IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 11 September 2023, accepted 2 October 2023, date of publication 6 October 2023, date of current version 13 October 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3322427

RESEARCH ARTICLE

Similarity Analysis of Ransomware Attacks Based on ATT&CK Matrix

ZHEYU SONG, YONGHONG TIAN, AND JUNJIN ZHANG
College of Data Science and Applications, Inner Mongolia University of Technology, Hohhot 010080, China

Corresponding author: Yonghong Tian (tyh@imut.edu.cn)

This work was supported in part by the Natural Science Foundation of Inner Mongolia under Grant 2020MS06026, and in part by the Science and Technology Planning Project of Inner Mongolian under Grant 2019GG303.

ABSTRACT In recent years, there has been an increasingly prevalent trend of ransomware attacks, with malicious organizations employing various techniques to gain system privileges and subsequently engaging in extortion through methods such as encrypting files or leaking information. Current research predominantly focuses on the analysis of ransomware using existing features, but there has been scarce exploration of the behavioral patterns associated with ransomware attacks. In light of this situation, we propose a ransomware attack similarity analysis method based on the ATT&CK matrix. To initiate this analysis, a substantial amount of network threat intelligence is sifted through to select reliable and comprehensive ransomware attack incidents. From these incidents, we extract attack tactics, techniques, and procedural information. Subsequently, we employ the TF-IDF algorithm to calculate the keyword weights within attack descriptions. Based on these weights, we utilize the cosine similarity algorithm to compare the similarity between attack events. This approach reveals critical technical and tactical information employed by the attacking organizations, enabling researchers to gain a deeper understanding of the behavioral patterns of the attackers. Finally, we propose countermeasures corresponding to the critical attack techniques employed by these malicious organizations. These countermeasures aim to enhance network security defenses and reduce the risks associated with ransomware attacks.

INDEX TERMS Cybersecurity, cyber threat intelligence, ransomware attack, similarity comparison, MITRE ATT&CK.

I. INTRODUCTION

During the COVID-19 pandemic, people have increasingly relied on the Internet as workplaces have shifted to home settings. This situation has provided cybercriminals with significant opportunities, leading to a surge in cyberattacks. Attackers have launched numerous phishing campaigns [1] using false COVID-19 information as bait, and there has been a substantial increase in malware attacks. Notably, ransomware has emerged as one of the most prevalent forms of malware attacks [2]. Unlike traditional malware, ransomware can inflict irreversible damage to the operating system or user files, persisting even after removal [3]. Statistics indicate that ransomware has caused $20 billion in losses in 2021, with projections suggesting this figure could escalate to $265 billion by 2031 [4]. A notable incident in June 2021 involved hackers associated with REvil exploiting a vulnerability in Kaseya's remote management software to launch ransomware attacks in 1,500 organizations worldwide [5]. This event led to the forced closure of hundreds of stores for the Swedish food chain Coop, resulting in significant financial losses for the company [6].

Ransomware attacks pose a significant and serious threat. Without proactive measures and swift responses, accurately estimating the extent of potential losses can be quite challenging. To mitigate this risk, we can employ machine learning techniques to analyze the currently collected TTPs used by contemporary leading attack groups. By doing so, we can develop proactive countermeasures that enhance our ability to respond effectively to ransomware attacks, thus minimizing potential damages.

The associate editor coordinating the review of this manuscript and approving it for publication was Mueen Uddin.

The analysis of TTPs is typically categorized into three distinct levels of detail. MITRE has developed the ATT&CK framework to optimize the utilization of this information [7]. This framework offers a comprehensive, high-level description of the behaviors employed by attackers, known as tactics. For instance, it may exemplify how an attacker gains network access through Remote Desktop Protocol (RDP). Building upon this, the framework furnishes a more intricate description of how attackers employ specific tactics, referred to as techniques. For instance, it may elucidate how attackers utilize brute-force password attacks on RDP to gain access to the network. Lastly, the framework provides highly detailed descriptions, denoted as procedures, outlining the techniques commonly employed by attackers. For instance, an attacker may exploit a list of passwords and potential usernames gathered through Open Source Intelligence (OSINT) to exploit RDP by utilizing a time-based password spraying attack. The ATT&CK framework enhances comprehension of attacker behaviors and assists organizations in identifying and responding to diverse threat scenarios.

Combining multiple TTPs employed by an attacker can serve as a distinctive signature, enabling responders to trace the origin of an attack back to a specific potential adversary. For instance, in a threat intelligence analysis of a ransomware attack, if the attacker initially gains access through a phishing email, and subsequently deploys Advanced IP Scanner and Cobalt Strike before initiating the ransomware attack, this unique combination was exclusive to the Egregor ransomware group at the time. It is rare to encounter other threat groups utilizing the same set of TTPs. Consequently, cyber extortion incident response experts regularly maintain awareness of the latest TTPs employed by various organizations. This knowledge enables rapid identification of adversaries based on their distinct TTP profiles.

Despite the extensive body of research on ransomware attacks [8], effectively tracking and countering this threat remains challenging. To tackle this issue, Modi et al. [9] introduced a method for detecting ransomware in encrypted web traffic, employing a machine learning classifier with 28 traffic-related features. Conversely, Zuhair et al. [10] proposed a multi-layer analysis model for classifying various ransomware families by leveraging a combination of static and dynamic features. However, these machine learning detection methods still encounter several issues, including inaccuracies in feature extraction and limitations in adapting to the latest attack vectors. To enhance the effectiveness of addressing ransomware attacks, our research employs a diverse set of similarity analysis methods and data analysis techniques to identify the TTPs employed by contemporary mainstream ransomware families and develop corresponding mitigation strategies. This approach aims to address the shortcomings of machine learning detection methods and enhance the traceability of ransomware attacks. In summary, this paper contributes to the field in the following ways:

1) We collected over 100 pieces of cyber threat intelligence (CTI) related to ransomware attacks from 2021 to 2022 and identified 12 attacks conducted by three distinct attacking organizations.
2) The TTPs employed by ransomware attack organizations are extracted through a comparison of keywords associated with various attack techniques in the ATT&CK matrix.
3) Utilizing a quantitative analysis approach, we calculate the tactical similarity among diverse attack activities by employing the TF-IDF algorithm and the cosine similarity method.
4) By identifying the core tactics utilized by different ransomware families and giving particular attention to the analysis of the attack techniques most commonly employed within each tactic, our objective is to propose effective mitigation strategies and detection methodologies.

The organization of this paper is structured as follows. In Section II, we delve into prior research related to ransomware attack detection and the utilization of TTPs. Section III provides a detailed exposition of the TTPs extraction process and the methodology for comparing the similarity of attack techniques. Subsequently, Section IV elaborates on the experimental process in comprehensive detail. Section V presents a comprehensive analysis and an in-depth discussion of the experimental results. In Section VI, we explicitly delineate certain limitations of the current research. Finally, Section VII offers a summary of the entire paper and a discussion of future research directions.

II. RELATED WORK

As technology continues to advance, so do ransomware attacks. Particularly with the emergence of cryptocurrency and onion routing techniques, the trend toward the modernization of ransomware attacks has become increasingly pronounced. These technologies enable attackers to communicate and make payments swiftly and anonymously, rendering ransomware attacks more prevalent and challenging to prevent. Currently, ransomware attacks primarily target systems such as laptops, mobile devices, and Internet of Things (IoT) devices. In this section, we focus on presenting activities related to ransomware attacks, discussing detection methods, and providing a literature review of utilizing TTPs to analyze cyberattacks.

A. RANSOMWARE ATTACK ACTIVITIES

Creating backups of critical files constitutes an effective measure to counter the encryption tactics employed by ransomware, enabling data recovery without succumbing to ransom demands. The significance of this countermeasure was exemplified in an attack by the Revil ransomware family on August 19, 2019. During this incident, a substantial volume of vital government files fell victim to ransomware encryption, causing severe disruptions to almost

all operational business systems. The attacker, in this case, demanded a hefty ransom of $2.5 million in exchange for the decryption key [11]. Although the business operations were severely impacted, the targeted city made a principled decision not to yield to the ransom demand. Instead, they successfully restored most of their data from backup archives. However, it is noteworthy that an increasing number of extortion organizations have come to realize that encrypting files alone may not suffice to extract a ransom. Consequently, they employ additional tactics to exert pressure on their victims. For instance, in December 2019, the Maze group launched an attack on Southwire, encrypting 878 devices and demanding a $6 million ransom. Southwire, however, refused to pay the ransom and promptly initiated the process of system restoration on the same day as the attack. The attackers, anticipating this response, retaliated by posting over 14GB of Southwire's pilfered data on their website. They further threatened to release 10% of the data every week until the ransom was paid [12]. Extortion organizations continually refine their extortion strategies, as observed in their inclusion of distributed Denial of Service (DDoS) attacks as part of their threat arsenal in 2021. This addition aims to exert greater pressure on victims to comply with ransom demands. Therefore, creating backups of crucial files stands as a fundamental step in safeguarding against ransomware attacks. Concurrently, it is imperative to remain vigilant and adaptive to the evolving strategies of ransomware organizations [13].

B. METHODS FOR DETECTING RANSOMWARE ATTAKS

1) LAPTOPS

Ransomware primarily targets laptop computers, resulting in a plethora of studies focusing on the detection and prevention of this particular ransomware variant. In recent years, machine learning algorithms have exhibited significant potential for ransomware detection. Masum et al. [14] introduced a framework based on feature selection to assess the performance of various machine learning algorithms in ransomware detection and prevention. Experimental results indicated that the random forest method surpassed other approaches in terms of detection performance. Addressing the limitation of supervised learning detection models in detecting new ransomware strains, Sharmeen et al. [15] proposed an adaptive detection framework capable of extracting features from novel ransomware variants and integrating them into the supervised learning detection model. Zhang et al. [16] presented a deep learning-based N-gram opcode static analysis framework employing a convolutional neural network (CNN) with a self-attention mechanism. This approach effectively captures rich contextual and semantic information from exceedingly long sequences, resulting in a significant enhancement in classification performance. Jethva et al. [17] devised a ransomware detection system that amalgamates machine learning and detection rules. This system employs machine learning models such as Support Vector Machines (SVM), random forest, and logistic regression to identify ransomware features, including registry key operations, API calls, and DLLs. Detection rules are used to monitor ransomware entropy and file signature changes.

2) MOBILE DEVICES

In the realm of network security, ransomware attacks targeting mobile devices are increasingly prevalent. Consequently, numerous researchers have proposed diverse methods to combat this issue. Among these methods, Ahmed et al. [18] introduced a behavior-based dynamic analysis framework. This framework leverages TF-IDF technology to identify the most informative features from samples and employs Support Vector Machine (SVM) and Artificial Neural Network (ANN) to develop and implement a machine learning-based detection model. The model demonstrates the ability to accurately identify specific behavioral characteristics associated with highly survivable ransomware (HSR) with remarkable precision and accuracy. Alzahrani et al. [19] presented an automated and lightweight ransomware behavior detection system. They enhance detection performance by enriching the dataset with various types of information and increasing the feature set through dynamic analysis. The authors also conducted experiments with multiple machine learning classification algorithms to ensure optimal precision and accuracy. On the other hand, Scalas et al. [20] proposed an API-based detection strategy relying on system Application Programming Interfaces (APIs). This strategy effectively distinguishes between general malware, ransomware, and legitimate software. It also enhances the interpretability of features detected by the classifier. Faris et al. [21] introduced an Android ransomware detection framework that utilizes machine learning and meta-heuristic algorithms. This framework combines a hybrid model of the Salp Swarm Algorithm (SSA) and Kernel Extreme Learning Machine (KELM). Experimental results demonstrate that the model exhibits superior detection performance compared to many robust classifiers.

3) IOT DEVICES

With the proliferation of IoT devices in domains such as smart homes, smart transportation, and smart cities, the substantial economic value generated by these devices is increasingly becoming a magnet for ransomware attacks. In response to the rising threat of ransomware attacks targeting IoT devices, Al-hawawreh et al. [22] introduced a detection model founded on stacked Variational Autoencoder (VAE). This model possesses the capability to discern the latent structure within system activity and effectively identify ransomware attacks. During the training process, the authors employed a VAE-based data augmentation method to generate new training data, thereby enhancing the detection performance of the model.

C. ANALYSIS OF CYBER ATTACKS USING TTPS

With the increasing frequency of network attacks, security researchers have initiated efforts to organize unstructured

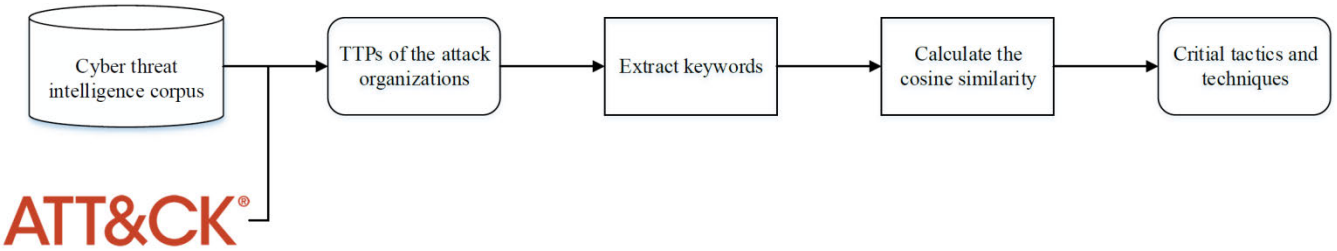

FIGURE 1. Workflow of the ransomware attack similarity analysis method.

information in threat intelligence, such as Indicators of Compromise (IOCs) or TTPs, into structured content. This restructuring enables the study of the constantly evolving landscape of network attacks. Various methods have been proposed to extract and classify TTPs in pursuit of this objective. Zhang et al. [23] introduced the EX-Action framework, which employs a multi-modal algorithm to identify and extract threat behaviors in Cyber Threat Intelligence. It quantifies differences in threat behaviors through the evaluation index of Normalized Mutual Information (NMI). Addressing challenges related to small training datasets and class imbalances in TTPs, Kim et al. [24] utilized data augmentation. Experimental results demonstrated that this approach effectively enhances the classification performance of TTPs, resulting in a performance improvement ranging from 60% to 80% when compared to the TRAM model. In the TTPs extraction process, machine learning methods, such as named entity recognition technology, can be employed to filter out irrelevant information [25]. Liu et al. [26] proposed a neural network-based TTPs extraction model using Transformers. This model incorporates an attention loop structure to simulate the relationship between tactics and techniques. It also features a hierarchical classification module for extracting tactics and techniques from threat reports. Additionally, You et al. [27] introduced a context-enhanced threat intelligence TTP mining framework. This framework classifies mined TTP information at the sentence level and organizes classified TTP elements into shareable threat intelligence using the Sigma detection rules format in STIX 2.1. Differing from previous document-level classification methods, this framework provides a more granular classification of TTP elements. These methods offer effective means of mining threat intelligence in the context of network security.

Existing research in this domain has rarely directly analyzed attack behaviors by assessing the similarity of TTPs. Shin et al. [28] adopted a methodology wherein they treated the occurrence of specific attack techniques within attack activities as binary values and established the similarity of attack activities by comparing these binary values. However, this approach relied solely on the presence or absence of attack techniques to deduce critical tactical information. It is important to note that certain sub-techniques under specific attack techniques might exhibit high similarity. Depending exclusively on this method may lead to reduced experimental accuracy. In contrast, our study enhances the accuracy of analysis by evaluating the semantic similarity of TTPs during the course of attack processes.

III. METHODOLOGY

The proposed framework aims to harness the similarities among various ransomware attacks for the identification of key attack tactics employed by malicious organizations. Subsequently, it seeks to develop mitigation measures for the detection and mitigation of ransomware attacks. An overview of the framework's overall flowchart is depicted in Fig. 1. Initially, relevant threat intelligence pertaining to ransomware attacks is collected from diverse open-source threat intelligence platforms. Information on TTPs within the attack process is extracted, relying on the keyword characteristics of attack techniques present in the ATT&CK matrix. Subsequently, employing the extracted attack techniques, corresponding descriptions of the attack process are identified within the ATT&CK framework. Following this, the framework utilizes TF-IDF to extract keywords from the text and transform word frequencies into vector representations for subsequent cosine similarity comparisons. Through this comparison of the similarity among multiple attack process descriptions, critical attack tactics are identified, and corresponding mitigation measures are determined. To provide a comprehensive understanding of the framework, each subsection below delineates the specific processes involved in each stage of the proposed framework. These processes encompass threat intelligence collection, TTPs information extraction, extraction of attack process descriptions, and the similarity comparison of multiple attack process descriptions.

A. COLLECT DATA

To ensure the reliability, timeliness, and completeness of the analysis regarding the attack process, we adopted a two-step strategy. Initially, we gathered cyber threat intelligence associated with ransomware attacks that transpired within the past two years. This data was procured from eight distinct threat intelligence providers, including AlienVault OTX [29] and The DFIR Report [30], among others. Subsequently, we meticulously screened all the collected data and narrowed our research focus to 16 ransomware attacks, which were linked to three attacking groups. This careful selection

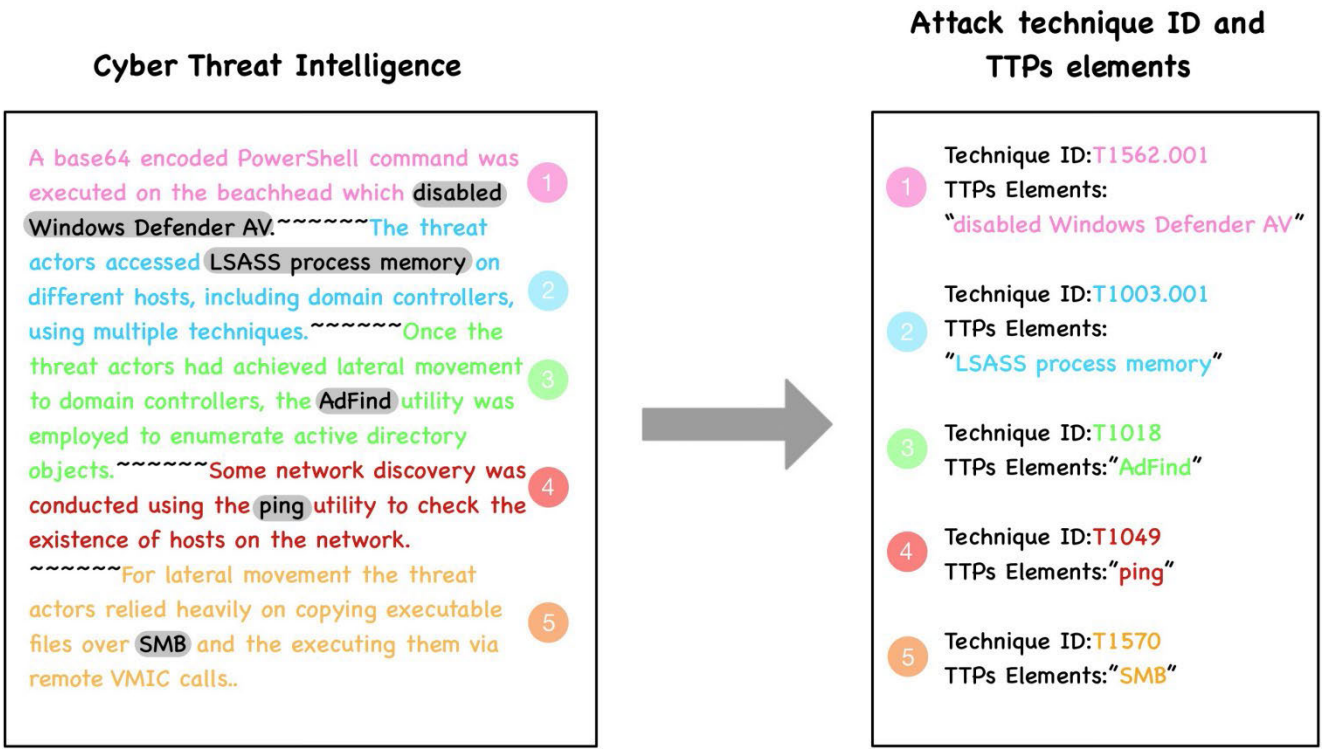

FIGURE 2. The manual extraction process of TTPs. The left figure is the original network threat intelligence, and the right figure is the corresponding attack technology ID and TTP elements. Different colors in the figure represent different TTPs.

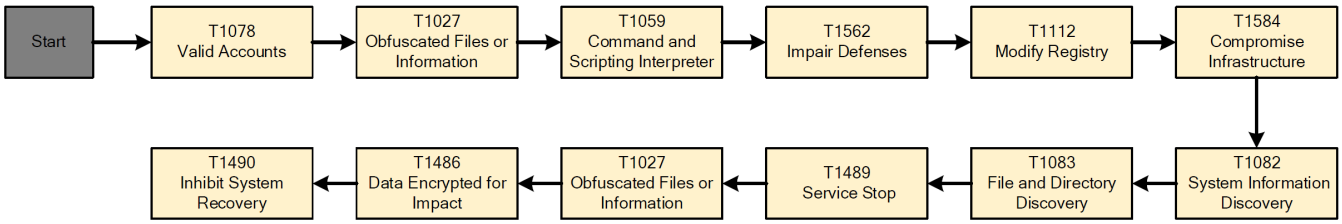

FIGURE 3. The critical attack path extracted from the offensive activities.

process ensured the robustness and relevance of our chosen dataset for in-depth analysis.

B. EXTRACT TTPS

During the TTPs extraction process, two essential steps must be followed. Firstly, we must ascertain the tactic employed in the current attack process by analyzing the description provided within the threat intelligence statement. Secondly, we need to pinpoint the specific attack technique by identifying the keywords present in the statement. It's noteworthy that the latest iteration of the ATT&CK framework encompasses 14 distinct tactics and 193 techniques. However, there exist certain keywords that can correspond to multiple attack techniques. For instance, when the keyword "SMB" appears, it may be associated with attack techniques such as remote services, lateral tool transfer, or exploiting public-facing applications. Directly relying on keywords for matching can significantly reduce the accuracy of such matches. Fig. 2 illustrates the process of TTPs extraction from cyber threat intelligence. The intelligence text on the left employs different colors to represent various TTPs descriptions, while certain indicators of compromise (IoC) and security terms within it are highlighted with a gray background. We refer to this type of content as TTPs elements. By analyzing the intelligence text on the left, we can derive the corresponding attack techniques and TTPs elements. Fig. 3 presents an attack chain where we have mapped the attack activities onto the ATT&CK framework, spanning from the initial access tactics involving trusted account credentials to the final stage of preventing system data recovery, thereby concluding the ransomware attack.

C. EXTRACT THE DESCRIPTION OF THE ATTACK TECHNIQUE

The ATT&CK framework offers an intricate description of attack techniques, which we can employ to extract a detailed

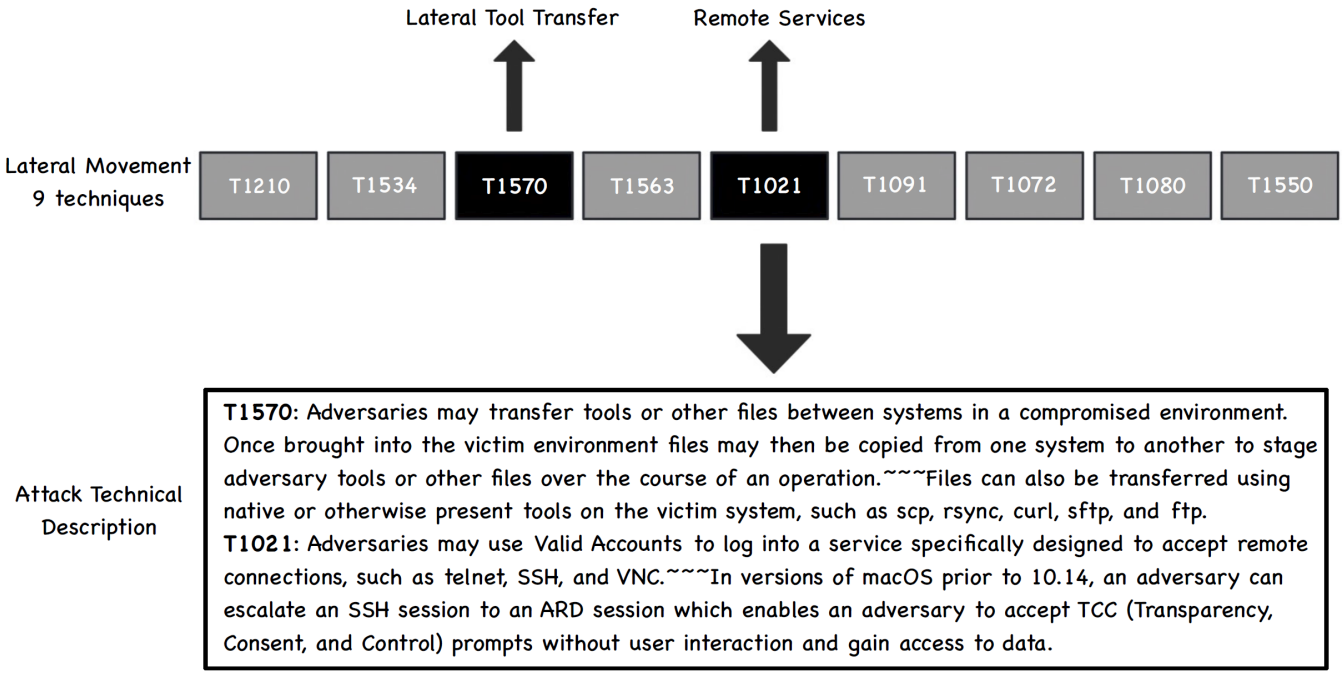

FIGURE 4. Extracting attack technique descriptions in tactical lateral movement. When lateral tool transfer and remote services are used, extract their descriptions from the ATT&CK framework.

account of the attack techniques employed by attacking groups within each tactic. Fig. 4 provides an overview of the specific extraction process related to tactical lateral movement. In an observed attack, we identified the presence of attack techniques T1570: Lateral Tool Transfer and T1021: Remote Services, both of which fall under this specific tactic. Consequently, the extracted attack technique descriptions are presented in the lower part of Fig. 3. For the sake of convenience in subsequent similarity comparison operations, we have organized the extracted attack technique descriptions under the respective agreed-upon tactics into a cohesive string.

D. COMPARE THE SIMILARITY OF THE ATTACK PROCESS

Through the comparison of the similarity among attack technique descriptions for each attack, we can ascertain the degree of similarity in attack tactics employed within each attack. This analysis enables us to pinpoint the most effective countermeasures for mitigating the attack. In instances where two attacks exhibit a higher consistency in the utilization of attack techniques, the resulting cosine similarity in the final attack description tends to be elevated. Conversely, when the two attacks do not employ the same attack techniques, the computed cosine similarity will be 0.

Prior to conducting a comparison of the similarity among attack technique descriptions, we employ the TF-IDF algorithm to compute the weight of keywords within the text. This calculation involves determining the frequency of a keyword's appearance in the current description and its overall frequency of occurrence in the two descriptions. The resulting TF-IDF value of the keyword signifies its importance within the current description. Furthermore, the TF-IDF value is directly proportional to the keyword's significance, resulting in a larger vector value. Utilizing the TF-IDF algorithm enhances our ability to conduct more robust similarity comparisons. The specific calculation method is elucidated in Formula (2).

$$\mathrm{tf}_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \tag{1}$$

$$W_{i,j} = \mathrm{tf}_{i,j} \times \log \frac{N}{n_i + 1} \tag{2}$$

In the equation, $W_{i,j}$ represents the TF-IDF score of word W_i in the attack description, $tf_{i,j}$ signifies the frequency of the word W_i occurring in the current description, N represents the total number of documents within the document set, n_i denotes the count of documents containing the word W_i, and the addition of $+1$ is included to prevent the denominator from reaching 0.

We initiated the process by vectorizing the attack technique descriptions based on their respective weights and subsequently calculated the similarity among these descriptions within each tactic. While exploring various algorithms for text similarity calculation, including the SimCSE [31] and PromCSE models [32], which have demonstrated commendable performance in comparing the similarity of short texts, it became evident that these methods were not suitable for our research due to the necessity of handling longer text data, as illustrated in Fig. 3. We computed the similarity of attack technique descriptions within each tactic using both the Euclidean distance algorithm and the cosine similarity algorithm. Cosine similarity is a metric for assessing

similarity by calculating the cosine value of the angle between two vectors. When vectors align in the same direction, it signifies text similarity, and as the angle becomes smaller, the similarity score increases, approaching 1. Conversely, when vectors have larger angles between them, they are less similar, resulting in a similarity score closer to 0 [33].

The calculation method for cosine similarity of tactics is depicted in Equation (3), where $\mathrm{sim}(\mathrm{X_i}, \mathrm{Y_i})$ represents the similarity of the i-th tactic in attack X and attack Y. In the formula, X_i and Y_i respectively denote the vector representations of the attack technique description in the i-th tactic within the corresponding attack.

$$\mathrm{sim}(\mathrm{X_i}, \mathrm{Y_i}) = \begin{cases} 0, \text{ one of } X_i \text{ or } Y_i \text{ is zero vector,} \\ \dfrac{\sum_n^{i=1} X_i Y_i}{\sqrt{\sum_n^{i=1} (X_i)^2} \times \sqrt{\sum_n^{i=1} (Y_i)^2}}, \text{ otherwise} \end{cases} \tag{3}$$

The Euclidean distance is a metric employed to gauge the distance or dissimilarity between vectors. It quantifies the dissimilarity between two vectors by computing the straight-line distance between their respective points. The smaller the Euclidean distance between two vectors, the closer they are positioned in space, indicating a higher degree of similarity. When two vectors occupy the exact same position in space, the Euclidean distance is 0, signifying complete similarity. Conversely, if two vectors are situated significantly apart in space, their Euclidean distance will be larger, indicating greater dissimilarity.

$$d(P, Q) = \sqrt{\sum_n^{i=1} (P_i - Q_i)^2} \tag{4}$$

IV. EXPERIMENT

Our analysis centers on the primary tactics employed in ransomware attacks and assesses the variety of techniques within each tactic. By computing the cosine similarity of the attack technique descriptions within each tactic, we can determine the level of diversity among these techniques. Higher similarity scores indicate lower diversity in attack techniques, whereas lower similarity scores imply greater diversity within the tactic's techniques. Simultaneously, the attack tactics with the highest similarity in each attack signify the weakest links in our cyber kill chain.

We have chosen 12 ransomware attack incidents from the gathered threat intelligence, spanning from 2021 to 2022. These incidents are associated with the Conti, Lockbit, and Hive ransomware attack groups. Specifically, Conti conducted 6 attacks, Lockbit carried out 3 attacks, and Hive executed 3 attacks. We have labeled these attacks as $C1 - C6$, $L1 - L3$, and $H1 - H3$ in chronological order. This dataset allows us to analyze the attack techniques employed by each ransomware attack group during various timeframes. For the purpose of similarity analysis, we have transformed the TTPs information from each attack into descriptions of attack techniques in the ATT&CK framework. If a particular tactic was not employed in an attack or was used solely in a single attack, the similarity values for that tactic will be entirely 0. To illustrate, in the case of Conti ransomware, the tactic TA0010: Exfiltration was never utilized in any of the six attack incidents, resulting in all similarity values in that column being 0.

The D3FEND knowledge graph, as referenced in [34], offers a viable solution for mitigating the attack techniques employed by APT organizations. Within this knowledge graph, D3FEND presents ten techniques designed to detect T1570: Lateral Tool Transfer. These techniques encompass: Protocol metadata anomaly detection, remote terminal session detection, user geolocation login pattern analysis, network traffic filtering, connection attempt analysis, file carving, network traffic community bias, per-host download-upload ratio analysis, client-server load analysis, and asset vulnerability enumeration. Each of these techniques possesses unique characteristics and can effectively detect various attack scenarios. By leveraging these techniques, we can enhance the security of our systems and reduce the vulnerability to attacks orchestrated by APT groups.

V. DISCUSSION

We conducted 66 similarity comparisons among the 12 attack incidents and computed the average similarity for each attack tactic, as presented in Table 1. Notably, the TA0008: Lateral Movement tactic exhibited the highest average similarity in ransomware attacks, reaching 0.5725. This tactic holds significant importance in the context of ransomware attacks and merits considerable attention from defenders. TA0008: Lateral Movement encompasses a variety of techniques employed to infiltrate and control remote systems. Although the specific techniques utilized may vary across different attacks, our experiments indicate that the T1570: Lateral Tool Transfer and T1021: Remote Services techniques are the most frequently employed. Attackers gain access to systems by leveraging T1021: Remote Services, allowing them to login through secure shell (SSH) or remote desktop protocol (RDP) and subsequently execute malicious software. Furthermore, they employ T1570: Lateral Tool Transfer to transfer attack tools within the target system. To detect adversary activities in ransomware attacks effectively, monitoring remote services and file transfers can be instrumental. Implementing security measures like multi-factor authentication and file sharing restrictions can help mitigate these two attack behaviors within the realm of ransomware.

In addition to analyzing the attacks comprehensively, we also conducted similarity comparisons for each ransomware family. Tables 2, 3 and 5 display the average cosine similarities of attack tactics for the Conti, Lockbit, and Hive ransomware families, respectively. To compare the tactical similarity across different attack campaigns, it was essential to perform pairwise statistical analysis on the experimental results for each pair of attack campaigns.

This approach ensured that we obtained comprehensive and accurate comparative data. Based on our experimental findings, we identified key tactics for the attack organizations

TABLE 1. The average cosine similarity of all attack activities.

	TA0001	TA0002	TA0003	TA0004	TA0005	TA0006	TA0007	TA0008	TA0009	TA0010	TA0011	TA0040
Avg.	0.1026	0.4481	0.1984	0.2162	0.2850	0.1710	0.3703	0.5725	0.0436	0.0239	0.0653	0.5640

TABLE 2. The cosine similarity of ransomware attacks initiated by conti.

	TA0001	TA0002	TA0003	TA0004	TA0005	TA0006	TA0007	TA0008	TA0009	TA0010	TA0011	TA0040
(C1,C2)	0	0.2840	0	0	0.2336	0	0.3880	0	0	0	0	0.8254
(C1,C3)	0	0.2613	0	0	0.2607	0.2197	0.5836	0.7240	0	0	0	1
(C1,C4)	0	0.4797	0	0.6885	0.5411	0	0.5144	1	0	0	0	1
(C1,C5)	0	0.2812	0.7106	0.3797	0.3003	0.2283	0.6375	0.7490	0	0	0	0
(C1,C6)	0	0.2760	0.2834	0.2381	0.5472	0.2263	0.4877	0.7240	0	0	0.3880	1
(C2,C3)	0	0.2166	0	0	0.2570	0	0.4591	0	0	0	0	0.8254
(C2,C4)	0	0.4247	0	0	0.1934	0	0.4956	0	0	0	0	0.8254
(C2,C5)	0	0.2561	0	0	0.2021	0	0.5843	0	0	0	0	0
(C2,C6)	0	0.4001	0	0	0.1979	0	0.5118	0	0	0	0	0.8254
(C3,C4)	0	0.6185	0	0	0.1985	0	0.6758	0.7240	0.6405	0	0	1
(C3,C5)	0.8465	0.6197	0	0	0.2599	0.7901	0.6172	0.7056	0	0	0	0
(C3,C6)	0	0.5985	0	0.3663	0.2291	0.8848	0.6201	1	0	0	0	1
(C4,C5)	0	0.6627	0	0.1710	0.2035	0	0.6121	0.7490	0	0	0	0
(C4,C6)	0	0.6477	0	0.1974	0.1848	0	0.7412	0.7240	0	0	0	1
(C5,C6)	0	0.4752	0.2748	0	0.3050	0.8828	0.5718	0.7056	0	0	0	0
Avg.	0.0564	0.4335	0.0846	0.1361	0.2743	0.2155	0.5667	0.5203	0.0427	0	0.0259	0.6201

TABLE 3. The cosine similarity of ransomware attacks initiated by lockbit.

	TA0001	TA0002	TA0003	TA0004	TA0005	TA0006	TA0007	TA0008	TA0009	TA0010	TA0011	TA0040
(L1,L2)	0.2718	0.2556	0.3287	0.2975	0.1907	0.2197	1	0.8247	0	0.3070	0.2689	0.7398
(L1,L3)	0.3708	0.5673	0.3307	0.3167	0.2607	0	0.2418	0.7551	0.1536	0	0.6496	0.6314
(L2,L3)	0.2381	0.3146	0.3491	0.3342	0.4620	0	0.2418	0.8840	0	0	0.1863	0.8294
Avg.	0.2936	0.3792	0.3362	0.3161	0.3045	0.0732	0.4945	0.8213	0.0512	0.1023	0.3683	0.7335

TABLE 4. The euclidean distance of ransomware attacks initiated by lockbit.

	TA0001	TA0002	TA0003	TA0004	TA0005	TA0006	TA0007	TA0008	TA0009	TA0010	TA0011	TA0040
(L1,L2)	12.0416	13.9284	17.6068	17.6918	20.5913	15.0665	0	8.8882	-	10.0000	16.1864	10.7238
(L1,L3)	14.7648	18.2483	20.3224	24.6982	24.7588	-	9.5394	10.6771	9.5917	-	11.7047	14.4222
(L2,L3)	11.1803	21.2368	19.2094	23.0434	21.6102	-	9.5394	10.6771	-	-	13.2288	10.1489

TABLE 5. The cosine similarity of ransomware attacks initiated by hive.

	TA0001	TA0002	TA0003	TA0004	TA0005	TA0006	TA0007	TA0008	TA0009	TA0010	TA0011	TA0040
(H1,H2)	0.1973	0.4096	0.2062	0.1879	0.2595	0.2397	0.2795	1	0	0	0	0
(H1,H3)	0	0.5804	0.2330	0.2140	0.3474	0	0.3245	0.4714	0	0	0	0.6592
(H2,H3)	0	0.3064	0.2428	0.2615	0.2509	0	0.2056	0.4714	0	0	0	0
Avg.	0.0658	0.4321	0.2273	0.2211	0.2859	0.0799	0.2699	0.6476	0.0000	0.0000	0.0000	0.2197

listed in Table 6, with each tactic being associated with the most commonly employed attack techniques. These insights contribute to a better understanding of adversary attack patterns. The tactic TA0040: Impact, linked to the Conti

TABLE 6. Critical tactics and frequently used techniques by the APT groups are described. Critical tactics and frequent techniques are considered as weakest link to mitigate.

APT Groups	Critical Tactics	Frequent Techniques
Conti	TA0007: Discovery TA0040: Impact	T1087: Account Discovery T1018: Remote System Discovery T1082: System Informeration Discovery T1482: Domain Trust Discovery T1486: Data Encrypted for Impact
Lockbit	TA0008: Lateral Movement TA0040: Impact	T1021: Remote System T1486: Data Encrypted for Impact
Hive	TA0002: Execution TA0008: Lateral Movement	T1059: Command and Scripting Interpreter T1021: Remote System

ransomware, exhibited a similarity score of 1 in 6 out of 15 similarity comparisons, while the tactic TA0007: Discovery had a similarity score exceeding 0.5 in 11 instances. The average similarity between these two tactics ranked the highest among the 12 tactics, underscoring their significance within the ransomware context. TA0007: Discovery involves gathering system information for subsequent attack actions, whereas TA0040: Impact aims to disrupt or modify data within the system, typically serving as the ultimate goal in ransomware attacks.

Regarding the 7 tactics outlined in Table 2, which include TA0001: Initial Access and TA0003: Persistence, more than half of the similarity results for Conti ransomware were 0. This implies that these tactics are infrequently employed by the ransomware in its attack operations. Conversely, several similarity values below 0.3 were observed in the results for TA0002: Execution and TA0005: Defense Evasion tactics. This suggests that the ransomware frequently alters the attack techniques within these tactics to fulfill their tactical objectives. Based on the data presented in Table 6, the Conti ransomware predominantly utilizes 5 specific attack techniques. Among these, T1087: Account Discovery and T1482: Domain Trust Discovery, both within TA0007: Discovery, can be mitigated by configuring Windows components to disable administrator enumeration and by reducing trust relationships. However, for the techniques T1018: Remote System Discovery and T1082: System Information Discovery in TA0007: Discovery, there are currently no effective mitigation measures available. Consequently, timely detection of these two techniques through monitoring operations like checking the /etc/hosts file and monitoring newly executed processes is essential to prevent further system damage. The technique T1486: Data Encrypted for Impact within tactic TA0040: Impact is typically employed as the final step in ransomware attacks to encrypt data within the system. To counteract this attack, we can implement Attack Surface Reduction (ASR) rules or maintain system data backups.

As per the data presented in Table 3, the average similarities for the Lockbit ransomware's tactics TA0008: Lateral Movement and TA0040 across 3 attack incidents were 0.8213 and 0.7335, respectively. In comparison to the similarity score of 0 for tactics TA0006, TA0009, and TA0010, it is plausible that Lockbit ransomware either does not pursue the objectives of these three tactics or employs different attack techniques throughout the attack process. By scrutinizing the frequency of usage of attack techniques within TA0008: Lateral Movement and TA0040, we discerned that the two most frequently employed attack techniques by Lockbit in ransomware attacks are T1021: Remote Service and T1486: Data Encrypted for Impact. Table 4 presents the Euclidean distances between the three attack campaigns initiated by the Lockbit ransomware family. According to the Euclidean distance principles elucidated in Section 2.4.2, larger distance values indicate greater dissimilarity between two vectors, while a distance of 0 signifies complete similarity between two vectors. However, it is not feasible to represent absolute dissimilarity between two vectors in the table by using infinity. Consequently, this approach cannot be utilized to compute the average similarity among all tactics.

Based on the data provided in Table 5, the average similarities for the Hive ransomware's tactics TA0002: Execution and TA0008: Lateral Movement across 3 attack incidents were 0.4321 and 0.6476, respectively. In contrast, the similarity results for other tactics were generally low. Tactics TA0009, TA0010, and TA0011 had a similarity score of 0, mirroring the situation observed with the Conti ransomware. The Hive ransomware predominantly employs two attack techniques in TA0002: Execution and TA0008: Lateral Movement, namely, T1059: Command and Scripting Interpreter and T1021: Remote Service. In summary, to effectively detect and mitigate ransomware attacks in a timely manner, it is advisable to configure additional detection strategies targeting the characteristics of TA0002: Execution, TA0007: Discovery, TA0008: Lateral Movement, and TA0040: Impact tactics. Enhancing the recognition capabilities for these tactics and attack techniques will subsequently improve the overall security of the system.

VI. LIMITION

This study has identified that the key tactics employed by various ransomware attack groups may not be identical. Our approach enables precise analysis of these key tactics, and with the aid of detection measures provided in the D3FEND knowledge graph, corresponding attack techniques can be promptly identified. However, it is essential to acknowledge that there is room for further refinement and development of our method. Firstly, addressing the scarcity of threat intelligence data sources is imperative. In addition to aggregating intelligence from open-source threat intelligence platforms, we can explore the direct extraction of threat intelligence from vulnerability lists reported in public code repositories, such as Github [35], GitLab [36], and Bitbucket [37]. Secondly, the issue of the reliability of threat intelligence sources requires attention. The collected intelligence may inadvertently include false information generated by malicious entities, thereby introducing erroneous TTPs data. Presently, research on data risk control remains relatively limited, and an effective mechanism for intelligence analysis is lacking. In reference to the findings of a previous study [38], which highlights the capability of fine-tuned GPT-2 models to generate counterfeit CTI text that is challenging even for professional security experts to discern, we can draw insights from deep learning techniques to investigate and establish a risk assessment model that offers enhanced adaptability and efficacy. Lastly, there are limitations associated with the current TTPs information extraction methods. While certain contemporary approaches advocate the use of deep learning techniques for extracting TTPs information from threat intelligence, many of these methods suffer from suboptimal accuracy. Therefore, we could contemplate integrating the keyword matching and deep learning methodologies outlined in this paper to enhance the precision of TTPs information extraction.

VII. CONCLUSION AND FUTURE WORK

The proliferation of sophisticated attack techniques employed by ransomware groups has emerged as a substantial challenge within the realm of cybersecurity, often leaving defenders grappling with a sense of helplessness. This threat continues to escalate and has already cast a substantial shadow over the cyber landscape. In response to this burgeoning menace, we introduce an innovative approach designed to facilitate the selection of effective defense strategies through an analysis of attack tactic similarities. Our research endeavors to furnish a pragmatic framework for cyber threat analysis, which can serve as a valuable resource for network security researchers in devising more robust defense strategies. In the course of our investigation, we scrutinized ransomware attack incidents orchestrated by three distinct APT groups: Conti, Lockbit, and Hive. Despite all three groups sharing the overarching goal of either encrypting or pilfering information, their specific attack methodologies exhibit noteworthy distinctions. Subsequent research initiatives will concentrate on enhancing the efficacy of information extraction from TTPs. To this end, we intend to embark on research endeavors that integrate deep learning models. Furthermore, we propose to employ graph neural networks and knowledge graphs to further analyze the empirical findings, enabling us to proactively predict the potential attack techniques that ransomware organizations might employ. This proactive stance equips us with the ability to respond promptly when ransomware attacks manifest.

REFERENCES

[1] C. Beaman, A. Barkworth, T. D. Akande, S. Hakak, and M. K. Khan, "Ransomware: Recent advances, analysis, challenges and future research directions," *Comput. Secur.*, vol. 111, Dec. 2021, Art. no. 102490.

[2] M. Almousa, S. Basavaraju, and M. Anwar, "API-based ransomware detection using machine learning-based threat detection models," in *Proc. 18th Int. Conf. Privacy, Secur. Trust (PST)*, Dec. 2021, pp. 1–7.

[3] T. McIntosh, A. S. M. Kayes, Y.-P.-P. Chen, A. Ng, and P. Watters, "Ransomware mitigation in the modern era: A comprehensive review, research challenges, and future directions," *ACM Comput. Surv.*, vol. 54, no. 9, pp. 1–36, Dec. 2022.

[4] D. Braue. *Global Ransomware Damage Costs Predicted to Exceed $265 Billion By 2031*. Accessed: Mar. 10, 2023. [Online]. Available: https://cybersecurityventures.com/global-ransomware-damage-costs-predicted-to-reach-250-billion-usd-by-2031

[5] L. Tung. *Kaseya Ransomware Attack: 1,500 Companies Affected, Company Confirms*. Accessed: Mar. 10, 2023. [Online]. Available: https://www.zdnet.com/article/kaseya-ransomware-attack-1500-companies-affected-company-confirms

[6] L. Abrams. *Coop Supermarket Closes 500 Stores After Kaseya Ransomware Attack*. Accessed: Mar. 10, 2023. [Online]. Available: https://www.bleepingcomputer.com/news/security/coop-supermarket-closes-500-stores-after-kaseya-ransomware-attack

[7] Mitre. *MITRE ATT&CK MITRE ATT&CK*. Accessed: Mar. 10, 2023. [Online]. Available: https://attack.mitre.org

[8] H. Oz, A. Aris, A. Levi, and A. S. Uluagac, "A survey on ransomware: Evolution, taxonomy, and defense solutions," *ACM Comput. Surv.*, vol. 54, no. 11s, pp. 1–37, Jan. 2022.

[9] J. Modi, I. Traore, A. Ghaleb, K. Ganame, and S. Ahmed, "Detecting ransomware in encrypted web traffic," in *Proc. Int. Symp. Found. Pract. Secur.* Cham, Switzerland: Springer, 2019, pp. 345–353.

[10] H. Zuhair, A. Selamat, and O. Krejcar, "A multi-tier streaming analytics model of 0-day ransomware detection using machine learning," *Appl. Sci.*, vol. 10, no. 9, p. 3210, May 2020.

[11] J. Bleiberg and E. Tucker. *Texas Ransomware Attack Shows What can Happen When Whole Towns are Targeted*. Accessed: Mar. 10, 2023. [Online]. Available: https://www.usatoday.com/story/tech/news/2021/07/26/texas-ransomware-attack-impact-cyberattack-cybersecurity-small-town-america/8090316002

[12] Secure Reading. *Data of Southwire Company Leaked by Maze Ransomware*. Accessed: Mar. 10, 2023. [Online]. Available: https://securereading.com/data-of-southwire-company-leaked-by-maze-ransomware

[13] J. Ji. *The New Trend of Ransomware: Triple Extortion*. Accessed: Mar. 10, 2023. [Online]. Available: https://nsfocusglobal.com/the-new-trend-of-ransomware-triple-extortion

[14] M. Masum, M. J. H. Faruk, H. Shahriar, K. Qian, D. Lo, and M. I. Adnan, "Ransomware classification and detection with machine learning algorithms," in *Proc. IEEE 12th Annu. Comput. Commun. Workshop Conf. (CCWC)*, Jan. 2022, pp. 0316–0322.

[15] S. Sharmeen, Y. A. Ahmed, S. Huda, B. Ş. Koçer, and M. M. Hassan, "Avoiding future digital extortion through robust protection against ransomware threats using deep learning based adaptive approaches," *IEEE Access*, vol. 8, pp. 24522–24534, 2020.

[16] B. Zhang, W. Xiao, X. Xiao, A. K. Sangaiah, W. Zhang, and J. Zhang, "Ransomware classification using patch-based CNN and self-attention network on embedded N-grams of opcodes," *Future Gener. Comput. Syst.*, vol. 110, pp. 708–720, Sep. 2020.

[17] B. Jethva, I. Traoré, A. Ghaleb, K. Ganame, and S. Ahmed, "Multi-layer ransomware detection using grouped registry key operations, file entropy and file signature monitoring," *J. Comput. Secur.*, vol. 28, no. 3, pp. 337–373, Apr. 2020.

[18] Y. A. Ahmed, B. Kocer, and B. A. S. Al-Rimy, "Automated analysis approach for the detection of high survivable ransomware," *KSII Trans. Internet Inf. Syst.*, vol. 14, no. 5, pp. 2236–2257, 2020.
[19] A. Alzahrani, H. Alshahrani, A. Alshehri, and H. Fu, "An intelligent behavior-based ransomware detection system for Android platform," in *Proc. 1st IEEE Int. Conf. Trust, Privacy Secur. Intell. Syst. Appl. (TPS-ISA)*, Dec. 2019, pp. 28–35.
[20] M. Scalas, D. Maiorca, F. Mercaldo, C. A. Visaggio, F. Martinelli, and G. Giacinto, "On the effectiveness of system API-related information for Android ransomware detection," *Comput. Secur.*, vol. 86, pp. 168–182, Sep. 2019.
[21] H. Faris, M. Habib, I. Almomani, M. Eshtay, and I. Aljarah, "Optimizing extreme learning machines using chains of salps for efficient Android ransomware detection," *Appl. Sci.*, vol. 10, no. 11, p. 3706, May 2020.
[22] M. Al-Hawawreh and E. Sitnikova, "Industrial Internet of Things based ransomware detection using stacked variational neural network," in *Proc. 3rd Int. Conf. Big Data Internet Things*, Aug. 2019, pp. 126–130.
[23] H. Zhang, G. Shen, C. Guo, Y. Cui, and C. Jiang, "EX-action: Automatically extracting threat actions from cyber threat intelligence report based on multimodal learning," *Secur. Commun. Netw.*, vol. 2021, pp. 1–12, May 2021.
[24] H. Kim and H. Kim, "Comparative experiment on TTP classification with class imbalance using oversampling from CTI dataset," *Secur. Commun. Netw.*, vol. 2022, pp. 1–11, Oct. 2022.
[25] P. Panagiotou, C. Iliou, K. Apostolou, T. Tsikrika, S. Vrochidis, P. Chatzimisios, and I. Kompatsiaris, "Towards selecting informative content for cyber threat intelligence," in *Proc. IEEE Int. Conf. Cyber Secur. Resilience (CSR)*, Jul. 2021, pp. 354–359.
[26] C. Liu, J. Wang, and X. Chen, "Threat intelligence ATT&CK extraction based on the attention transformer hierarchical recurrent neural network," *Appl. Soft Comput.*, vol. 122, Jun. 2022, Art. no. 108826.
[27] Y. You, J. Jiang, Z. Jiang, P. Yang, B. Liu, H. Feng, X. Wang, and N. Li, "TIM: Threat context-enhanced TTP intelligence mining on unstructured threat data," *Cybersecurity*, vol. 5, no. 1, pp. 1–17, Dec. 2022.
[28] Y. Shin, K. Kim, J. J. Lee, and K. Lee, "Focusing on the weakest link: A similarity analysis on phishing campaigns based on the ATT&CK matrix," *Secur. Commun. Netw.*, vol. 2022, pp. 1–12, Apr. 2022.
[29] *AlienVault OTX*. Accessed: Mar. 12, 2023. [Online]. Available: https://otx.alienvault.com/dashboard/new
[30] *The DFIR Report*. Accessed: Mar. 12, 2023. [Online]. Available: https://thedfirreport.com
[31] T. Gao, X. Yao, and D. Chen, "SimCSE: Simple contrastive learning of sentence embeddings," 2021, *arXiv:2104.08821*.
[32] Y. Jiang, L. Zhang, and W. Wang, "Improved universal sentence embeddings with prompt-based contrastive learning and energy-based learning," in *Proc. Findings Assoc. Comput. Linguistics, EMNLP*, 2022, pp. 3021–3035.
[33] Y. Januzaj and A. Luma, "Cosine similarity—A computing approach to match similarity between higher education programs and job market demands based on maximum number of common words," *Int. J. Emerg. Technol. Learn.*, vol. 17, no. 12, pp. 258–268, Jun. 2022.
[34] P. E. Kaloroumakis and M. J. Smith, "Toward a knowledge graph of cybersecurity countermeasures," The MITRE Corp., Tech. Rep., 2021, p. 11.
[35] Github. Accessed: Mar. 15, 2023. [Online]. Available: https://github.com
[36] *GitLab*. Accessed: Mar. 15, 2023. [Online]. Available: https://about.gitlab.com
[37] *Bitbucket*. Accessed: Mar. 15, 2023. [Online]. Available: https://bitbucket.org
[38] P. Ranade, A. Piplai, S. Mittal, A. Joshi, and T. Finin, "Generating fake cyber threat intelligence using transformer-based models," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2021, pp. 1–9.

ZHEYU SONG is currently pursuing the master's degree in cyberspace security with the Inner Mongolia University of Technology. His research interests include cyber threat intelligence, APT attack, and deep learning.

YONGHONG TIAN received the B.S. and master's degrees in engineering from the Inner Mongolia University of Technology, in 1998 and 2002, respectively. He is currently a Professor with the Inner Mongolia University of Technology. His research interests include cyber threat intelligence, natural language processing, and deep learning.

JUNJIN ZHANG is currently pursuing the master's degree in software engineering with the Inner Mongolia University of Technology. Her research interests include natural language processing, deep learning, and cyber threat intelligence.