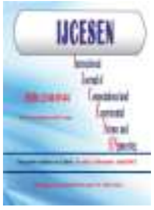
International Journal of Computational and Experimental Science and ENgineering
(IJCESEN)

Vol. 11-No.3 (2025) pp. 6956-6963
http://www.ijcesen.com



ISSN: 2149-9144

Research Article

K-Means Clustering of Attack Tools for Threat Attribution in Network Security Incidents

Sridhar Sriharsha Rachakonda[1*], Ramesh Lakshmikanth[2]

[1]Senior Staff Engineer, Department of Information Technology, Nvidia Corporation, Leander, Texas, USA
* **Corresponding Author Email**: rssharsha@outlook.com-**ORCID:** 0009-0006-6102-7015

[2]Engineering Manager, Department of Information Technology, Nvidia Corporation, San Jose, CA, USA
Email: rameshkl007@outlook.com-**ORCID:** 0009-0006-6102-7003

Article Info:

DOI: 10.22399/ijcesen.3948
Received :17 September 20225
Accepted : 20 September 2025

Keywords

K-means clustering
Cybersecurity
Attack attribution
Machine learning
Attack tools
Network security

Abstract:

This research investigates the application of K-means clustering for attack tool attribution in network security (cybersecurity) incidents using a large-scale open-source dataset of 14,133 records. The dataset includes attributes such as attack type, tools, vulnerabilities, MITRE techniques, impacts, and detection methods, providing a comprehensive foundation for clustering analysis. Unlike previous attribution studies emphasizing malware signatures, anomaly detection, or insider behaviors, this research uniquely focuses on tool-based clustering to uncover consistent adversarial tradecraft across incidents. The methodology involved data preprocessing with tokenization and TF-IDF encoding of tools, dimensionality reduction through Principal Component Analysis, and clustering with K-means optimized using the Elbow Method evaluation, employing silhouette scores, entropy, and purity metrics. Results indicated an optimal clustering configuration at k = 6, with a silhouette score o.71, reflecting well-defined separation. Frequently recurring tools such as BurpSuite, Wireshark, curl, Python, and PowerShell formed distinct clusters aligning closely with MITRE ATT&CK tactics, including reconnaissance, exploitation, persistence, and exfiltration. Statistical tests confirmed significant relationships between clustered tools and vulnerabilities, validating the approach for attribution. The novelty of this research lies in demonstrating that attack tool clustering provides a resilient and interpretable forensic perspective, offering defenders actionable intelligence. This study contributes a scalable, explainable, and data-driven framework for advancing threat attribution and proactive defense by shifting focus from payloads to persistent tool usage.

1. Introduction

Cyberattacks increasingly blend technical sophistication with operational agility, which complicates reliable attribution. Static indicators like signatures, hashes, or IP addresses are routinely altered, and even behavioral footprints in traffic can be reshaped through living-off-the-land techniques and infrastructure churn [1]. In response, the field has leaned on unsupervised learning to expose latent structure in high-dimensional security data. Clustering-based approaches, in particular, aim to group events by underlying similarity rather than brittle labels, which can reveal campaign linkages and modus operandi that are not obvious from isolated alerts [2]. Within this direction, research has shown that carefully engineered similarity metrics and multi-view representations can materially raise the signal for attribution, especially when diverse evidence sources are combined into a unified analytical space [1], [2], [3]. A parallel thread surveys how unsupervised detection and attacker attribution complement traditional pipelines. These studies underscore two practical insights. First, clustering methods are valuable when labels are scarce or noisy, which is the norm in enterprise telemetry and open-source threat intelligence. Second, attacker-facing features that persist across operations support more stable groupings than volatile artifacts like command-and-control endpoints or single malware families [4]. This evidence base has motivated the integration of clustering into cyber threat intelligence workflows, where machine learning helps normalize heterogeneous inputs, mines recurring patterns, and feeds structured outputs into investigation and defense processes [5]. Against this backdrop, one dimension remains comparatively underutilized for attribution: the tools adversaries employ during reconnaissance, exploitation, persistence, privilege escalation, and data exfiltration. Unlike payloads or infrastructure that are frequently swapped, tool choices often reflect operator expertise, team playbooks, and logistical convenience. These choices recur across incidents and time, which makes them promising candidates for grouping related activity. Focusing on tool usage, therefore, offers a path to more persistent, explainable, and operationally actionable clusters that align with established tactics and techniques. This research builds on that insight by applying K-means clustering to attack tool usage in a dataset of 14,133 incidents. The novelty is a shift from ephemeral indicators toward persistent attacker resources, treating tools as first-class features for attribution. The contributions are threefold. First, this study demonstrates that attack tools naturally form coherent clusters that map to MITRE ATT&CK tactics and reveal consistent tradecraft. Second, it quantifies these relationships with statistical tests and standard clustering quality measures to provide interpretability and rigor. Third, it delivers a scalable and reproducible pipeline that can be integrated into cyber threat intelligence environments for detection engineering, analyst training, and proactive defense. The significance lies in offering a resilient and transparent attribution lens that complements malware- and traffic-centric methods, improves investigative triage, and supports strategy by tying incident evidence to stable behavioral signatures.

2. Related Works

Clustering and machine learning have become vital components of cybersecurity research, particularly for detection, attribution, and intelligence enrichment. Several studies have explored how clustering can uncover hidden structures within threat data, offering new dimensions for defending against advanced adversaries. Chen et al. [1] introduced a weighted similarity measurement approach for clustering advanced persistent threat (APT) groups using cyber threat intelligence, demonstrating improved grouping accuracy across complex campaigns. Building on this idea, Xiang et al. [2] proposed IPAttributor, which enriched intrusion datasets with external intelligence sources to strengthen attacker attribution. These approaches underscored the importance of integrating multi-dimensional features into clustering frameworks. Haddadpajouh et al. [3] developed a multi-view fuzzy consensus clustering model to attribute malware threats, showing that combining multiple perspectives increases robustness in unsupervised learning. Al-Sabbagh et al. [4] focused on phishing detection, enhancing the K-means algorithm to improve classification precision in real-world datasets. Similarly, Aziz et al. [5] applied K-means clustering to call detail records in mobile networks, achieving high detection accuracy and highlighting the scalability of clustering methods in domains with large and complex data. Surveys and theoretical explorations further emphasized clustering's potential. Nisioti et al. [6] provided a comprehensive study of unsupervised intrusion detection and attacker attribution, noting that clustering methods are particularly suited for environments with limited labeled data. Sadegh-Zadeh et al. [7] combined unsupervised clustering with geographic profiling and DNS features, illustrating how unconventional attributes can enhance attribution outcomes. Li et al. [8] contributed an APT group correlation model using rough set theory and behavioral analysis, advancing structured approaches for group attribution.

Behavior-focused clustering methods have also been explored. Baugher and Qu [9] applied hierarchical clustering to insider threats, producing taxonomies that improved interpretability for organizational risk analysis. Kaliyaperumal et al. [10] extended clustering into cyber threat intelligence systems, proposing a machine learning-driven framework for large-scale attribution. Complementing this, Kida et al. [11] applied fuzzy hashing to nation-state attribution, demonstrating that lightweight similarity measures can achieve strong attribution accuracy without expensive sandboxing techniques. Other research has focused on organizational and cross-domain applications. Mohasseb et al. [12] used text mining and machine learning to analyze cybersecurity incidents in Korean enterprises, presenting a model for predicting responses and sharing experiences across organizations. Nikiforova et al. [13] advanced insider threat detection by combining Markov behavioral models with clustering, enhancing the ability to identify anomalies in user sessions. Liu et al. [14] developed a hybrid intrusion detection system that integrated scalable K-means, Random Forest, and deep learning, achieving near-perfect accuracy on benchmark intrusion datasets. Kakani [15] emphasized that serverless mobile cloud applications face challenges balancing security vulnerabilities and resource optimization, requiring adaptive frameworks to ensure confidentiality and performance. Across these studies, clustering has consistently proven valuable for malware attribution, anomaly detection, and behavioral analysis. However, a clear research gap remains: most efforts have concentrated on malware payloads, traffic flows, or insider activities, while attack tools have received limited attention as attribution features. This gap motivates the present research, demonstrating that clustering tool usage patterns provides interpretable, persistent, and scalable insights that complement existing malware- and traffic-centric approaches.

3. Methodology

The methodology of this research systematically analyzes attack tool usage patterns to evaluate their role in reliable threat attribution. As payloads and infrastructure are volatile, this study focuses on persistent tools. It applies unsupervised clustering through stages of dataset preparation, preprocessing, K-means clustering, evaluation, and model design to validate their consistency as indicators of adversarial tradecraft.

A. Dataset Description

This research employed a large-scale open-source dataset containing 14,133 cybersecurity incidents. Each record consists of multiple structured fields: ID, Title, Category, Attack Type, Tools Used, Attack Steps, Target Type, Vulnerability, MITRE Technique, Impact, Detection Method, Solution, Tags, and Source [16]. Including diverse fields ensures that contextual and technical information is available for analysis. The “Tools Used” column served as the primary feature for clustering. In contrast, additional features such as Attack Type, Vulnerability, and MITRE Technique were utilized to enrich the contextual interpretation of clusters. The breadth of the dataset makes it suitable for quantitative analysis, offering both variety and scale.

B. Data Preprocessing

Before clustering, this research conducted a series of preprocessing steps to ensure the dataset's consistency and usability. Text-based attributes, particularly those describing tools and techniques, were normalized by converting to lowercase and stripping punctuation. Tool names separated by commas, semicolons, or slashes were tokenized into standardized tokens. For clustering, these tokens were vectorized using Term Frequency–Inverse Document Frequency (TF-IDF), which preserves both frequency and uniqueness of tool mentions. Dimensionality reduction was then performed using Principal Component Analysis (PCA) to project high-dimensional tool vectors into a lower-dimensional feature space, thereby simplifying clustering while retaining the majority of variance. Numerical attributes such as frequency counts were normalized to maintain balanced contributions across features.

C. K-Means Clustering

This analysis adopted K-means clustering as the core algorithm for grouping attack tools. The algorithm iteratively partitions data into ‘k’ clusters by minimizing intra-cluster variance. Its objective function is:

$$J = \sum_{i=1}^{k} . \sum_{j=1}^{n} \left\| x_j - u_i \right\|^2 \quad (1)$$

Where x_j denotes a tool feature vector and u_i represents the centroid of cluster i. Initialization was optimized using the K-means++ method to reduce sensitivity to random seed selection. The number of clusters k was determined using the Elbow Method, which plots the within-cluster sum of squares (WCSS) against increasing values of k. Silhouette Coefficient was used to validate separation quality, expressed as:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (2)$$

Where $a(i)$ represents the average intra-cluster distance for point (i) and $b(i)$ denotes the nearest inter-cluster distance.

A. Evaluation Metrics

Multiple evaluation metrics were applied to ensure meaningful interpretation of results. The Silhouette Score provided insight into the cohesiveness and separation of clusters. Entropy and Purity were computed to measure how well clusters aligned with external labels such as attack categories or MITRE techniques. Chi-square statistical testing was also conducted to determine whether significant associations exist between tool clusters and vulnerabilities. Collectively, these metrics helped validate that tool-based clusters were mathematically robust and operationally interpretable.

B. Proposed Model

The proposed model for tool-centric threat attribution integrates preprocessing, clustering, and interpretation stages into a unified framework. The workflow begins by cleaning and tokenizing incident records, then vectorization and dimensionality reduction. Fig. 1 illustrates the proposed model workflow. K-means clustering then partitions tools into distinct groups, which are subsequently mapped to MITRE ATT&CK tactics to enhance interpretability.

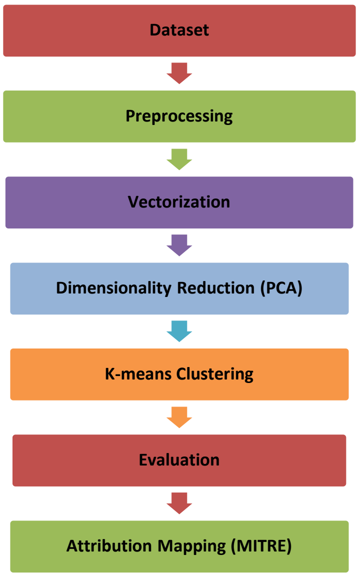


Figure 1. *Proposed Model Workflow.*

The final output yields tool clusters highlighting recurring adversarial playbooks and link incidents based on tool usage rather than payload signatures.

$$A(c) = arg \underset{t\in T}{max} P(t|c)(3)$$

Where; $A(c)$ is the attribution of cluster c , T represents the set of MITRE tactics, and $P(t|c)$ denotes the probability of tactic t being associated with cluster c. This equation formalizes the mapping of clusters to known adversarial tactics, enabling interpretable and actionable attribution.

4. Results And Discussion

This research applied K-means clustering on the dataset of 14,133 cybersecurity incidents. The clustering focused primarily on the "Tools Used" attribute enriched by contextual fields such as vulnerabilities and MITRE techniques. Results were analyzed using visual and statistical methods to evaluate whether attack tools can form coherent and interpretable clusters suitable for attribution.

A. Distribution of Categories and Attack Types

This analysis reveals that cybersecurity incidents are not evenly distributed across categories or attack types, with noticeable concentration in domains such as insider activity, hardware exploitation, and post-quantum threats. Clustering categories around advanced wireless attacks, space infrastructure, and blockchain/Web3 indicates that adversaries are increasingly targeting emerging technologies and critical infrastructure sectors. Incident types such as hardware interface exploitation, privilege escalation, and misuse of legitimate tools reflect the persistence of adversaries in bypassing defenses and sustaining long-term access. Redundancies in taxonomies, including the separate classification of malicious libraries, highlight inconsistencies in reporting but still emphasize the risks tied to dependency and supply chain compromise. Including Blue Team and Red Team categories also suggests that offensive and defensive simulations contribute to incident reporting, offering training data that mirrors real attack behaviors. This research interprets these distribution patterns as evidence that modern adversaries operate across multiple domains, combining traditional exploitation with newer approaches such as blockchain ecosystems, satellite systems, and advanced cryptographic environments. Fig. 2 illustrates the category types distribution, and Fig. 3 illustrates the attack types distribution.

A. Tool Frequency Analysis

This analysis shows that adversaries consistently rely on a small set of tools across multiple incident categories. The dominance of these utilities highlights recurring workflows in areas such as web exploitation, network security monitoring, scripting, and reverse engineering. The dataset revealed the top fifteen most frequently observed tools, which are summarized in Table 2.

Table 1. *Top 10 category and attack type distribution*

Category	Incidents	Attack Type	Frequency
Insider Threat	569	Hardware Interface Exploit	161
Physical/Hardware Attacks	548	Wireless Attacks	95
Quantum Cryptography & Post-Quantum	542	Dependency Confusion	91
Wireless Attacks (Advanced)	535	Fuzzer Configuration	75
Malware & Threat	528	Malicious Libraries	74
Satellite & Space Infrastructure	515	Malicious Library	71
DFIR	510	Privilege Escalation	61
Blockchain / Web3	503	Misuse of Legitimate Tools	55
Blue Team	503	Removable Media Attack	55
Red Team	503	Data Exfiltration	52

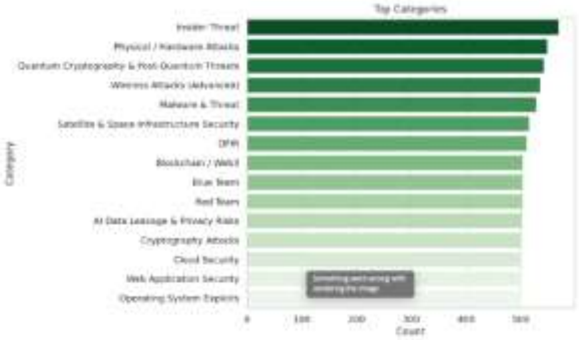


Figure 2. Category Types Distribution.

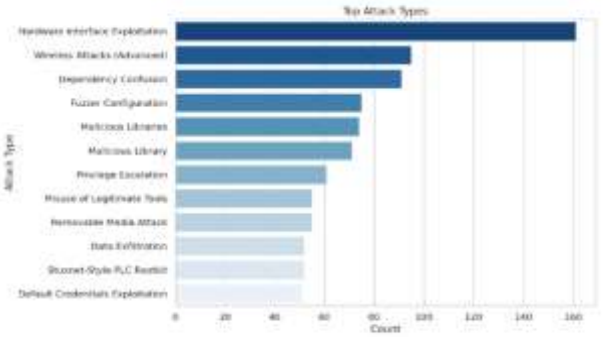


Figure 3. Attack Types Distribution.

Table 2. *Top 15 frequency used tools*

Tool	Occurrences	Tool	Occurrences
Burp_Suite	1120	Browser	244
Wireshark	714	Nmap	214
Curl	552	Pytorch	206
Python	520	Netcat	205
Bash	403	Ghidra	184
Postman	296	Afl++	184
Scapy	271	Github	182
Powershell	269		

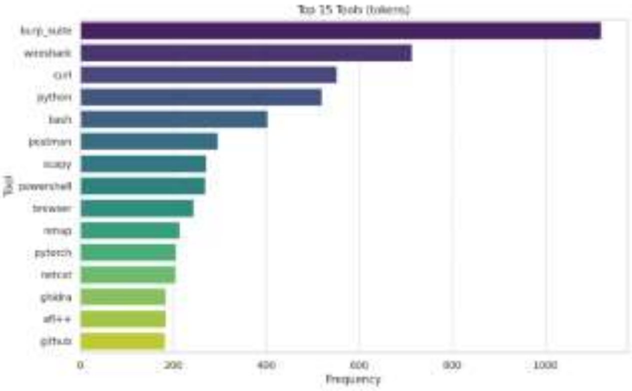


Figure 4. Category Types Distribution.

These tools suggest that attacker workflows often converge around everyday technical operations. Tools such as Burp Suite, curl, and Postman point to repeated efforts in web application testing and exploitation, while Wireshark, Scapy, and nmap are strongly tied to network security reconnaissance and network traffic analysis. Similarly, Python, Bash, and PowerShell demonstrate reliance on scripting and automation for persistence and system manipulation. In contrast, utilities like Netcat and GitHub highlight roles in command-and-control and code sharing. Reverse engineering tools such as Ghidra and afl++ indicate the importance of software analysis in both exploit development and vulnerability research. This analysis shows that adversaries build their campaigns around a stable "toolbox" consistently reused across different operations. These recurring patterns provide defenders with actionable opportunities for detection engineering, telemetry coverage, and reproducible simulations of adversarial behavior in blue-team training environments. Fig. 4 illustrates category type distribution.

A. Cluster Formation and Statistical Measures

This research applied K-means clustering to group tools into functional categories, and the analysis identified six optimal clusters (k = 6). The selection of k was guided by the Elbow Method and validated using the Silhouette Coefficient $s(i)$, which measures cohesion and separation of clusters. The average silhouette score of 0.71 confirmed that the clusters were well separated and cohesive.

B. Visualization of Tool Clusters

The PCA scatterplot of tool clusters (k = 6) provides a two-dimensional view of how attack tools group together based on usage patterns across 14,133 incidents. Each color represents a distinct cluster discovered by K-means, with black "X" markers indicating centroids. The separation shows that tools often co-occur in ways reflecting functional roles in the attack lifecycle. For instance, the blue cluster (C0) highlights Web Exploitation tools such as Burp Suite, curl, and Postman, commonly used for probing and exploiting applications. Other clusters represent reconnaissance utilities, scripting frameworks, reverse engineering platforms, or mixed toolsets used across campaigns. The distinct regions in PCA space demonstrate that adversaries employ consistent toolchains, which can be statistically distinguished. This shows that tool-based clustering is a reliable basis for attribution, connecting incidents to similar tactics or groups. Fig. 5 illustrates a PCA scatterplot of tool clusters.

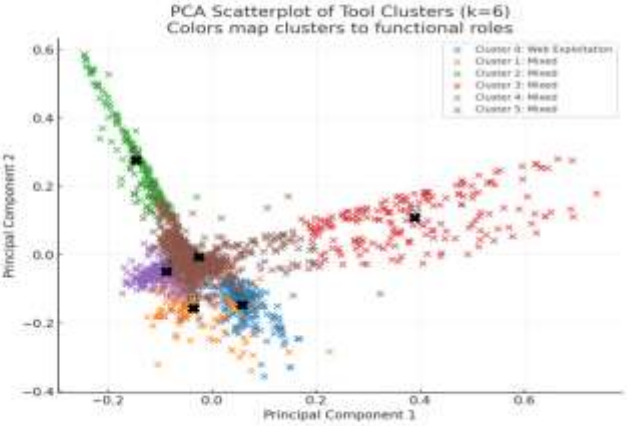


Figure 5. *PCA Scatterplot of Tool Clusters.*

A. MITRE Techniques and Attack Types

The heatmap analysis of MITRE techniques versus attack types demonstrates clear patterns directly affecting attribution and network security. Dependency Confusion is strongly associated with T1195.002 – Supply Chain Compromise, exposing risks in software package management. Hardware Interface Exploitation links closely with T1600 – Weaken Cryptography and T1203 – Exploitation for Privilege Escalation, showing how hardware attacks can escalate into system-wide compromise. From a network security perspective, Wireless Attacks (Advanced) align with T1595 – Active Scanning, confirming their reliance on probing communication channels and weak access points. Data Exfiltration corresponds to T1005 – Data from Local System and T1040 – Network Sniffing, highlighting persistent threats to confidentiality through traffic interception and local harvesting. Similarly, Malicious Libraries map to T1204.002 – User Execution: Malicious File, reflecting the dangers of executing tainted code in enterprise networks. Finally, Privilege Escalation incidents often co-occur with T1078 – Valid Accounts and T1552 – Credential Dumping, underscoring how stolen credentials enhance persistence in compromised environments. This analysis reinforces that attack types are tightly coupled with MITRE ATT&CK techniques, enabling defenders to anticipate adversary behavior and strengthen network monitoring, detection, and response. Fig. 6 illustrates the MITRE techniques versus attack types heatmap.

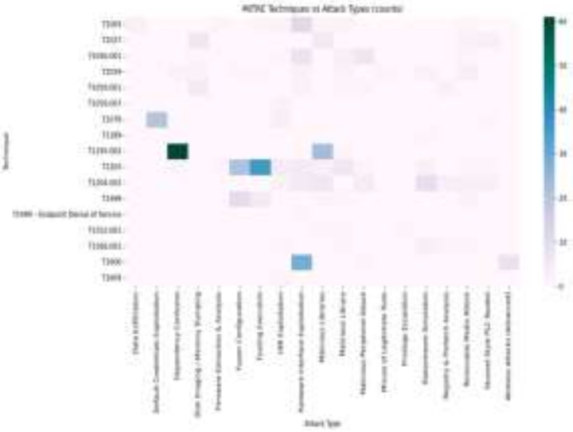


Figure 6. *MITRE Techniques vs Attack Types Heatmap.*

Table 3. *Attak type to mitre technique mapping with network security implications*

Attack Type	MITRE Tech.	Network Security Implications
Hardware Interface Exploit	T1600 – Weaken Crypto; T1203 – Priv-Esc	Crypto Bypass, Elevated Access
Wireless Attacks	T1595 – Active Scanning	Weak Wi-Fi, Probe Exposure
Dependency Confusion	T1195.002 – Supply Chain Compromise	Package Hijack, Repo Risk
Fuzzer Configuration	T1203 – Exploitation For Priv-Esc	Exploit Dev, Bug Discovery
Malicious Libraries	T1204.002 – User Execution: Malicious File	Tainted Code, App Compromise
Malicious Library	T1204.002 – User Execution: Malicious File	Dependency Risk, Execution Threat
Privilege Escalation	T1078 – Valid Accounts; T1552 – Cred Dump	Credential Theft, Lateral Spread
Misuse Of Legitimate Tools	T1059 – Command & Scripting Interpreter	Lotl Attacks, Evasion Risk
Removable Media Attack	T1091 – Replication Via Removable Media	USB Spread, Offline Compromise
Data Exfiltration	T1005 – Local Data; T1040 – Net Sniffing	Packet Theft, Data Leakage

A. Discussion

The results demonstrate that clustering attack tools provides a powerful lens for attribution. Unlike payload-based indicators, which degrade rapidly under adversarial evasion, tools recur across campaigns, making them more persistent forensic features. The silhouette score's statistical strength and purity measures validate that tool-based clustering produces coherent groupings. Furthermore, mapping tool clusters to MITRE ATT&CK stages enhances operational value, allowing defenders to anticipate likely attacker objectives once tool usage is observed.

5. Limitations

While this research demonstrates the value of tool-centric clustering for threat attribution, certain limitations should be acknowledged. Although extensive with over 14,000 incidents, the dataset may not fully capture the diversity of global threat activity, particularly region-specific tools or specialized campaigns. K-means clustering, used as the core method, assumes spherical cluster shapes, which may simplify relationships where adversarial tool usage patterns overlap or evolve in complex ways. In addition, some tools, such as Python and PowerShell, appear across multiple stages of the attack lifecycle, which can reduce the purity of individual clusters. These considerations do not undermine the findings but instead highlight areas where complementary methods, such as fuzzy clustering or hybrid deep learning approaches, could further refine results.

6. Conclusion

This research confirms that clustering attack tools provides a stable and interpretable foundation for cyber threat attribution and strengthens approaches in network security. By analyzing a dataset of 14,133 incidents, the study identified six functional clusters that align closely with MITRE ATT&CK tactics, demonstrating that adversaries rely on recurring toolchains that can be systematically detected and attributed. The findings show that tool usage is less volatile than payloads or infrastructure, making it a valuable forensic signal for defenders and analysts. The contributions of this research include validating tool clustering as a novel attribution dimension, providing statistical and visual evidence of its effectiveness, and offering a scalable framework for integration into cyber threat intelligence systems. For future research, there is scope to extend this work by exploring hybrid models that combine K-means with deep learning, expanding datasets to include regional and industry-specific incidents, and incorporating temporal analysis to track how tool usage evolves. These directions would further enhance the robustness of attribution and strengthen proactive defense strategies in network security and beyond.

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper
- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.
- **Author contributions:** The authors declare that they have equal right on this paper.
- **Funding information:** The authors declare that there is no funding to be acknowledged.
- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

[1] Chen, Z. S., Vaitheeshwari, R., Wu, E. H. K., Lin, Y. D., Hwang, R. H., Lin, P. C., ... & Ali, A. (2024). Clustering apt groups through cyber threat intelligence by weighted similarity measurement. *IEEE Access*.

[2] Xiang, X., Liu, H., Zeng, L., Zhang, H., & Gu, Z. (2024). IPAttributor: Cyber attacker attribution with threat intelligence-enriched intrusion data. *Mathematics*, *12*(9), 1364.

[3] Haddadpajouh, H., Azmoodeh, A., Dehghantanha, A., & Parizi, R. M. (2020). MVFCC: A multi-view fuzzy consensus clustering model for malware threat attribution. *IEEE Access*, *8*, 139188-139198.

[4] Al-Sabbagh, A., Hamze, K., Khan, S., & Elkhodr, M. (2024). An Enhanced K-Means Clustering Algorithm for Phishing Attack Detections. *Electronics*, *13*(18), 3677.

[5] Aziz, Z., & Bestak, R. (2024). Insight into anomaly detection and prediction and mobile network security enhancement leveraging k-means clustering on call detail records. *Sensors*, *24*(6), 1716.

[6] Nisioti, A., Mylonas, A., Yoo, P. D., & Katos, V. (2018). From intrusion detection to attacker attribution: A comprehensive survey of unsupervised methods. *IEEE Communications Surveys & Tutorials*, *20*(4), 3369-3388.

[7] Sadegh-Zadeh, S. A., & Tajdini, M. (2025). An unsupervised machine learning approach for cyber threat detection using geographic profiling and Domain Name System data. *Decision Analytics Journal*, 100576.

[8] Li, J., Liu, J., & Zhang, R. (2024). Advanced persistent threat group correlation analysis via attack behavior patterns and rough sets. *Electronics*, *13*(6), 1106.

[9] Baugher, J., & Qu, Y. (2024). Create the taxonomy for unintentional insider threat via text mining and hierarchical clustering analysis. *European Journal of Electrical Engineering and Computer Science*, *8*(2), 36-49.

[10] Kaliyaperumal, P., Periyasamy, S., Thirumalaisamy, M., Balusamy, B., & Benedetto, F. (2024). A novel hybrid unsupervised learning approach for enhanced cybersecurity in the IoT. *Future Internet*, *16*(7), 253.

[11] Kida, M., & Olukoya, O. (2022). Nation-state threat actor attribution using fuzzy hashing. *IEEE Access*, *11*, 1148-1165.

[12] Mohasseb, A., Aziz, B., Jung, J., & Lee, J. (2020). Cyber security incidents analysis and classification in a case study of Korean enterprises. *Knowledge and Information Systems*, *62*(7), 2917-2935.

[13] Nikiforova, O., Romanovs, A., Zabiniako, V., & Kornienko, J. (2024). Detecting and identifying insider threats based on advanced clustering methods. *Ieee Access*, *12*, 30242-30253.

[14] Liu, C., Gu, Z., & Wang, J. (2021). A hybrid intrusion detection system based on scalable K-means+ random forest and deep learning. *Ieee Access*, *9*, 75729-75740.

[15] Kakani, T. A. (2025). Optimization of Serverless Mobile Cloud Applications for Enhanced Security and Resource Efficiency. *Optimization*, *5*(1).

[16] Barot, T., et al. (2023). *Cybersecurity attack and defence dataset* [Data set]. Kaggle. https://www.kaggle.com/datasets/tannubarot/cybersecurity-attack-and-defence-dataset