mathematics

MDPI

Article

IPAttributor: Cyber Attacker Attribution with Threat Intelligence-Enriched Intrusion Data

Xiayu Xiang [1,†], Hao Liu [2,†], Liyi Zeng [1], Huan Zhang [1,2] and Zhaoquan Gu [1,2,*]

[1] Department of New Networks, Peng Cheng Laboratory, Shenzhen 518000, China; xiangxy@pcl.ac.cn (X.X.); zengly@pcl.ac.cn (L.Z.); zhangh07@pcl.ac.cn (H.Z.)

[2] School of Computer Science, Harbin Institute of Technology (Shenzhen), Shenzhen 518000, China; 22s051004@stu.hit.edu.cn

* Correspondence: guzhaoquan@hit.edu.cn

† These authors contributed equally to this work.

Abstract: In the dynamic landscape of cyberspace, organizations face a myriad of coordinated advanced threats that challenge the traditional defense paradigm. Cyber Threat Intelligence (CTI) plays a crucial role, providing in-depth insights into adversary groups and enhancing the detection and neutralization of complex cyber attacks. However, attributing attacks poses significant challenges due to over-reliance on malware samples or network detection data alone, which falls short of comprehensively profiling attackers. This paper proposes an IPv4-based threat attribution model, IPAttributor, that improves attack characterization by merging a real-world network behavior dataset comprising 39,707 intrusion entries with commercial threat intelligence from three distinct sources, offering a more nuanced context. A total of 30 features were utilized from the enriched dataset for each IP to create a feature matrix to assess the similarities and linkage of associated IPs, and a dynamic weighted threat segmentation algorithm was employed to discern attacker communities. The experiments affirm the efficacy of our method in pinpointing attackers sharing a common origin, achieving the highest accuracy of 88.89%. Our study advances the relatively underexplored line of work of cyber attacker attribution, with a specific interest in IP-based attribution strategies, thereby enhancing the overall understanding of the attacker's group regarding their capabilities and intentions.

Keywords: cyber threat intelligence; attacker attribution; APT; community discovery

MSC: 68M25

check for updates

Citation: Xiang, X.; Liu, H.; Zeng, L.; Zhang, H.; Gu, Z. IPAttributor: Cyber Attacker Attribution with Threat Intelligence-Enriched Intrusion Data. *Mathematics* **2024**, *12*, 1364. https://doi.org/10.3390/math12091364

Academic Editor: Antanas Cenys

Received: 1 April 2024
Revised: 27 April 2024
Accepted: 28 April 2024
Published: 30 April 2024

1. Introduction

More recently, the landscape of cyber threats has undergone significant transformation, with a notable shift toward more collaborative forms of cyber attacks [1]. A typical manifestation of this trend is the strategy employed by Advanced Persistent Threat (APT) groups, which often utilize an array of IP addresses to probe and compromise a single target. This approach not only demonstrates the coordinated and multi-faceted nature of modern cyber attacks but also serves to fulfill their strategic objectives by leveraging collective resources. Such an evolution represents a marked departure from earlier, more isolated methods, indicating a complex and interconnected threat environment. As a result, the challenges of cybersecurity have intensified, presenting formidable obstacles at a global scale. This macroscopic perspective underscores the urgent need to address the escalating complexities and collaborative nature of cyber threats in the contemporary security discourse.

Traditional cybersecurity defense measures have primarily focused on identifying known attack signatures or performing rule-based matching, a method that falls under reactive defense. While this strategy has proven effective against known threats, its effectiveness is significantly limited when facing the constantly evolving and increasingly complex landscape of unknown threats. Particularly in the context of APTs, which often operate in a highly covert and organized manner, relying solely on traditional defense mechanisms makes it challenging to comprehensively identify and effectively protect against these sophisticated new threats [2]. Therefore, in response to contemporary cybersecurity challenges, there is an urgent need to implement more advanced and proactive defense measures. It should not only detect and respond to known threats but also anticipate and mitigate emerging threats in a dynamic and collaborative cyber environment.

In this context, the significance of CTI has surged, establishing itself as a fundamental element of an advanced, proactive defense strategy [3–5]. CTI not only delivers extensive data on known threats but also, more crucially, provides profound insights into the behavioral patterns, attack strategies, and evolving trends of attackers. This enhanced capability goes beyond the confines of traditional defense mechanisms, offering security professionals the predictive tools needed to proactively detect and counteract complex and cooperated threats. Through comprehensive analysis and real-time monitoring, CTI transforms security decision-making from being merely reactive—based on historical or current events—to a proactive stance that anticipates potential future risks. This proactive shift enables organizations to mitigate threats more effectively by refining security policies, enhancing defense structures, and bolstering protections for sensitive assets before potential threats can take effect.

However, existing research has not fully capitalized on threat intelligence in the proactive detection of coordinated cyber attackers. Typically, the attribution process for a cyber attack commences with a technical analysis of the data produced during the incident [6], often focusing on malware sample investigation [7–13]. Malware utilized in APT campaigns facilitates remote manipulation and data theft from infected devices, highlighting its crucial role in delineating APT group profiles. Li et al. [7] introduces a machine learning-based methodology for classifying APT groups by analyzing malware characteristics, employing behavioral data tagged with APT organizational identifiers, derived from the dynamic analysis of APT malware collected from Internet of Things (IoT) devices. Beyond malware analysis, network behavior data is supplemental for cyber attack attribution [14–17], aiding the identification of distinct patterns in the attackers' Tactics, Techniques, and Procedures (TTP) from log data. Bai et al. [15] presents a novel strategy centered on network behavior for cyber attack grouping, aiming to refine attack classification accuracy by analyzing attackers' unique behavioral patterns. A critical limitation of these studies is their reliance on samples and logs, which often lack a more comprehensive context required for accurate attacker attribution. This gap hinders the understanding of the attackers' operational patterns and goals. Addressing this issue necessitates the integration of diverse threat intelligence sources, enriching the contextual landscape for a more effective evaluation of cyber threats.

This paper presents IPAttributor, an IPv4-based cyber attacker attribution model that refines attack characterization by integrating a real event network behavior dataset containing 39,707 alarm entries with commercial threat intelligence from three distinct sources. This amalgamation provides a proactive and comprehensive context for cyber attacker analysis. A detailed set of 30 features is extracted from this enriched dataset for each IP address, aiding a feature matrix construction. This matrix enables the detailed pairwise evaluation of similarities and relationships among associated IP groups. To further enhance the analysis, a dynamic weighted clustering algorithm is employed to delineate communities of attackers within this dataset. These communities are indicative of attacker groups sharing similar behavior patterns or origins, thereby improving the analytical insight obtained during this event investigation. The efficacy of IPAttributor in pinpointing attackers with shared origins has been validated through experiments, underscoring its utility as an effective tool in cybersecurity endeavors. The synergy of in-depth network behavior analysis and integrated threat intelligence offers a solid framework for attributing cyber attackers and unraveling the intricate dynamics of attacker communities.

The contributions of this paper are elaborated as follows:

- Threat Intelligence Enrichment: The paper enhances the analytical dataset by integrating and evaluating intelligence-based features, which are extracted from commercial threat intelligence, and behavior-based features, which are extracted from network intrusion data. This enriched dataset provides a comprehensive foundation for nuanced attack characterization and analysis.
- Cyber Attacker Attribution Model: A robust model for cyber attacker attribution is proposed, specifically focusing on APT groups. Our model capitalizes on the detailed features identified within the enriched dataset, employing pairwise similarity analyses and clustering techniques to discern and delineate attacker communities. This approach facilitates the efficient and precise pinpointing of the origins and associations of these groups.
- Empirical Validation with Real-world Data: Extensive experiments are conducted using a real-world dataset that includes a substantial number of alarm entries, combined with paid commercial threat intelligence. These experiments validate the effectiveness of the proposed approach, demonstrating its ability to accurately attribute cyber attacks and uncover the dynamics of attacker groups in a practical setting.

The rest of this paper is organized as follows. Section 2 reviews the related work and its limitations. Section 3 introduces the system and its core components. A comprehensive evaluation is presented in Section 4. Finally, Section 5 concludes this paper.

2. Related Work

In related works, we categorize cyber attacker attribution into three main types: malware-based, behavior-based, and intelligence-based. This tripartite classification forms the basis for our systematic review of existing research in the field. Our investigation reveals the strengths and weaknesses of each approach, highlighting the gaps and challenges that persist in current attribution methodologies.

2.1. Malware-Based Methods

Malware serves as a pivotal strategic tool within APT attacks. Consequently, malware characteristics have emerged as crucial identifiers for APT organizations, playing a significant role in attacker attribution. Malware attribution can be regarded as a classification problem; thus, machine learning-based methods have been widely applied in this field. Li et al. [7] introduced an integrated multi-classification model for attributing APT malware, wherein highly discriminative features were chosen based on the chi-square value of the high-dimensional feature vector. Wang et al. [8] utilized API calls to filter functions and generate paths. Subsequently, they employed time series models to extract critical local path features. After vectorizing the features, the malware was attributed to different attack organizations using a classification model. Nataraj et al. [9] introduced a method for classifying malware based on standard image features. This approach is grounded in the observation that images from the same malware family exhibit significant similarity in both layout and texture. In the analysis of malware homology, it is a common strategy for attackers to employ various malware variants as a means to circumvent detection and obscure classification. To tackle this challenge, Sahoo et al. [10] developed multi-view descriptions of malware by extracting and mixing opcodes, bytecodes, and header features. These descriptions serve to counteract obfuscation techniques effectively. Additionally, the researchers implemented four distinct machine learning classifiers to facilitate organization attribution. Addressing the imbalanced distribution of malware in practical contexts, Li et al. [11] developed a malware classification framework that employed multimodal fusion with an adaptive weighting scheme. Their methodology commenced with the augmentation of the malware feature space through the inclusion of diverse descriptors, encompassing byte-level, structural, and semantic attributes. The self-learning weights adaptively adjust during the training process, thereby enhancing the model's ability to accurately attribute new malware samples to the correct family, particularly in scenarios characterized by imbalanced datasets.

In addition to employing machine learning algorithms to train systems with malware features, graph techniques also serve as avenues for representing malware behavior and achieving attribution. Ding et al. [12] proposed the use of generic behavior graphs to represent the behavioral characteristics of malware families. They employed a graph-matching algorithm based on maximum weight subgraphs for detecting malware. Ki et al. [13] employed DNA sequence alignment algorithms to extract API call sequence patterns from diverse categories of malware. Their findings indicate that malware within the same family often share numerous common call subsequences.

2.2. Behavior-Based Methods

Cyber attackers often leave distinct characteristics and traces during their activities, which can be leveraged to trace and identify the perpetrators or the origins of the attacks. However, due to the sensitive nature of openly sourced network behavioral datasets, the academic community faces a shortage of authoritative data and research in this area. Wang et al. [14] introduced a high-speed network traffic analysis detection technology, which entails capturing and parsing network traffic packets in high-speed network environments. This method enables the efficient detection of WebShells and tracing of attackers through feature code matching. Bai et al. [15] extracted intrinsic attribute features and behavioral characteristics of attackers from alerts to derive profound behavioral patterns. They devised a feature matrix similarity calculation method based on this analysis and subsequently attributed alert information to specific attack organizations using community discovery algorithms. In the realm of industrial control systems, Wang et al. [16] employed statistical features and function code sequence features of packets from industrial control honeypots to quantify attack behaviors. They utilized multiple clustering methods to model feature representations and analyzed homologous attacks, which are created by the same attackers. Zhang et al. [17] developed a network attack-defense game model, based on game theory principles, to analyze attacker behavior and attribute network attacks. This model quantifies the payoffs of attack and defense to enhance the accuracy of attacker attribution.

2.3. Intelligence-Based Methods

CTI provides detailed information on cyber threats, including attack methodologies and the tools used by adversaries. By analyzing CTI, security professionals can attribute cyber attacks to specific organizations by matching observed activities with known profiles and behaviors of threat actors.

Ren et al. [2] proposed a cyber threat platform that supports threat knowledge representation, extraction, and practice, and includes the intrusion analysis diamond model to profile attack organizations. The diamond model provides a structured method for analyzing and understanding the relationships between adversaries, their capabilities, the infrastructure they use, and the victims they target. In this work, researchers expanded the “victim” dimension of the diamond model and refined the classification of “capability” to sketch different attack events launched by APT attack groups. Noor et al. [18] extracted attack patterns from CTI reports and profiled network threat actors. Subsequently, they conducted training and testing of multiple machine learning classifiers utilizing these profiles. Wang et al. [19] developed a threat intelligence knowledge map to extract tactical and technical intelligence from data sources such as malware, IP addresses, and domains. Subsequently, they classified nodes, representing threat intelligence reports, into categories corresponding to advanced persistent threat organizations using a graph attention neural network, a type of machine learning model that focuses on important features in graph-structured data. Xiao et al. [20] introduced an attribution methodology for APT actors, employing multi-modal and multi-level feature fusion. The multi-modal features, encompassing attribute type, natural language text, and topological relationship characteristics, are extracted from heterogeneous attribute graphs formed from APT reports and their Indicators of Compromise (IoC). The multi-level features stem from multi-layer heterogeneous graph attention networks constructed based on APT reports. Perry et al. [21] devised an innovative algorithm for representing threat intelligence text. They proposed the use of an enhanced binary bag-of-words model to generate vectors representing reports and then utilized XGBoost and binary classifiers to classify intelligence into known threat actors or unknown threat actors. Building upon Perry et al.'s text representation methodology, Naveen et al. [22] employed a neural network embedding model, Word2Vec, to embed known vocabulary and perform fuzzy matching embedding for unknown terms. Subsequently, they devised a multi-layer neural network model for predicting the classification of threat actors.

2.4. Discussion

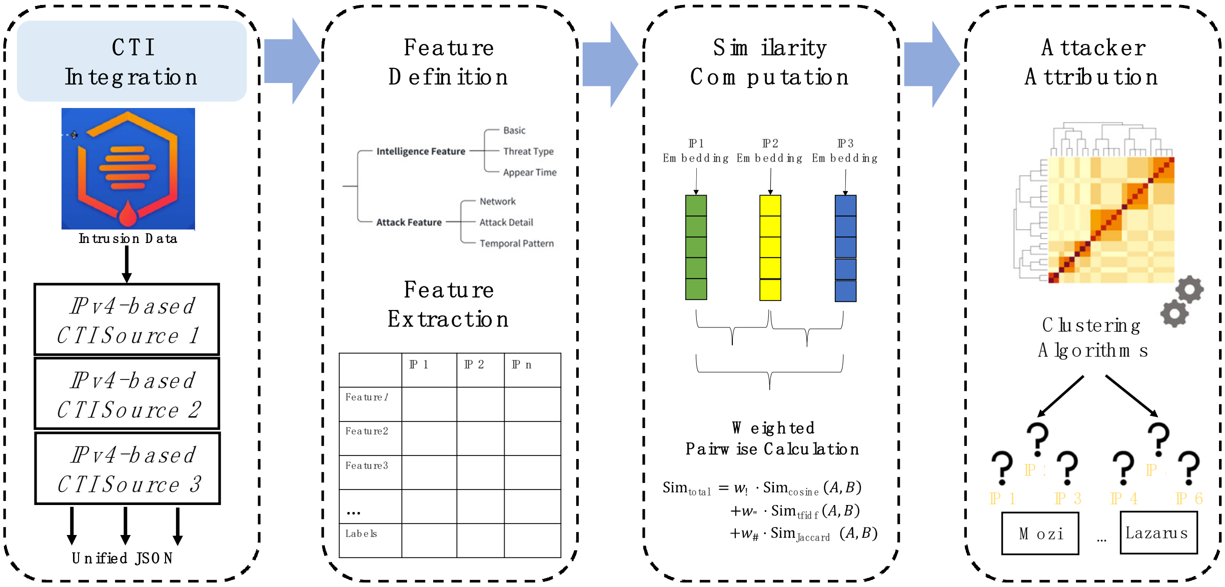
The domain of cybersecurity attribution plays a pivotal role in tracing the origins and perpetrators of cyber attacks. Malware-based methods attribute similar malware to specific malware families based on shared code structures, patterns, or features. Nonetheless, intricate relationships exist among malicious attacker families, as attackers might employ malware scripts sourced from the same family, rendering it challenging to pinpoint the actors of attack behavior. Behavior-based methods seek to delineate the behavioral profiles of attackers using data derived from single traffic or log sources. Nevertheless, these sources may contain deliberately obfuscated false leads, complicating the attribution process. Current intelligence-based methods primarily focus on extracting attacker attributes and attributing threats to attack organizations based on open-source threat intelligence reports. However, these methodologies often lack contextual grounding in real-world scenarios and fail to integrate both internal and external threat intelligence, thereby limiting their ability to provide a comprehensive understanding of attackers' origins and motivations.

3. Methods

Security logs serve as fundamental data sources for threat detection. However, an over-reliance on alarm entries alone falls short of supporting effective threat attribution. This limitation stems from the fact that logs primarily provide a snapshot of the prevailing situation without a comprehensive context. In light of this, our study not only incorporates real attack detection datasets but also enriches the analysis by integrating commercial threat intelligence from three distinct sources. This amalgamation yields more nuanced and detailed information, which is instrumental for conducting threat attribution analysis. To address the gaps and shortcomings in the current research, this paper introduces an advanced analysis framework for attacker origin, anchored in network behavior and threat intelligence datasets. Through a multi-dimensional analysis of attack behaviors, the framework elucidates the intricate patterns of attacker activities. Furthermore, it leverages sophisticated algorithms to facilitate the clustering of attackers, culminating in a robust attribution analysis methodology.

3.1. Overview

The system's framework contains four modules, shown in Figure 1. The first module focuses on the integration of threat intelligence, triggered by real network-based intrusion logs. The second module formally defines the feature types, and outputs a feature matrix. The third module computes pairwise IP similarity, utilizing feature embedding techniques. The fourth module employs a clustering algorithm to perform attribution analysis of the cyber attackers, which allows for the identification of the various attacking groups.



	IP 1	IP 2	IP n
Feature1			
Feature2			
Feature3			
...			
Labels			

Figure 1. IPAttributor System Framework.

3.2. CTI Integration

This module commences with a detailed examination of the complex operations conducted by attackers, as recorded in the intrusion log, which initially comprises 39,707 alarm entries. These entries, secured through appropriate permissions, lay the foundation for an in-depth analysis. To enhance analytical capabilities, an extensive preprocessing phase was undertaken to refine the dataset, aiming to improve its clarity and pertinence.

The preliminary phase of this process involved the identification and elimination of duplicate records, addressing the issue of redundant representation of certain intrusion attempts or activities. This elimination was accomplished by scrutinizing the log entries, focusing on identical attributes such as timestamps, source and destination IP addresses, and alarm signatures within a brief timeframe. Concurrently, extraneous or irrelevant data were methodically expunged, employing domain-specific heuristics to identify and discard non-essential log entries. Thereafter, the dataset was subjected to a normalization procedure to standardize the various log attributes, ensuring a consistent and analytically viable format. Following these rigorous preprocessing measures, the dataset was reduced to a more precise and manageable total of 24,879 entries.

Upon preprocessing the intrusion logs, 892 attackers' source IPs were extracted to initiate the CTI-acquiring process. For our research, we accessed three commercial CTI sources. Retrieving intelligence from these sources was facilitated through the use of Application Programming Interfaces (APIs). These APIs enabled an automated query and retrieval mechanism, allowing us to efficiently collect relevant IP threat data. The intelligence obtained from these sources was primarily in JSON (JavaScript Object Notation) format, which facilitated the integration and analysis of the data within our research framework. This step allowed for a structured and systematic analysis of the cyber threats associated with the identified IP addresses, enriching the intrusion data with threat intelligence.

Following the retrieval of CTI from multiple sources, an integration phase was conducted to amalgamate the data from three distinct vendors. This step was crucial in synthesizing a comprehensive threat profile for the implicated IP addresses. Each vendor contributes a unique perspective to the security threat associated with the IP address, offering diverse insights that enhance the depth of the threat analysis. To ensure the integrity and utility of the integrated dataset, we implemented a deduplication step. While repeated data points such as geographical location were identified and merged to eliminate redundancies, we preserved the distinct features provided by each vendor. This meticulous approach to integration allowed for the retention of the unique value offered by each source, thereby enriching the overall dataset with a multifaceted view of the threat landscape.

3.3. Feature Engineering

The second module of our study focused on feature engineering. In this phase, we first defined and categorized features within the enriched dataset to facilitate a layered classification approach. The classification was divided into two primary categories: intelligence features and attack features.

Intelligence features encompassed attributes directly related to the attacker's profile and the nature of the threat, including the attacker's IP address, geographical location, and the type of threat posed. These features were instrumental in constructing a comprehensive profile of the threat actors and their operational domains. Conversely, attack features were defined to capture the technical specifications of the cyber attack incidents, encompassing the attack payload, descriptive narratives of the incidents, and the network characteristics involved in the attack. These features provided a granular view of the attack methodologies and the technical vectors employed. Key features identified in our study are systematically presented in Table 1, illustrating the dual categorization and providing a foundational framework for subsequent analytical processes. This structured approach to feature engineering was designed to enhance the dataset's utility for deeper analytical insights.

Table 1. Feature definition and classification.

Feature Category	Feature Type	Feature
intelligence feature	basic	location carrier
	threat type	malicious label ttps http request
	appear time	occurrences
attack feature	network	ips ports
	execution	payload behavior
	temporal pattern	time frequency

In intelligence features, Basic Intelligence encompasses foundational data about the threat source, including location details and carrier information. Location data pinpoints the geographical origin of the IP address, while carrier information identifies the network provider hosting the IP. This category forms the bedrock of threat intelligence, offering essential context about the attacker's infrastructure. Threat-type intelligence reflects the historical threat activities associated with the IP, including the types of attacks and the associated malicious families. This category extends to the specific tactics used, such as the payload delivered, user agent strings, and HTTP request details (body and parameters). It serves as a crucial element for attacker identification and understanding the modus operandi, enriching the profile of past IP-related threats. Appear Time intelligence is segmented into 12 intervals over the past year, providing a temporal analysis of threat occurrences. This helps in understanding the frequency and timing of past activities, offering a time-based perspective that enriches the overall threat intelligence. Such temporal insights are valuable for identifying patterns and predicting future threat behaviors.

In attack features, Network Attributes are key characteristics of network communications, defining the source and destination of data flow and the services involved. These include the Source IP, identifying where the traffic originates, and the Destination IP, pinpointing the target. Additionally, the Destination Port specifies the service or application targeted at the recipient, while the Source Port helps trace the origin of the activity. Together, these components are crucial for mapping network pathways and analyzing traffic patterns. Attack Execution Details refer to the specific aspects of how a cyber attack is carried out, capturing the method and nature of the attack. Key components include the Payload, which is the actual malicious data or code transmitted during the attack, revealing the attack's intent and mechanism. Behavior describes the qualitative aspect of the attack, offering insights into how the attack manifests in network or system logs. These details are crucial for understanding the technical execution of cyber threats. Temporal Patterns in the cyber security context refer to the timing and frequency of network attacks, highlighting when and how often certain events occur. It includes the Activity Pattern, which tracks the occurrence of incidents over specific time intervals, represented in hourly segments across 24 h. These patterns provide insights into the regularity and timing of attacks, identifying peak activity times and potential automated behavior.

In the final step of feature engineering, we constructed a feature matrix to systematically analyze attacker behaviors. The matrix columns represent each distinct IP address identified as a potential source of threat, while rows indicate the individual features derived from the previous phase of CTI integration. The matrix facilitated a comprehensive and efficient portrayal of attacker behaviors, enabling a detailed examination of the characteristics and patterns associated with each IP address. The feature matrix serves as a pivotal tool for analyzing the multidimensional aspects of cyber threats. It allows for the aggregation of disparate data points into a coherent framework, where the interactions and correlations between different features can be analyzed. By transforming qualitative intelligence and attack features into a quantifiable format, the matrix enhances the analytical capabilities, enabling the identification of trends and patterns in cyber threat activities.

3.4. Similarity Computation

In the context of similarity analysis, our first step is to transform raw data features into structured, interpretable formats suitable for computation. This transformation, known as embedding, involves converting various types of data into numerical or vector representations that encapsulate the inherent characteristics of the data. Subsequently, we employ a variety of algorithms tailored to the nature of these embedded features to calculate their similarities. This approach ensures a nuanced analysis that accurately reflects the underlying patterns and relationships within the data.

To encode the semantic content of features within our dataset, we employ the Word2Vec [23] embedding technique. This method involves training a Word2Vec model on a corpus of malicious payloads [24] to acquire a high-dimensional vector representation for each word or token. This representation is crucial as it captures the contextual relationships and semantic similarities among different data fields, facilitating a nuanced analysis of the malicious content. By projecting data into a 100-dimensional continuous vector space, we enhance our capacity to quantify and analyze the textual content of cyber threats. This improvement is pivotal in augmenting the detection and classification of complex attack vectors. For less complex features, we employ straightforward encoding techniques such as one-hot encoding or label encoding, which are sufficient for capturing the necessary categorical distinctions without the need for semantic depth.

Upon completing the feature embedding process, we proceed to compute the similarity between IP addresses, employing diverse methods tailored to the types of data features.

For time features and payload: We utilize cosine similarity to assess the similarity between time features and payloads. This method calculates the cosine of the angle between two vectors in an n-dimensional space, effectively capturing the semantic relationships between time sequences and payload data. The cosine similarity is calculated as follows:

$$\mathrm{Sim}_{\mathrm{cosine}}\,(A,B) = \frac{A \cdot B}{\parallel A \parallel \parallel B \parallel} \tag{1}$$

where $A \cdot B$ is the dot product of vectors A and B, and $\parallel A \parallel$ and $\parallel B \parallel$ are the magnitudes (or Euclidean norms) of the vectors.

For alert information: We employ the TF–IDF (Term Frequency–Inverse Document Frequency) [25] method to evaluate the similarity of textual content in alert messages. This technique quantifies the importance and relevance of words by analyzing their frequency in a document against their distribution across the entire dataset, providing insight into text-based similarities. TF–IDF similarity Sim_{tfidf} is structurally similar to cosine similarity.

For threat intelligence labels: The Jaccard coefficient is used to measure the similarity between sets of threat intelligence labels. By calculating the proportion of the intersection to the union of two label sets, the Jaccard coefficient is particularly suited for analyzing the similarity in categorical data and label sets. The Jaccard coefficient is calculated as follows:

$$\mathrm{Sim}_{\mathrm{Jaccard}}\ (A, B) = \frac{|A \cap B|}{|A \cup B|} \tag{2}$$

where $|A \cap B|$ is the number of elements in the intersection of sets A and B, and $|A \cup B|$ is the number of elements in the union of A and B.

To compute the sum of these features, the similarity scores are aggregated into a single metric, a weighted sum based on the relevance of each feature, as shown in Equation (3):

$$Sim = w_1 \cdot \mathrm{Sim}_{\mathrm{cosine}}\ (A, B) + w_2 \cdot \mathrm{Sim}_{\mathrm{tfidf}}\ (A, B) + w_3 \cdot \mathrm{Sim}_{\mathrm{Jaccard}}\ (A, B) \tag{3}$$

where w_1, w_2, and w_3 are weights that reflect the relative importance of various features. The choice of weights depends on the specific context and importance of each similarity measure adaptive to the overall analysis.

In Equation (3), we only compute the weighted sum for an individual data source to assess the similarity score between two IP addresses. However, to accommodate expansions in data sources, we need to extend this approach to integrate results from multiple sources, such as the alarm info and threat intelligence from three sources in our case. By doing so, we can maintain the analytical capabilities of the existing data sources while seamlessly integrating additional sources in the future, thus forming a comprehensive assessment of similarity. Specifically, we will calculate the weighted sum for each data source and then aggregate these sums into a single similarity metric. The similarity S_{ij} is calculated as:

$$\mathrm{Sim}_{\mathrm{xy}}\ = \textstyle\sum_{i=1}^{n} w_i \cdot Sim_i(x, y) \tag{4}$$

Formally, $Sim_{\mathrm{i}}\ (x, y)$ represents the similarity measure between IP address x and y in the ith data source, where n is the total number of data sources; w_i is the corresponding weight assigned to each data source, indicating its importance.

3.5. Attacker Attribution

To enhance the detection of potential attack organizations within the entire dataset, we will employ a clustering algorithm. This approach is aimed at facilitating improved community discovery, enabling us to identify and analyze clusters of data that may represent coordinated attack behaviors. By leveraging the inherent patterns and relationships within the data, the clustering algorithm can segregate the dataset into distinct groups based on their similarities. This method not only helps in pinpointing anomalous activities that signify potential threats but also aids in understanding the underlying structure of the data, leading to more effective monitoring and preemptive security measures against organized cyber threats.

The initial step before executing clustering is to construct a similarity matrix, which is based on the previously calculated pairwise similarities between IP addresses. This matrix serves as a foundational element that encapsulates the relationships and degree of resemblance between the various IP entities. By systematically arranging the similarity scores, the matrix provides a comprehensive view of how each IP is related to the others, thereby enabling a structured and informed approach to subsequent clustering processes. This construction of a similarity matrix is crucial as it directly influences the effectiveness and accuracy of the clustering algorithm in identifying and grouping related IP addresses.

Building on the foundational similarity matrix, which quantifies pairwise similarities between IP addresses, we employ the K-means clustering algorithm to categorize these IP addresses into distinct groups. The algorithm initiates by selecting 'K' initial centroids, which can be chosen either randomly or based on predefined heuristic criteria. It then proceeds iteratively: each IP address is assigned to the nearest centroid based on the calculated Euclidean distance within the similarity matrix. This step clusters IPs with high mutual similarity. After each assignment, the centroids are recalculated by determining the arithmetic mean of all IPs within each cluster. This iterative reassignment and recalculation continues until the centroid positions stabilize, indicating that the clusters have converged and no further significant shifts occur in their composition. This method not only segments IP addresses into coherent groups reflecting their similarity-based relationships but also utilizes the detailed insights from the similarity matrix to ensure that the clustering is both accurate and meaningful in delineating underlying patterns among the IPs.

We present the dynamic weighted threat segmentation algorithm, an adaptive method that fine-tunes its focus based on the varying significance of each feature for every data source. This algorithm goes beyond traditional clustering by employing a heuristic optimization technique to dynamically allocate weights to features, based on expert feedback, thereby sharpening the resolution of our similarity assessments. Through iterative refinement, the algorithm determines the optimal congregation of IP addresses into clusters that signify potential coordinated attack behavior. The complete algorithm is illustrated in Algorithm 1.

Algorithm 1 Dynamic weighted threat segmentation algorithm

```
Input:
Dataset $X = \{X_1, X_2, \ldots, X_n\}$; each $X_i$ is a multi-featured representation of an IP.
  Security threshold $\delta$; heuristic parameter for the optimization algorithm.
Number of clusters $K$.
Output:
Clustering result $U_1, U_2, \ldots, U_k$.
Optimal weights $W^* = \{W_1^*, W_2^*, \ldots, W_n^*\}$.
Procedure:
Initialize weights randomly within range $[0, 1]$, ensure sum to 1
for $(x, y)$, where $1 \leq x, y \leq n$ do
Sim$_{\mathrm{xy}}$ $= \sum_{i=1}^{n} w_i \cdot Sim_i(X_x, X_y)$
end for
Apply heuristic optimization $J(W; E) = f(W) + \lambda \cdot g(E, W)$
Recompute Sim$_{\mathrm{xy}}$ with optimized weights $W^*$ after convergence
Initialize $K$ centroids $\{\mu_1, \mu_2, \ldots, \mu_k\}$
repeat
    Assign $X_i$ to cluster $U_l$ if $min_l ||X_i - \mu_l||$
        Update $\mu_l = \frac{1}{U_l} \sum X_i \in X_i$
        if $\mu < threshold\ \varepsilon$
break
until clusters are stable
for each $U_l$ do:
        Optimize $\mu_l$ using the updated similarity matrix $S$
end for
return final clusters $U_1, U_2, \ldots, U_k$ and optimal weights $W^*$
```

4. Evaluation

In this section, we conduct cyber attacker attribution based on the enriched dataset, verify the performance of the IPAttributor, and proceed with a detailed case study.

4.1. Environment

We implement the proposed approach in Python 3.7. All experiments are run on two NVIDIA GeForce RTX 2080 GPU machines, and the video memory size is 16 GB. The CPU is an Intel(R) Core(TM) i7-9700K CPU, and the total memory of the machine is 32 GB.

4.2. Dataset

To more effectively validate the efficacy of the methodology proposed in this paper, we opted not to employ traditional open-source datasets. Instead, we utilized a real-world network intrusion log, enriched by data from three paid commercial intelligence sources. This approach allowed us to compile a more comprehensive dataset, which holds particular importance for research in the field of cybersecurity. It addresses the prevalent issue of the gap between theory and practice in cybersecurity datasets, offering a more relevant and empirically grounded basis for analysis.

The data acquisition process involved collecting syslog outputs from security devices over a period of one week. These logs were initially filtered to extract fields related to attacker IP addresses, which were then used to query threat intelligence services. The queries to these services were made via APIs, facilitating real-time data retrieval and integration. All the gathered data were stored in a MongoDB database, ensuring robust data management and easy retrieval for further analysis.

The initial dataset comprised 39,707 logs, which, after data preprocessing, were reduced to 24,879 logs. We further extracted 892 unique IPv4 addresses and used them as search criteria for querying various threat intelligence sources. This process yielded multidimensional threat intelligence data. We integrated all unique intelligence fields, excluding redundant threat information, into the corresponding logs, culminating in a dataset with 30 distinct features. Typical data fields are illustrated in Table 1.

4.3. Evaluation Metrics

To carry out a quantitative assessment of cyber attribution, the following metrics with their mathematical definitions are provided below.

Accuracy (Acc) is used to assess whether attacker IPs can be accurately classified into their respective organizations, which is achieved by identifying the proportion of similar IPs correctly assigned.

$$\text{Acc} = \frac{n_{correct}}{N} \tag{5}$$

Here, $n_{correct}$ denotes the number of IPs correctly classified into organizations, meaning that IPs identified as part of the same community in the experimental results are also affiliated with the same organization according to label knowledge. N represents the total number of IPs.

To assess the quality of the clustering solution, we will implement the silhouette coefficient, which quantifies how well each data point lies within its cluster relative to other clusters. This coefficient assists in determining the compactness and separation of the clusters formed. The silhouette coefficient $s(i)$ for each data point is expressed as follows:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{6}$$

For a set of clusters, the overall mean silhouette coefficient is calculated as follows:

$$S = \frac{1}{N}\sum\nolimits_{i=1}^{N} s(i) \tag{7}$$

where $a(i)$ is the mean distance of point i to the other points in the same cluster, $b(i)$ is the mean distance of point i to the points in the nearest cluster that i is not a part of, and N is the total number of points. A higher average S value close to 1 signifies well-fitted clusters, while a low or negative average S value would indicate overlapping or poorly separated clusters.

4.4. Similarity Computation Results

Setup: In our experiment, we calculated the pairwise similarity between IPs and determined a composite similarity score based on the weighted sum of these calculations. Specifically, we randomly selected four IPs as reference points and computed their similarity with all other IPs, identifying the nine IPs with the highest similarity scores for each reference IP. We then compared these results with existing label information to assess the accuracy of the similarity-based classification.

Evaluation: In Table 2, we calculated the pairwise similarities between IP addresses and measured the accuracy of classifying them into their associated attack organizations. The analysis was performed using four randomly chosen reference IPs, and the attribution accuracy was found to be 77.78%, 66.67%, 88.89%, and 77.78%. The average accuracy across these IPs was determined to be 77.78%.

Table 2. IP similarity accuracy result.

	IP 1 **111.XXX.XXX.136**	**IP 2** **122.XXX.XXX.163**	**IP 3** **111.XXX.XXX.148**	**IP 4** **101.XXX.XXX.111**
Acc	77.78%	66.67%	88.89%	77.78%
Average Acc	77.78%			

The findings from our experiment suggest that the methodology can be highly effective, as seen with the highest accuracy of 88.89%, and the overall average accuracy, which is indicative of a generally robust model. Similar observations were noted in Reference [15], where the authors achieved comparable accuracy results around 90%. However, our dataset is fundamentally different, and our research incorporates additional threat intelligence sources and features. Consequently, our work represents an improvement over previous efforts.

Upon analyzing the experimental outcomes, it is evident that leveraging an enriched dataset augmented with threat intelligence contributes significantly to the effectiveness of identifying similar IPs. The method's success, demonstrated by the high accuracy rate in certain instances, underscores the potential of applying a multifaceted data approach to enhance cybersecurity measures. This technique shows particular promise in distinguishing between benign and malicious network behavior by analyzing the nuanced similarities among IP addresses.

4.5. Attacker Attribution Results

4.5.1. Clustering Comparison

Setup: We conducted a detailed comparative analysis of K-means and spectral clustering algorithms. To facilitate a comprehensive understanding of the underlying patterns and distinctions between these clustering methodologies, we employed t-SNE (t-distributed stochastic neighbor embedding) for high-dimensional data visualization. This approach allowed us to project the multi-dimensional features into a two-dimensional space, enhancing the interpretability of the clustering results. The experiments were systematically structured to explore the impact of varying the number of cluster centers on the clustering outcome. Through this experimental design, we aimed to elucidate the behavioral dynamics and performance efficacy of K-means and spectral clustering across different cluster configurations, providing an in-depth visual and quantitative analysis of their clustering capabilities.

Evaluation: Upon reviewing the t-SNE visualizations provided for both spectral clustering and K-means, a comparative evaluation can be formulated. Figure 2 represents spectral clustering, and displays a discernible delineation of data clusters, albeit with overlapping regions where the cluster boundaries are less clearly defined. Conversely, Figure 3 represents K-means, and exhibits a more pronounced separation of clusters with distinct boundaries and centrally located centroids, indicative of higher intra-cluster homogeneity and inter-cluster separation. These observations suggest that in the context of our dataset and the chosen dimensional reduction technique, K-means delivers a more effective clustering solution with cleaner division among the data points. This visual assessment advocates for the superiority of K-means in yielding cohesive and well-partitioned clusters, as is evident from the t-SNE visualized data distribution. Thus, K-means is selected as the base clustering algorithm for our dynamic weighted clustering algorithm.

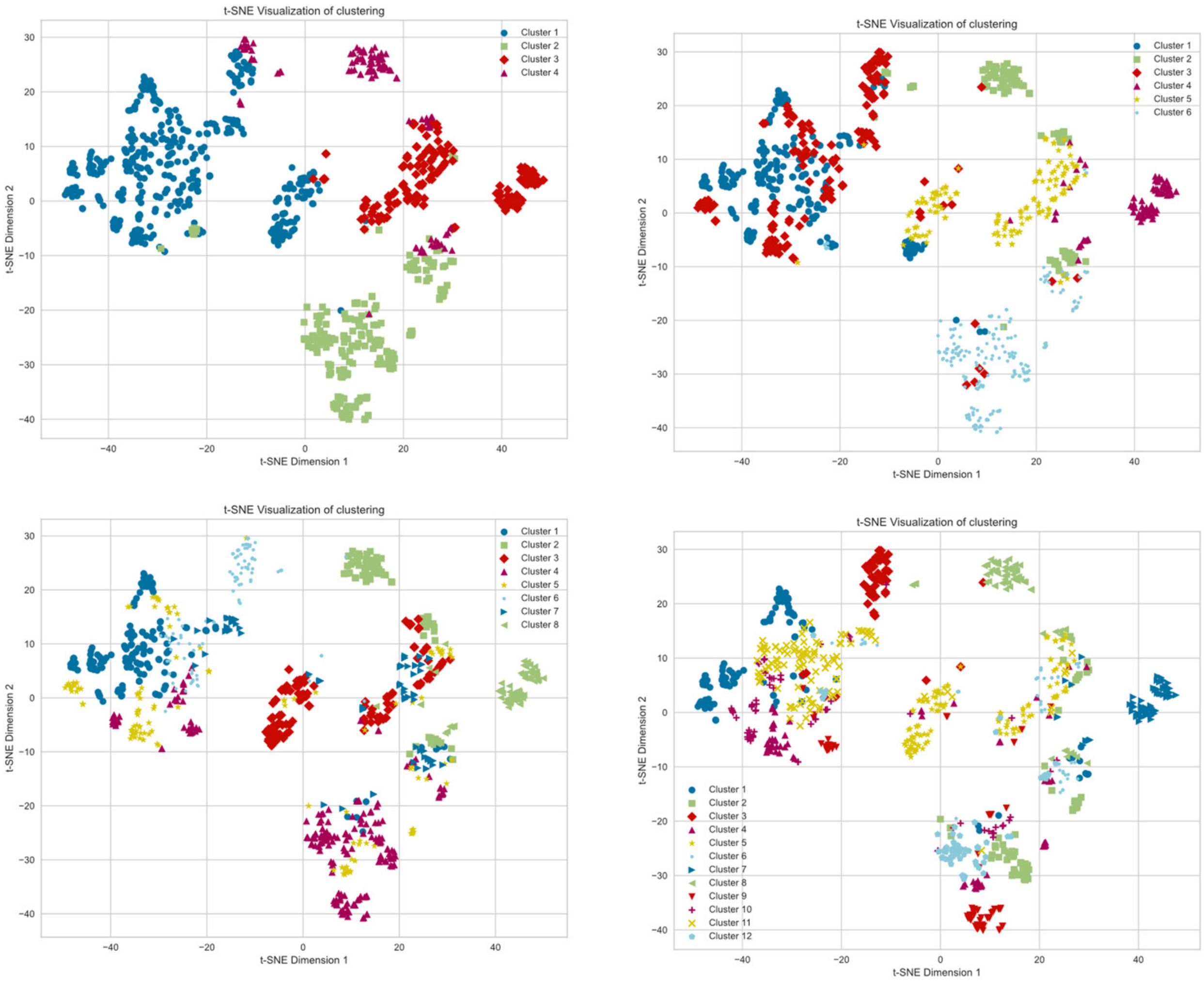


Figure 2. Spectral clustering t-SNE visualization.

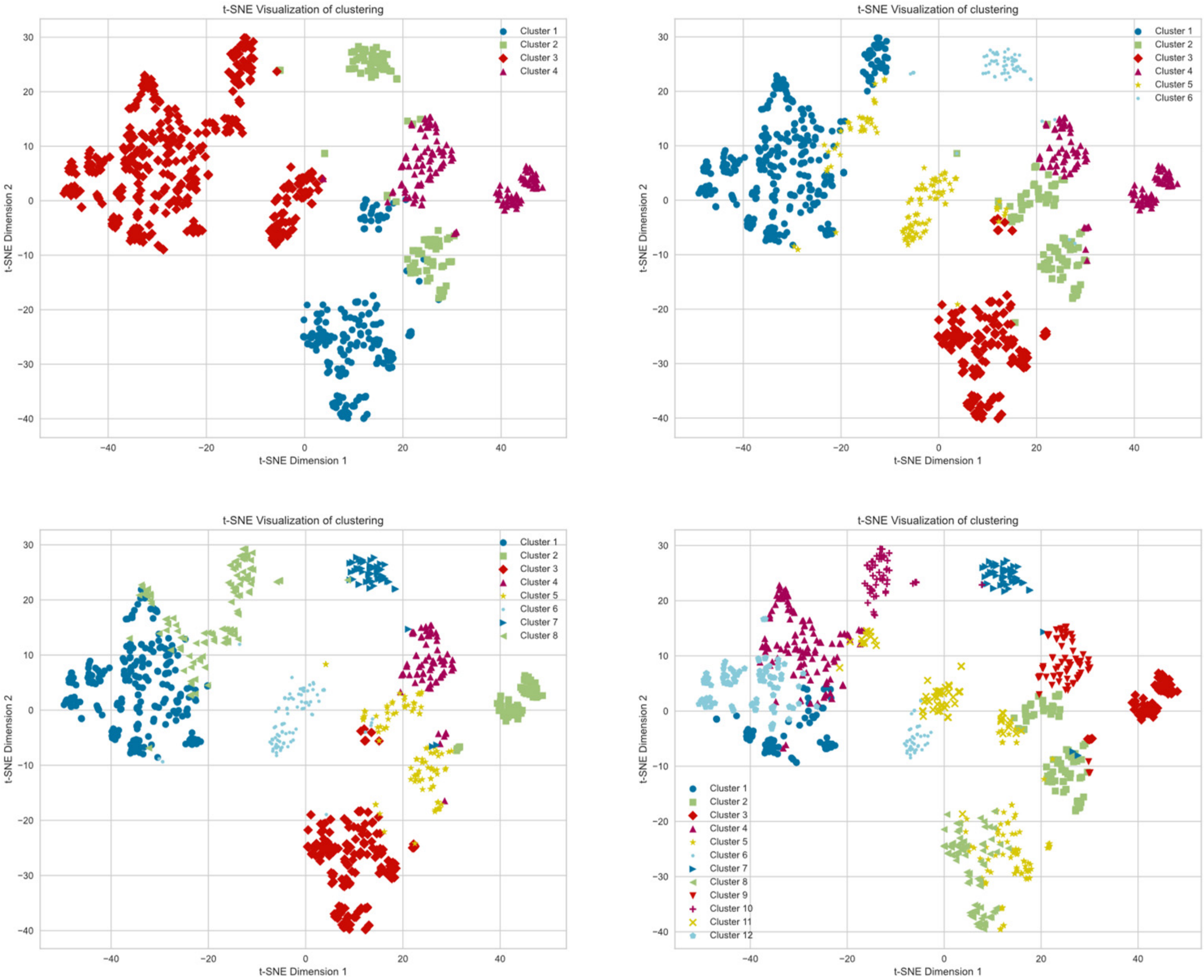


Figure 3. K-means t-SNE visualization.

4.5.2. Optimal Attribution Results

Setup: The subsequent phase of the experiment meticulously evaluated the silhouette systems generated through the dynamic weight selection algorithm. This involved an intricate analysis based on a spectrum of weights and multiple clustering centroids. This comprehensive examination can successfully corroborate the efficacy of the proposed method, and the validation was achieved by quantitatively assessing the silhouette scores, which provide insight into the cohesion and separation of the clusters formed.

Evaluation: According to Figure 4, the algorithm is capable of dynamically selecting the weights among different data sources, leading to the identification of various clustering centers and the calculation of multiple silhouette coefficients. This process results in an enhanced clustering outcome. Notably, the optimal silhouette coefficient achieved was 0.44, where the corresponding weights for alarm log, CTI source 1, CTI source 2, and CTI source 3 are 0.6, 0.1, 0.1, and 0.2. This coefficient indicates the degree of fit of the data points within their respective clusters, suggesting that the algorithm effectively discriminates between clusters, thus optimizing the homogeneity within clusters and the heterogeneity between them. And a total of four communities were found with our cyber attacker attribution model.

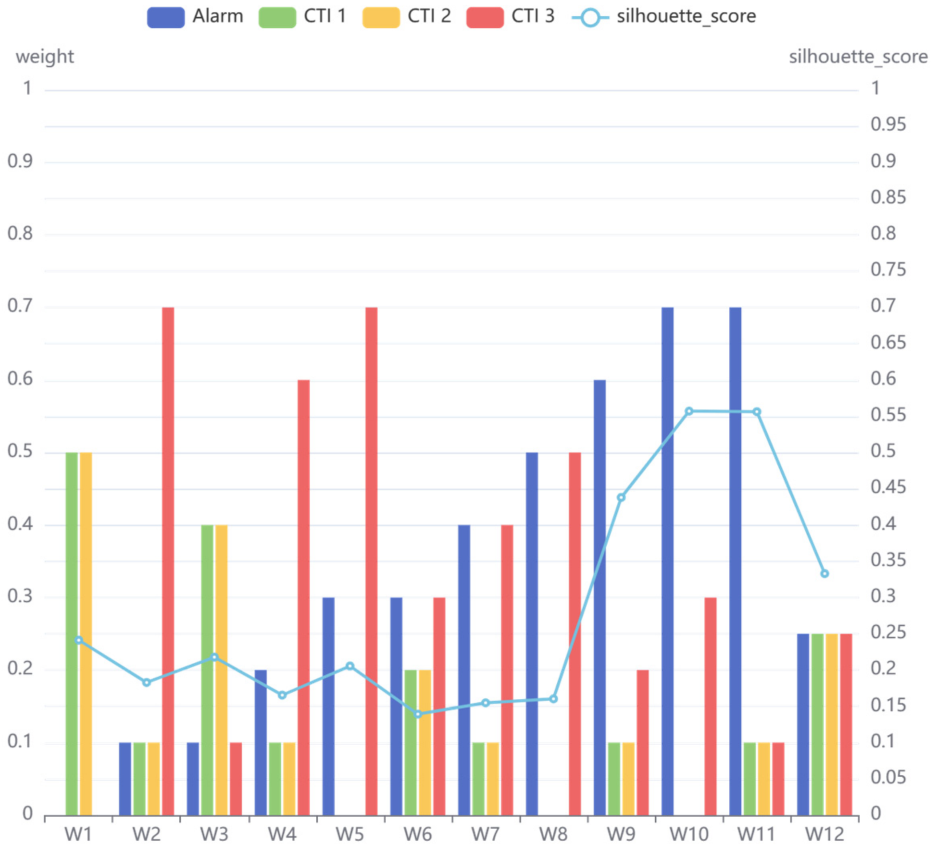


Figure 4. Dynamic weight selection and corresponding silhouette coefficient.

The effectiveness of the methodology presented in this section is substantiated by the empirical results, which illustrate that the dynamic weight selection algorithm possesses the capability to adaptively adjust to the diverse characteristics of the data. This adaptability is crucial for optimizing cluster formation, leading to a significant enhancement in the overall clustering performance. The algorithm's core feature, the dynamic weight selection mechanism, is instrumental in this improvement. It finely tunes the influence of each data source, aligning it with its relevance to the specific clustering task at hand. This tailored approach ensures that the clustering algorithm can effectively discern and group similar data points while distinguishing between dissimilar ones, thus maximizing the efficacy of the clustering process. Through these mechanisms, the proposed method demonstrates a robust and flexible approach to clustering, capable of handling varied data scenarios and yielding optimized results.

4.6. Similarity Computation Case Analysis

Following the completion of the accuracy calculation for IP similarity, this section of our experiment will employ a case study on a selected group of IP addresses. Through this detailed examination, we aim to further substantiate the effectiveness of our methodology.

In Table 3, we present IP address 111.XXX. XXX.148 as the reference IP, followed by a list of other IPs with high similarity scores. Key attack and intelligence features are provided for these IPs, including records of attack behavior, the corresponding Internet service providers, and malicious labels from a commercial intelligence source.

The importance of an enriched dataset is exemplified by analyzing the data presented in Table 3, which details the characteristics and threat types associated with each IP address. Initially, relying solely on an attack detection dataset may lead to incomplete attribution because such datasets typically do not offer comprehensive insights into the behaviors and associations of the attackers. For instance, although IPs such as 111. XXX. XXX.150 and 111.

XXX. XXX.169 exhibit 'Normal Access Behavior', without additional context, their potential threat cannot be accurately assessed.

Table 3. High-similarity IPs detail information.

Source IP	Attack Behavior	Carrier Information	Threat Types
111. XXX. XXX.148	'IP Pool Scan'	'China Mobile'	['SSH', 'Web App Attack']
111. XXX. XXX.151	'Normal Access Behavior'	'China Mobile'	['Web Spam', 'SSH', 'Web App Attack']
111. XXX. XXX.150	'Normal Access Behavior'	'China Mobile'	['Brute-Force', 'SSH', 'Web App Attack']
111. XXX. XXX.139	'IP Pool Scan'	'China Mobile'	['DDos Attack', 'Port Scan']
111. XXX. XXX.152	'IP Pool Scan'	'China Mobile'	['Hacking', 'Web App Attack', 'Port Scan']
123. XXX. XXX.17	'IP Pool Scan'	'China Telecom'	['SSH']
52. XXX. XXX.65	'Normal Access Behavior'	'Ningxia West Cloud Data'	['SSH', 'Port Scan', 'Web App Attack']
162. XXX. XXX.6	'Normal Access Behavior'	'DigitalOcean, LLC'	['SSH', 'Port Scan', 'Exploited Host']
111. XXX. XXX.160	'Normal Access Behavior'	'China Mobile'	['Web App Attack', 'SSH']
111. XXX. XXX.169	'Normal Access Behavior'	'China Mobile'	['Web App Attack', 'Port Scan']

Furthermore, threat intelligence with mere geographical location information, like the carrier data showing multiple IPs associated with 'China Mobile' or 'China Telecom', is insufficient for reliable homology typing. Geolocation can signal potential regions of interest but fails to uncover the complex layers of a cyber attack, including the specific threat actors and their operational tactics.

By merging multidimensional intelligence, such as attack behavior descriptors and specific threat types—ranging from 'IP Pool Scan' to 'DDoS Attack' and 'Web App Attack'—our method effectively pinpoints IPs related to the reference IP. For example, the reference IP 111. XXX. XXX.148, associated with an 'IP Pool Scan' attack and linked to 'China Mobile', is contextualized with other IPs sharing similar threat patterns and carrier information, enhancing the accuracy of cyber threat attribution. This unified approach not only reinforces the necessity of an enriched dataset but also highlights its efficacy in identifying and understanding complex cyber threats in a nuanced and actionable manner.

Building on the foundation of the three existing intelligence sources, we incorporated additional intelligence sources and performed further comparisons on the IPs exhibiting high similarity. Specifically, we conducted comparisons based on the subnet information provided by the additional intelligence sources, as well as the profiles of the attacking organizations. This allowed us to deepen our understanding and attribution of the cyber threats associated with these IPs.

Following the computation of IP similarity, we identified a set of IPs that were analogous to our reference IP. To validate the accuracy of our similarity calculations, we cross-referenced these similar IPs with an additional intelligence source, shown in Table 4. This subsequent step confirmed that our identified IPs were indeed listed within this new source's subnet information, thereby corroborating the effectiveness of our similarity assessment methodology. This cross-validation with an external dataset provides a robust endorsement of our approach, establishing the reliability of our findings in the context of cybersecurity threat analysis.

In Figure 5, we present a granular comparative analysis showcasing the detailed attributes of two IP addresses, both associated with the Mozi Botnet family. These IPs are visually linked in the figure, with the timeline indicating that their malicious activity was first detected on 26 December 2020, and persisted until 26 March 2024. This persistence underlines the enduring threat posed by these entities. The graphic representation details the techniques utilized by the attackers, specifically noting the number of 'Scan' actions performed. It is observed that the upper IP conducted 'Scan 159' actions, affected 'Ports 2', and reached 'Targets 1k+', while the lower IP was more aggressive, with 'Scan 542', targeting the same number of ports, 'Ports 2', but hitting a higher number of targets, 'Targets

3k+'. These comparable yet distinct patterns underscore our methodology's effectiveness in discerning and correlating cyber threats.

Table 4. Subnet information for 111.XXX. XXX.148.

Source IP	Location	Last Seen	Carrier Information	Threat Types
111. XXX. XXX.150	'Henan, China'	2024-03-25	'China Mobile'	['Zombie', 'Exploit']
111. XXX. XXX.151	'Henan, China'	2024-03-25	'China Mobile'	['Brute Force', 'Web Spam' , 'Exploit']
111. XXX. XXX.152	'Henan, China'	2024-03-25	'China Mobile'	['Scan', 'Exploit', 'Dynamic IP']
111. XXX. XXX.154	'Henan, China'	2024-03-25	'China Mobile'	['Scan', 'Exploit', 'Dynamic IP']
111. XXX. XXX.153	'Henan, China'	2024-03-25	'China Mobile'	['Scan', 'Exploit', 'Dynamic IP']
111. XXX. XXX.155	'Henan, China'	2024-03-25	'China Mobile'	['Scan', 'Dynamic IP']
111. XXX. XXX.156	'Henan, China'	2024-03-25	'China Mobile'	['Scan', 'Exploit', 'Dynamic IP']
111. XXX. XXX.157	'Henan, China'	2024-03-25	'China Mobile'	['Scan', 'Exploit', 'Dynamic IP']
111. XXX. XXX.158	'Henan, China'	2024-03-25	'China Mobile'	['Scan', 'Dynamic IP']
111. XXX. XXX.159	'Henan, China'	2024-03-25	'China Mobile'	['Scan', 'Exploit', 'Dynamic IP']
111. XXX. XXX.160	'Henan, China'	2024-03-25	'China Mobile'	['Scan', 'Exploit', 'Dynamic IP']

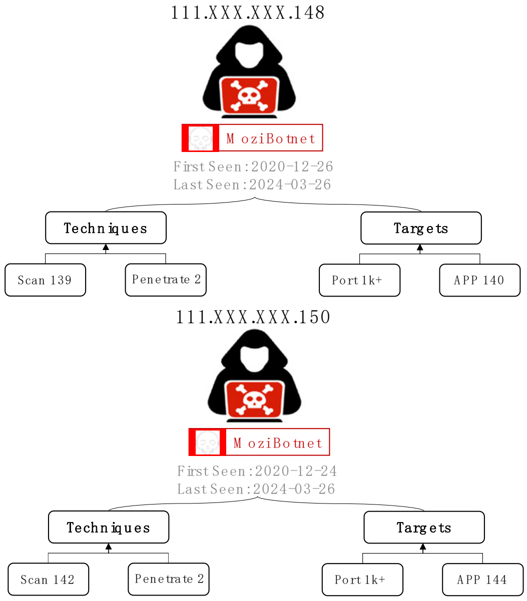


Figure 5. APT group attribution.

Zooming out from these specifics, this comparative analysis validates the heightened accuracy of our threat attribution process. The systematic integration of multidimensional data gleaned from multiple intelligence sources is critical, as it significantly sharpens our capacity to pinpoint and link these IPs to their originating cybercriminal factions. This robust, multi-sourced intelligence paradigm enriches the precision of our cyber threat attribution efforts, emphasizing the critical role that a thorough data analysis framework plays in the realm of effective cybersecurity investigations.

5. Conclusions

The paper's significant contribution is the introduction of IPAttributor, a sophisticated IPv4-based cyber attacker attribution model. This model not only enhances the analytical landscape by merging comprehensive threat intelligence with detailed network intrusion data but also utilizes a refined clustering approach to accurately identify and characterize communities of attackers. Moreover, its robustness and precision are demonstrated through rigorous validation with a large-scale, real-world dataset, achieving the highest accuracy of

88.89% and silhouette coefficient of 0.44, showcasing its potential as a valuable model in cybersecurity attribution analysis.

The primary limitation of this study is the challenge of conducting a comparative analysis due to the lack of publicly available, comparable datasets or closely related work in the field. This situation stems from the specialized nature of cyber attacker attribution, where datasets are often proprietary or classified, and methodologies are not universally shared or standardized. The absence of accessible, equivalent datasets and established benchmarks in this niche area hinders the ability to perform a detailed comparative study, thus constraining a comprehensive evaluation of the IPAttributor's performance against other existing models. This limitation underscores the need for future research to focus on creating or accessing open datasets and establishing benchmark standards for cyber attacker attribution, enabling more robust and comparative validations of models.

Author Contributions: Conceptualization, X.X. and H.L.; methodology, X.X.; software, H.L.; validation, X.X. and H.L.; formal analysis, X.X.; investigation, H.Z.; resources, X.X.; data curation, H.L.; writing—original draft preparation, X.X.; writing—review and editing, L.Z.; visualization, X.X.; supervision, Z.G.; project administration, Z.G.; funding acquisition, Z.G. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China (Grant No. 62372137), Major Key Project of PCL (PCL2022A03), and Zhejiang Provincial Natural Science Foundation of China (LZ22F020007).

Data Availability Statement: Due to privacy restrictions, the data supporting all the findings in the article are not accessible.

Acknowledgments: We would like to express our gratitude for the constructive suggestions offered by the anonymous reviewers.

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

References

1. Milajerdi, S.M.; Gjomemo, R.; Eshete, B.; Sekar, R.; Venkatakrishnan, V.N. HOLMES: Real-Time APT Detection through Correlation of Suspicious Information Flows. In Proceedings of the 2019 IEEE Symposium on Security and Privacy (SP), San Francisco, CA, USA, 20–22 May 2019.
2. Ren, Y.; Xiao, Y.; Zhou, Y.; Zhang, Z.; Tian, Z. CSKG4APT: A Cybersecurity Knowledge Graph for Advanced Persistent Threat Organization Attribution. *IEEE Trans. Knowl. Data Eng* **2023**, *35*, 5695–5709. [CrossRef]
3. Sun, N.; Ding, M.; Jiang, J.; Xu, W.; Mo, X.; Tai, Y.; Zhang, J. Cyber Threat Intelligence Mining for Proactive Cybersecurity Defense: A Survey and New Perspectives. *IEEE Commun. Surv. Tutor.* **2023**, *25*, 1748–1774. [CrossRef]
4. Jia, Y.; Gu, Z.; Du, L.; Long, Y.; Wang, Y.; Li, J.; Zhang, Y. Artificial Intelligence Enabled Cyber Security Defense for Smart Cities: A Novel Attack Detection Framework Based on the MDATA Model. *Knowl.-Based Syst.* **2023**, *276*, 110781. [CrossRef]
5. Du, L.; Gu, Z.; Wang, Y.; Wang, L.; Jia, Y. A Few-Shot Class-Incremental Learning Method for Network Intrusion Detection. *IEEE Trans. Netw. Serv. Manag.* **2024**, *21*, 2389–2401. [CrossRef]
6. Tsagourias, N.; Farrell, M. Cyber Attribution: Technical and Legal Approaches and Challenges. *Eur. J. Int. Law* **2020**, *31*, 941–967. [CrossRef]
7. Li, S.; Zhang, Q.; Wu, X.; Han, W.; Tian, Z. Attribution Classification Method of APT Malware in IoT Using Machine Learning Techniques. *Secur. Commun. Netw.* **2021**, *2021*, 9396141. [CrossRef]
8. Wang, Q.; Yan, H.; Zhao, C.; Mei, R.; Han, Z.; Zhou, Y. APT Attribution for Malware Based on Time Series Shapelets. In Proceedings of the 2022 IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), Wuhan, China, 9–11 December 2022.
9. Nataraj, L.; Karthikeyan, S.; Jacob, G.; Manjunath, B.S. Malware Images: Visualization and Automatic Classification. In Proceedings of the 8th International Symposium on Visualization for Cyber Security, Pittsburgh, PA, USA, 20 July 2011.
10. Sahoo, D. Cyber Threat Attribution with Multi-View Heuristic Analysis. In *Handbook of Big Data Analytics and Forensics*; Choo, K.-K.R., Dehghantanha, A., Eds.; Springer International Publishing: Cham, Switzerland, 2022; pp. 53–73.
11. Li, S.; Li, Y.; Wu, X.; Otaibi, S.A.; Tian, Z. Imbalanced Malware Family Classification Using Multimodal Fusion and Weight Self-Learning. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 7642–7652. [CrossRef]
12. Ding, Y.; Xia, X.; Chen, S.; Li, Y. A Malware Detection Method Based on Family Behavior Graph. *Comput. Secur.* **2018**, *73*, 73–86. [CrossRef]
13. Ki, Y.; Kim, E.; Kim, H.K. A Novel Approach to Detect Malware Based on API Call Sequence Analysis. *Int. J. Distrib. Sens. Netw.* **2015**, *11*, 659101. [CrossRef]

14. Wang, Y.; Huan, P.; Jing, T.; Song, Y. Research and Implementation on WebShell Comprehensive Detection and Traceability Technology Based on High-speed Network. *Netinfo Secur.* **2021**, *21*, 65–71.
15. Bai, B.; Feng, Y.; Liu, B.; Wang, X.; He, S.; Yao, D.; Liu, Q. Research on Network Behavior-based Cyberattack Grouping Method. *J. Cyber Secur.* **2023**, *2023*, 66–80.
16. Wang, Y.; Huan, P.; Jing, T.; Song, Y. Same origin attack analysis based on features of industrial control system function code. *Comput. Eng.* **2020**, *46*, 36–42.
17. Zhang, X.; Zhang, H.; Ma, J.; Sun, P.; Wang, J. Cyber attack attribution method based on signaling game model. *Comput. Eng. Des.* **2023**, *44*, 1616–1620.
18. Noor, U.; Anwar, Z.; Amjad, T.; Choo, K.-K.R. A Machine Learning-Based FinTech Cyber Threat Attribution Framework Using High-Level Indicators of Compromise. *Future Gener. Comput. Syst.* **2019**, *96*, 227–242. [CrossRef]
19. Wang, T.; Yan, H.; Lang, B. Threat intelligence report attribution based on attention mechanism. *J. Beijing Univ. Aeronaut. Astronaut.* **2022**, *2022*, 1–13.
20. Xiao, N.; Lang, B.; Wang, T.; Chen, Y. An advanced persistent threat actor attribution method based on multimodal and multilevel feature fusion. *arXiv* **2024**, arXiv:2402.12743.
21. Perry, L.; Shapira, B.; Puzis, R. NO-DOUBT: Attack Attribution Based On Threat Intelligence Reports. In Proceedings of the 2019 IEEE International Conference on Intelligence and Security Informatics (ISI), Shenzhen, China, 1–3 July 2019.
22. Naveen, S.; Puzis, R.; Angappan, K. Deep Learning for Threat Actor Attribution from Threat Reports. In Proceedings of the 2020 4th International Conference on Computer, Communication and Signal Processing (ICCCSP), Chennai, India, 28–29 September 2020.
23. Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Efficient Estimation of Word Representations in Vector Space. *arXiv* **2013**, arXiv:1301.3781.
24. Leskovec, J.; Rajaraman, A.; Ullman, J. *Mining of Massive Data Sets*; Cambridge University Press: Cambridge, UK, 2020.
25. Nie, F.; Wang, X.; Huang, H. Clustering and Projected Clustering with Adaptive Neighbors. In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 24–27 August 2014.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.