in cybersecurity, they are primarily oriented toward large-scale environments. Most rely on synthetic datasets, standardized collections for benchmarking, or generalized threat models that lack specificity to real-world operational contexts. Consequently, their findings have limited applicability when considering environments such as Internet Service Providers.

This study contributes to the field by applying clustering techniques directly to real-world ISP data, focusing on malware detections originating from Albania over a 15-month period. It introduces a two-level clustering framework: time-series clustering to capture temporal trends in malware activity and feature-based clustering to group ISPs by their distinct vulnerability profiles. Beyond methodological innovation, the research demonstrates practical value by providing insights that can guide customized defence mechanisms, improve the allocation of resources, and reinforce regulatory compliance within the regional ISP ecosystem.

METHODOLOGY

This research applies a comprehensive methodological framework, shaped around a quantitative and explanatory approach. Several complementary techniques were employed data collection, statistical analysis, clustering, and visualization each contributing to a clearer understanding of ISP behaviour. The process unfolded in three stages:

Stage I: Data Collection and Preparation. The dataset was compiled from the annual reports of the National Authority for Cyber Security (NCSA) in Albania and supplemented by monitoring conducted through the Shadowserver platform. Seven ISPs were included, with monthly malware detection counts recorded over a 15-month period (May 2021–July 2022). Since the dataset was complete and required no preprocessing, the analysis could proceed with a high level of confidence in the accuracy of the values.

Stage II: Statistical Analysis and Visualization. To establish a baseline understanding of the dataset, descriptive statistics (mean, standard deviation, minimum, and maximum values) were calculated for each ISP, following approaches similar to those outlined in [15] to ensure robust group comparisons. These were followed by graphical representations

boxplots and similarity graphs which helped highlight differences in variability and common behavioural trends. This stage not only summarized the dataset but also offered preliminary insights into how ISPs may align or diverge in their security patterns.

Stage III: Clustering Analysis and Identification of Patterns. Finally, time-series clustering was used to explore malware detection patterns across the ISPs. The elbow method was applied to determine the optimal number of clusters, which allowed ISPs with comparable vulnerability profiles to be grouped together. Interpreting these clusters provided insights into recurring weaknesses and shared exposures, paving the way for recommendations on sector-wide improvements.

PROPOSED APPROACH AND RESULTS

When analysing similarity patterns in time-series data, it is often useful to adopt more than one analytical perspective. In this study, two approaches were applied: clustering by temporal cohorts and clustering based on ISP-specific characteristics. While the relatively small sample size inevitably introduces some limitations, the analysis nevertheless reveals meaningful trends. More importantly, it provides a framework that can be adapted to similar contexts where larger datasets are available. Accordingly, the hypotheses presented below are formulated without being constrained by the sample size, though it is worth noting that in several cases the results point to clear and consistent patterns.

ISP clusters can be formed using various criteria, such as the demographic characteristics of their users or specific attributes of the providers themselves speed, customer service, pricing models, and similar factors as noted by [6]. Given the structure of the dataset and the research objectives, the following hypotheses were developed to guide the analysis:

H1. Significant clustering occurs across three distinct time intervals. This assumes that ISPs can be grouped according to the number of malware detections observed within their networks. By testing this, the study seeks to confirm whether malware activity follows recurring temporal patterns.

H2. ISP-specific behaviours influence clustering outcomes. Here the expectation is that features such as security measures, response speed, and user demographics shape the way ISPs cluster. Providers with weaker safeguards or particular traffic characteristics may appear in groups with higher detection counts.

H3. Clustering patterns are consistent across consecutive years. If clusters remain stable across time, it would suggest that the underlying drivers of ISP vulnerability do not change rapidly. Stability would also make clustering a useful tool for anticipating future trends, although shifts in preventive measures could still alter detection levels.

H4. Malware detections vary according to identifiable monthly patterns. The hypothesis is that detection counts do not fluctuate randomly but reflect seasonal dynamics, special events, or operational adjustments by the ISPs.

H5. Malware detection follows cyclical trends. This assumes that certain months repeatedly show higher or lower levels of malware activity, indicating cycles that may be linked to attack timing, periodic system updates, or recurring malicious campaigns. Recognizing these cycles could help ISPs anticipate and prepare for peaks in activity.

Building on the correlations and behavioural patterns identified, it is possible to design forecasting systems that predict the intensity of future malware detections by analyzing temporal relationships in historical data. For instance, authors in [9] demonstrated the effectiveness of classical time-series models, such as ARIMA, which significantly outperformed naïve forecasting methods and dynamically updated predictions to improve accuracy. Similarly, authors in [10] applied ARIMA, SARIMA, GARCH, and Bootstrapping models to predict SSH attacks on honeypots, showing the utility of such approaches for anticipating attacks and identifying regional attack patterns. More recently, authors in [11] applied deep sequence models particularly the LSTM Encoder–Decoder architecture and found notable improvements in anomaly prediction, with accuracy gains of more than 11% when outlier adjustments were included.

However, given the relatively low dimensionality and limited consistency of our dataset, advanced time-series models such as ARIMA or LSTM were not applied in this study. These models typically require larger datasets and stronger temporal dependencies to deliver reliable forecasts. Instead, the aim of our analysis was not to predict exact values but to derive insights into ISP behaviour and examine how malware detection patterns vary across months and seasons.

Time cohort clusters

As noted in the data presentation, the relatively small sample size limits the use of certain time-series metrics—such as autocorrelation and stationarity tests—that are more suitable for larger datasets when analysing attack frequency trends. Nevertheless, the aim of this study is to maximize the insights that can be drawn from the available data in order to better anticipate and mitigate high-risk behaviour in the ISPs under consideration, as well as in providers with comparable characteristics. The dataset, which spans 15 months from May 2021 to July 2022 at a monthly frequency, is evenly distributed across two calendar years, allowing for meaningful observation of temporal patterns.

Figure 1 illustrates the behaviour of the seven ISPs over the 15-month observation period. The results show that while some ISPs maintain relatively stable patterns, others display more pronounced fluctuations in malware detections. Notably, ISP 2 and ISP 4 exhibit a symmetrical trend: when one records higher values, the other tends to show lower values, and vice versa. This inverse relationship highlights the dynamic nature of their activity over time.

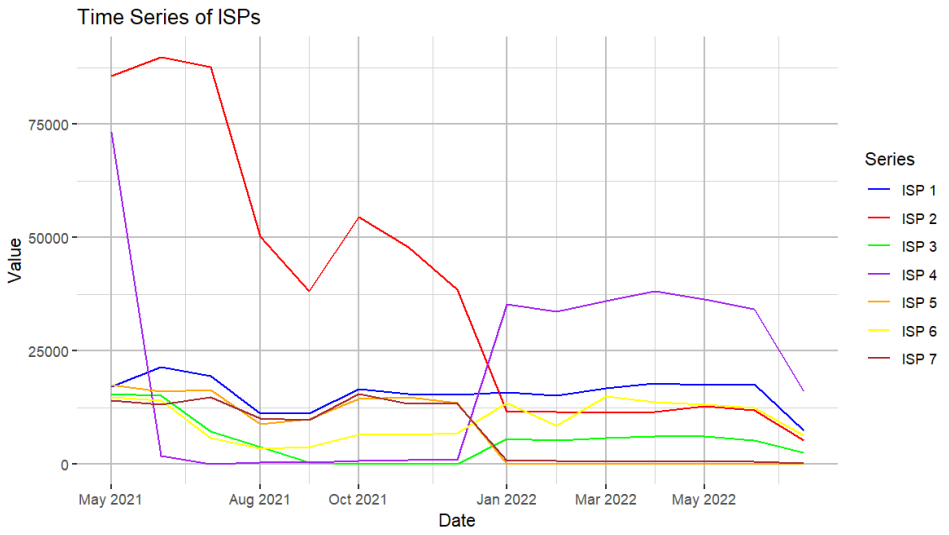


Figure 1. Time Series of ISPs

Descriptive statistics

Data for this study were collected from seven Internet Service Providers (ISPs) operating in Albania. The dataset records monthly counts of malware detections for each provider and was complete, requiring no pre-processing. Descriptive statistics were then computed to summarize the dataset and generate initial insights to guide the modelling process, see Table 1.

Table 1. Descriptive statistics

ISP	Mean	SD	Min	Max
ISP 1	15742.1	3521.5	7448.0	21526.0
ISP 2	37903.9	30625.9	5223.0	89745.0
ISP 3	5269.7	4800.3	51.0	15494.0
ISP 4	20559.3	22237.1	114.0	73210.0
ISP 5	7447.3	7506.1	3.0	17508.0
ISP 6	9630.2	4257.9	3442.0	15024.0
ISP 7	7243.2	6579.6	263.0	15551.0

Boxplots offer an effective way to visualize malware detection trends across ISPs over time. They provide a concise summary of distribution, variability, and anomalies, making it easier to identify patterns and compare behaviours among providers. The results show that ISP 2 and ISP 3 experienced notable shifts in the range of malware detections during the observation period. This shift appears linked to notifications issued by the National Cyber Security Authority, which alerted ISPs to detected malware. These communications likely encouraged the adoption of stronger security measures and protocols, leading to a marked decline in detections. Similar behavioural adjustments are evident among other ISPs, particularly in the second year of observation, suggesting that the Authority's

proactive engagement and guidance played a broader role in improving cybersecurity practices.

The boxplot of normalized data, used here to allow comparisons across ISPs, supports the hypothesis that malware detections for each provider vary in predictable monthly patterns. By examining the monthly descriptive statistics, these recurring trends can be identified and used to better understand the behaviour of individual ISPs over time, see Figure 2.

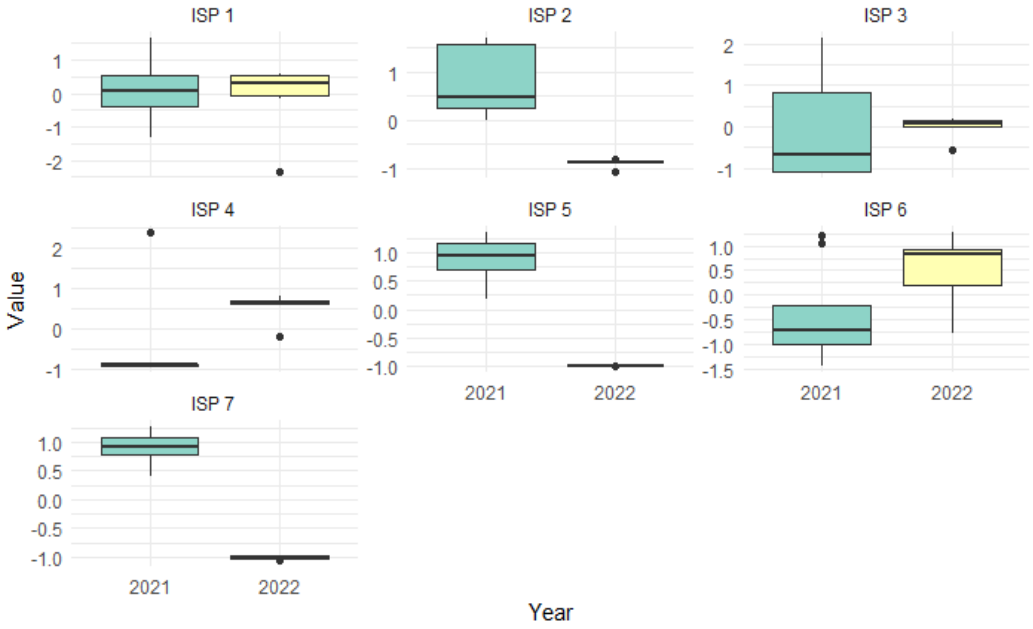


Figure 2. Normalized data Box Plot of ISPs by year

Identifying time cohort clusters involves grouping time-series data that display similar patterns or trends across defined intervals. The time-series plot (Figure 1) reveals non-stationary behaviour across all ISPs. To analyse these patterns, we applied K-means clustering using Lloyd's algorithm, a widely recognized implementation of this method, to the monthly malware detection counts of the ISPs. The objective was to identify cohorts of ISPs with comparable detection behaviours over time. For each month, the total malware detection counts were grouped into clusters, with each cluster representing providers that shared similar temporal patterns. This approach offers valuable insights for strengthening cybersecurity monitoring and response strategies. To determine the optimal number of clusters, we employed the elbow method. Although the sample size was relatively small, both the visual inspection and the elbow method pointed to three as the most appropriate number of clusters.

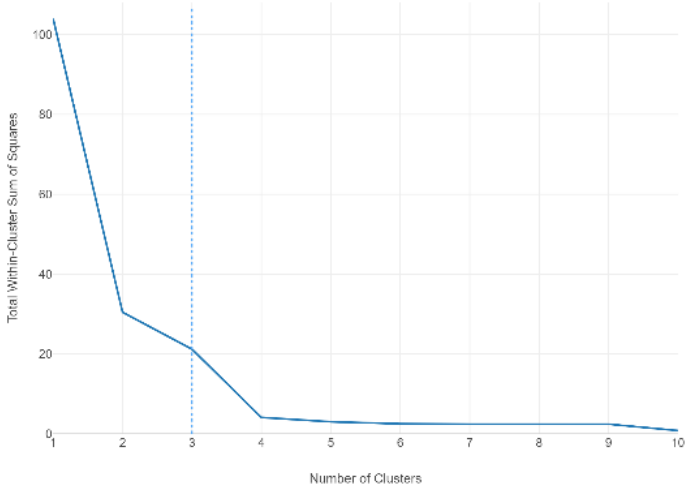
Figure 3 depict the optimal number of clusters and temporal distribution of Cohort clusters. Time-series clustering revealed three distinct clusters of time cohorts:

- *Cluster 1: July 2021, January 2022, and July 2022*

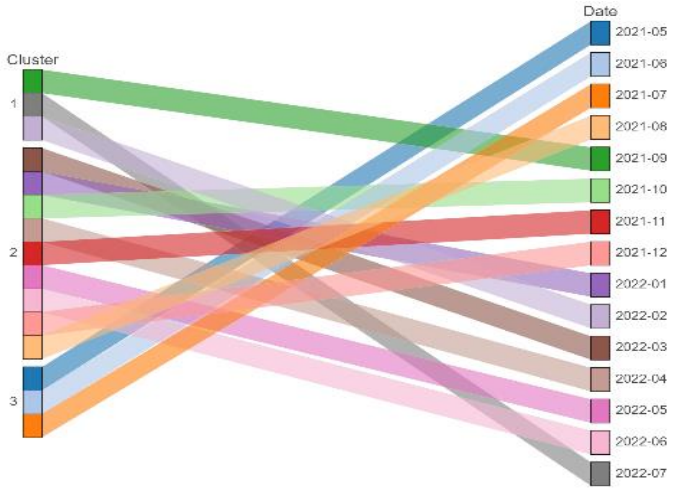
- *Cluster 2: August 2021 to June 2022 (excluding January 2022)*
- *Cluster 3: May and July 2021*

These intervals correspond to key phases in the dataset: the initial observation period, a mid-term phase during which ISPs began strengthening their practices in response to notifications from the National Cyber Security Authority, and a later stage where the impact of these measures became more evident. This clustering highlights the progression of ISP responses to detected malware and illustrates how their security behaviours evolved over time.

To extend the analysis, normalized boxplots were produced for each ISP across the identified periods. By normalizing the data, we were able to compare distributions directly, allowing clearer identification of similarities and differences in malware detection patterns among providers.



(a)



(b)

Figure 3. (a) - Optimal number of clusters and (b) - temporal distribution of Cohort clusters.

In the Figure 4, the normalized boxplots indicate that ISP behaviours within each time period followed broadly similar patterns by allowing both commonalities and anomalies to be identified. In the first period (Cluster 3), the range of malware detection counts was relatively narrow, suggesting limited variation across ISPs, with the exception of ISP 4—an outlier also noted in the time-series plot. In contrast, Cluster 2 displayed a wider range of values, reflecting considerable fluctuation in detection counts among providers. Another notable observation is that the clusters are closely tied to mid-year intervals, with January and July consistently grouped together. This pattern suggests that specific actions or events may have occurred during these months, shaping the malware detection trends observed.

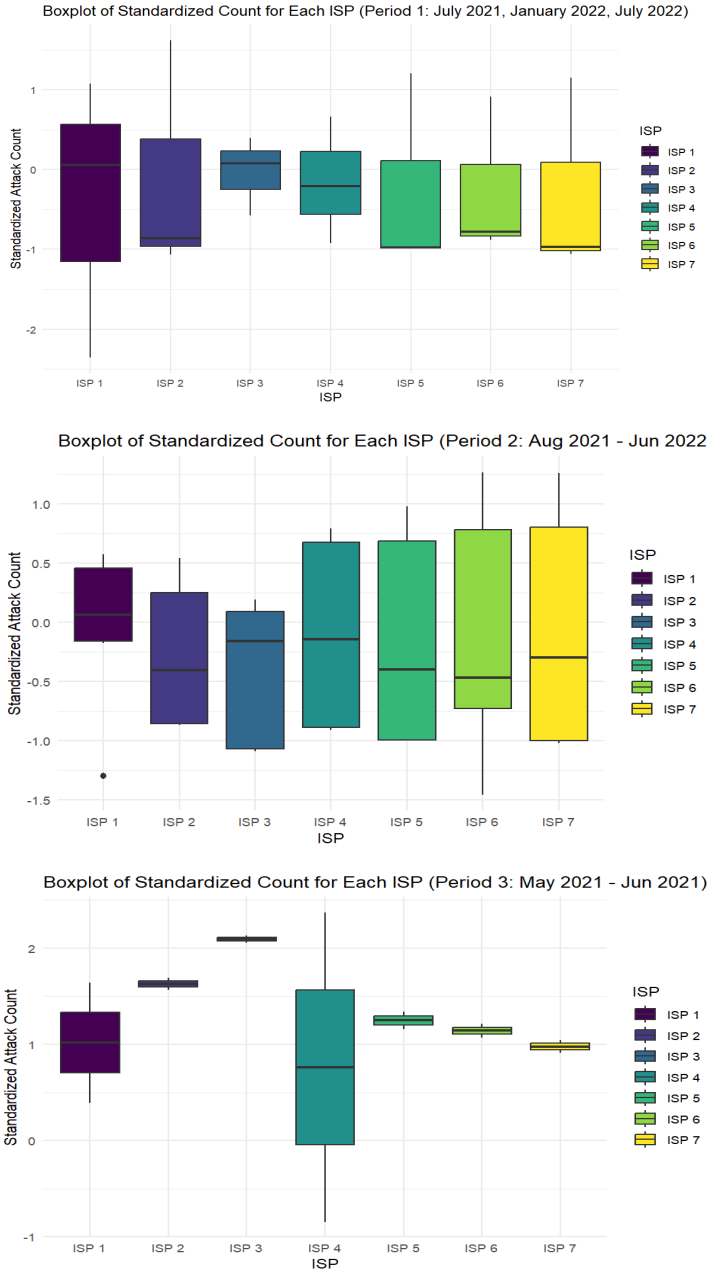


Figure 4. Normalized data Boxplot of ISPs across time cohort clusters

ISPs Clusters

The similarity graph illustrates how ISPs relate to one another based on the clustering results, see Figure 5. ISPs that appear closely connected in the graph exhibit comparable patterns in the number of malware detections. From this visualization, three distinct groups can be identified: a main cluster consisting of ISP1, ISP3, ISP5, ISP6, and ISP7, and two separate groups represented individually by ISP2 and ISP4. This outcome supports our hypothesis that ISP-specific factors such as security measures, response times, and user base characteristics play a significant role in shaping clustering behaviour.

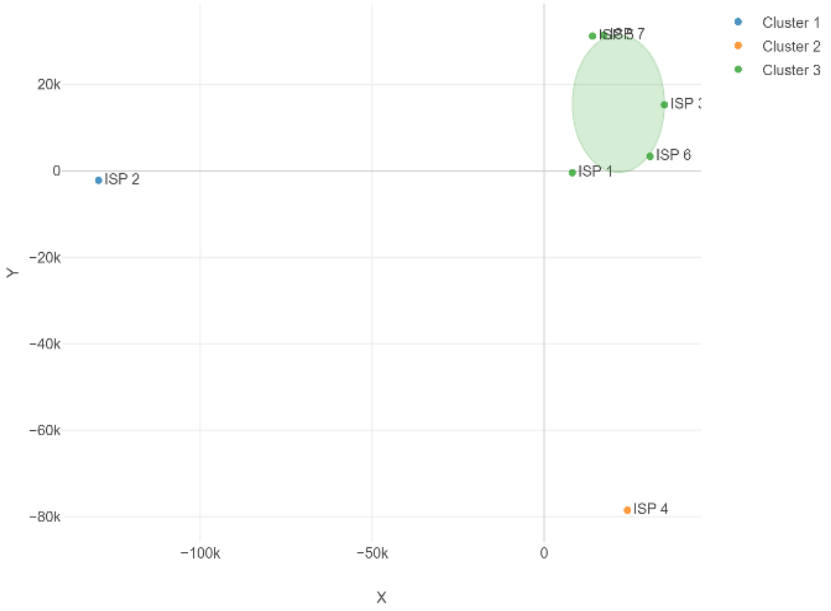


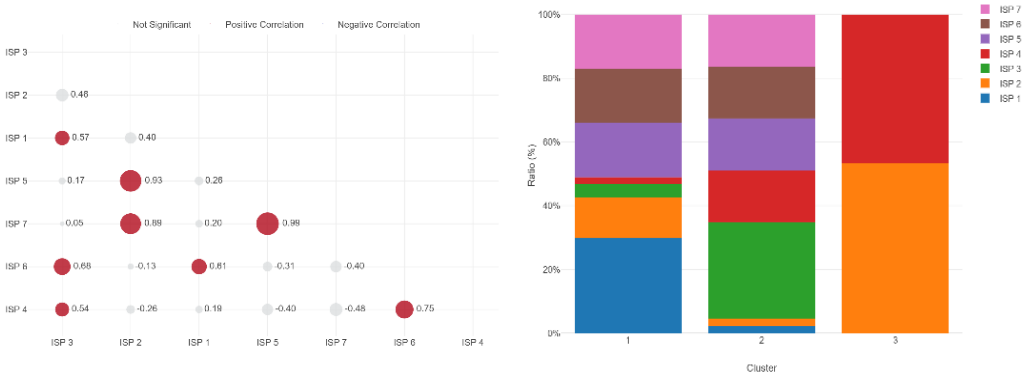
Figure 5. 2D Similarity Map of ISPs Based on Clustering Results

A similarity map, presented as a 2D graph, was used to visualize the clustering of ISPs. In this map, each point represents an ISP, and the distance between points reflects the degree of similarity in their malware detection patterns. This visualization makes it possible to identify clusters, detect outliers, and highlight shared issues across the ISP landscape. The results show that ISP 2 and ISP 4 fall outside the main cluster, exhibiting distinct behaviours when compared to the other ISPs, which form a more cohesive group. Such maps are particularly useful for strategic planning, competitive benchmarking, and pinpointing areas where targeted interventions may be necessary. Interestingly, while the similarity map suggests that ISP 2 and ISP 4 behave differently from the others, correlation values for malware detections tell a more nuanced story.

Figure 6 illustrates these correlations across ISPs for each year, providing a complementary perspective to the clustering results.

Figure 6. Correlation Matrices of ISP Malware Detections

Since the dataset does not contain a complete year of observations for 2022, the clusters identified for that year cannot be interpreted in isolation with full reliability. To address this limitation, the analysis instead focuses on the combined data from 2021 and 2022, providing a more comprehensive basis for identifying potential clustering patterns (Figure 7).



(a) (b)

Figure 7. (a) - Overall Correlation Matrix of ISP Malware Detections counts (b) - Cluster Distribution by ISP Association

The first two graphs in Figure 6 (left to right) present the yearly correlations of malware detections for 2021 and 2022, while Figure 7 shows the overall correlation across the full observation period. The results reveal significant positive correlations among several ISPs, indicating shared vulnerabilities or exposure to similar malicious activity. This pattern suggests that certain malware types or system weaknesses may be affecting multiple providers simultaneously, likely due to common infrastructure dependencies or sector-wide trends.

Notably, ISPs 2, 5, and 7 display very high correlations, with values between 0.89 and 0.93, reflecting strong similarities in their detection patterns. In comparison, ISPs 1, 3, 4, and 6 exhibit moderate correlations ranging from 0.54 to 0.75, suggesting comparable but more varied detection trends. These findings are consistent with the Cluster Distribution by ISP Association plot, which identifies three main clusters: ISPs 1, 2, and 3 most closely aligned with Cluster 1, ISPs 4 and 5 grouped in Cluster 2, and ISPs 6 and 7 forming Cluster 3.

The observed distinctions are likely influenced by several factors, including differences in the implementation and effectiveness of cybersecurity measures, the extent to which ISPs respond to notifications from the National Cyber Security Authority, targeted malware campaigns, and variations in user bases and traffic patterns. Addressing these vulnerabilities is essential for strengthening the overall cybersecurity posture of the sector. Coordinated action and the adoption of shared security practices can play a key role in mitigating common threats, ultimately fostering a more resilient defence against malware activity. Achieving this will require sustained collaboration among ISPs, the systematic application of best practices, and continuous monitoring to adapt to an evolving threat landscape.

CONCLUSION

This study highlights the value of clustering analysis as a strategic approach for enhancing the cybersecurity posture of Internet Service Providers (ISPs) in Albania. By analysing malware detection patterns over a 15-month period, two complementary methods—time-series clustering and characteristics-based clustering were employed to capture both temporal dynamics and distinct vulnerability profiles across ISPs. Although the dataset was relatively small and context-specific, the analysis identified three key periods of heightened malware activity and three distinct ISP groupings. These findings demonstrate that even in data-limited environments, statistical and clustering techniques can generate actionable insights. More broadly, the results underscore the need for sector-wide collaboration, the implementation of shared defence mechanisms, and the adoption of tailored security strategies to address common vulnerabilities effectively

Beyond its immediate findings, this study offers a transferable methodological framework that can be applied to other national contexts or sectors where real-world detection data are available. The results demonstrate that even relatively limited datasets, when examined through rigorous clustering techniques, can support strategic decision-making, improve resource allocation, and reinforce regulatory compliance. Future research should expand on this foundation by incorporating larger datasets, integrating contextual factors such as network architecture and incident response practices, and applying advanced forecasting models to anticipate shifts in malware activity. Such efforts would enhance predictive capabilities and enable more proactive, data-driven defence strategies at both national and regional levels.

AUTHOR CONTRIBUTIONS

Conceptualization, K.P.; Methodology, K.P. and E.G.; Validation, K.P. and E.G.; Investigation, K.P.; Resources, K.P.; Data Curation, K.P.; Writing – Original Draft Preparation, K. and E.G.; Writing – Review & Editing, E.G.; Supervision, E.G.

CONFLICT OF INTERESTS

There is no conflict of interest associated with this publication.

REFERENCES

1. Pashaj K., Gjika E., & Basha L. The Importance of Critical Information Infrastructure Protection – Case of Albania. *Journal of Natural Sciences*, **2024**, 35, 278-293.
2. Pashaj K, Tomco V, Gjika E. From threat to response: The evolution of cybersecurity in Albania. *Smart Cities and Regional Development Journal*, **2025**, 2(1), 1-10
3. Pashaj, K., Tomco, V. & Gjika, E. Strategic Monitoring and Proactive Measures for Mitigating Network-Level Threats and Malware in Albania. *Proceedings Book of the 9th International Engineering Symposium on Advanced Engineering Days,* Tabriz, Iran**, 2024**, pp 901-903.
4. N. Moustafa, B. Turnbull and K. -K. R. Choo, "An Ensemble Intrusion Detection Technique Based on Proposed Statistical Flow Features for Protecting Network Traffic of Internet of Things," *in IEEE Internet of Things Journal*, **2019**, 6(3), 4815-4830.
5. Fouladi, R. F., Afzal, S., & Shabtai, A. Applying Machine Learning Techniques to Detect and Prevent DDoS Attacks in ISP Networks. *IEEE Transactions on Information Forensics and Security*, **2020,** 15, 1248-1260.
6. Paulino, E.P., & Esteban, G.C. Work from home connection: a cluster analysis based on the Internet service attributes towards subscribers' profile. *Digital Transformation and Society*, **2023**, 2(1), pp. 60-77.
7. Moustafa, A.A., Bello, A., & Maurushat, A. The Role of User Behaviour in Improving Cyber Security Management. *Frontiers in Psychology*, **2021**, 12, 561011.
8. Aslan Ö, Aktuğ SS, Ozkan-Okay M, Yilmaz AA, Akin E. A Comprehensive Review of Cyber Security Vulnerabilities, Threats, Attacks, and Solutions. *Electronics*, **2023**, 12(6), 1333.
9. Werner, G., Yang, S., McConky, K. Time series forecasting of cyber-attack intensity, CISRC '17: Proceedings of the 12th Annual Conference on Cyber and Information Security Research. NY, USA, 04 April **2017**, pp. 1-3.
10. Zuzčaľk, M., & Bujok, P. Using Honeynet Data and a Time Series to Predict the Number of Cyber Attacks, *Computer Science and Information Systems*, **2020**, 18(4), 1197-1217.
11. Saha, S., Haque A., and Sidebottom, G. Deep Sequence Modeling for Anomalous ISP Traffic Prediction," ICC 2022 - *IEEE International Conference on Communications,* Seoul, Republic of Korea, **2022,** pp. 5439-5444.
12. Artioli P, Maci A, Magrì A. A comprehensive investigation of clustering algorithms for User and Entity Behavior Analytics. *Front Big Data*. **2024**, 9, 1375818.

13. El Maouaki, W., Innan, N., Marchisio, A., Said, T., Bennai, M., & Shafique, M. (2024). Quantum Clustering for Cybersecurity," 2024 IEEE International Conference on Quantum Computing and Engineering (QCE), Montreal, QC, Canada, **2024**, pp. 5-10.
14. Sufi, F., & Alsulami, M. Mathematical modeling and clustering framework for cyber threat analysis across industries. *Mathematics*, **2025**, *13*(4), 655.
15. Erilli, N. A. A Trimean and Asymmetry-Based Statistical Permutation Test for Group Comparisons. *International Journal of Innovative Technology and Interdisciplinary Sciences*, **2025**, 8(2), 324–335.