Application Scenarios of Using Knowledge Graph

Subjects: Computer Science, Artificial Intelligence

Contributor: Kai Liu, Fei Wang, Zhaoyun Ding, Sheng Liang, Zhengfei Yu, Yun Zhou

In dynamic complex cyber environments, Cyber Threat Intelligence (CTI) and the risk of cyberattacks are both increasing. This means that organizations need to have a strong understanding of both their internal CTI and their external CTI. The potential for cybersecurity knowledge graphs is evident in their ability to aggregate and represent knowledge about cyber threats, as well as their ability to manage and reason with that knowledge. While most existing research has focused on how to create a full knowledge graph, how to utilize the knowledge graph to tackle real-world industrial difficulties in cyberattack and defense situations is still unclear.

Keywords: application scenarios ; Knowledge Graph ; Cybersecurity

1. Introduction

Knowledge graph technology has sparked a lot of research interest in recent years, thanks to its introduction by Google. In the cybersecurity domain, research on KG can be divided into two categories, study of the KG construction techniques and research of applications. Construction technique studies concentrate on the information extraction technologies, knowledge representation methods, knowledge fusion technologies, and knowledge reasoning methods in graphs [1], such as accurately attaching entities and relationships to the KG after extracting them from the textual corpus and reasoning some new triplets out of such KG.

2. Situation Awareness and Security Assessment

2.1. Security Awareness

As the integration of equipment and services deployed on company networks has gotten increasingly intricate, assessing the overall security posture of an internal network as well as understanding its situation has become a demanding undertaking for human administrators. These multi-stage and multi-host attack scenarios must be addressed by an enterprise network security administrator. CSKG can play a vital role in situation awareness and security assessment. MITRE proposed a situation awareness system CyGraph [2], which is primarily oriented toward network warfare task analysis, visual analysis, as well as the management of cyber knowledge. CyGraph, as a KG with four layers: network infrastructure, security posture, cybersecurity threats, and mission readiness, aims to analyze attack paths, predict critical vulnerabilities, analyze the intrusion alarm, and analyze visual queries by bringing separated data and log events into an ongoing total picture. CyGraph provides some query-driven demonstration cases for stating its effectiveness but no corresponding datasets. Chen [3] combined an existing indicator system and situation detection model and proposed an attack situation detection scheme based on KG. It provides a novel feature for improving the accuracy of cybersecurity situation detection by abstracting the attack events (e.g., historical events, Internet news) as graph description. Wang [4] proposed a KG-based Network Security Situation Awareness model (KG-NSSA) to address the two classic problems: network attack scenario discovery and situation understanding. Different from the traditional alert-based attack scenario discovery approach, which is susceptible to a high level of redundancy and false positives, the scheme can effectively reflect the network attack scenario in the asset node situation with similarity estimation and the attribute graph mining method. Yi et al. [5] utilized the domain knowledge-based reasoning method to realize multi-source intelligence automatic correlation analysis and understand the satellite cyber situation.

2.2. Security Assessment

Wu et al. [6] proposed a novel ontology as well as a graph-based methodology for the task of security assessment. The ontology, which may be instantiated for specific networks, is designed to standardize the representation of security knowledge, such as assets, vulnerabilities, and attacks. Using the ontological model's inference capabilities, an efficient framework for generating attack graphs, identifying the possible attacks caused by published vulnerabilities, and assessing network security is proposed. The attack graph is a graphical representation and explanation of the attributes defined that are included in the method’s final output. This included an attack process flow diagram to clearly depict how an attacker might invade and compromise numerous goals of a test network across multiple hosts and multiple stages. This facilitates the enterprise security manager to complete security risk assessment tasks as well as respond to new threats. Inspired by the existing cybersecurity ontologies, such as STUCCO [7], UCO [8], and Cyber Intelligence Ontology [9], Kiesling et al. [10] proposed a publicly available CSKG with concrete instance information and illustrated its applicability for security assessment by two SPARQL [11] query example scenarios. It will query which data assets might be exposed in the local system model based on matching the information of organization-specific assets to a continually updated stream of existing vulnerabilities in order to estimate the possible effect of a newly discovered vulnerability. Pang et al. [12] proposed a security evaluation methodology of power IOT terminals based on KG with three dimensions: terminal assets, vulnerability, and intrusion alert, based on application scenarios as well as the power IOT terminal’s threat characteristics. The approach performs a correlation analysis of the cybersecurity monitoring information of the IoT terminal with independent power and provides a terminal threat index that reflects the security condition of the terminal. This should have been a good attempt, unfortunately, the research did not give the details of the KG and method.

3. Threats Discovery

3.1. Attack Prediction

Narayanan et al. [13] developed a cognitive system that combines input from conventional sensors, dynamic internet textual sources, and KGs to identify cybersecurity issues early. The researchers extended UCO so that it can reason over inputs from multiple network sensors, such as intrusion detection systems (IDS), Snort, and so on, as well as the knowledge from the cyber-kill chain. To express rules between entities, the Semantic Web Rule Language (SWRL) was utilized. The aggregator module was designed to combine alerts into a reasoning model. They proved its ability to identify newer attacks by putting it to the test against custom-built ransomware akin to WannaCry and displaying the timeline of the attack as well as the system’s response activities. Unfortunately, this solely described the system’s architecture and did not include any additional information or data. Sun et al. [14] suggested a prediction approach of a 0-day attack route based on a cyber defense KG to address the challenge of attack prediction induced by the 0-day vulnerability. The KG was generated from three aspects (i.e., threat, assets, and vulnerability), which supported transforming the task of attack prediction into a KG link prediction problem. A path ranking algorithm was used to create the 0-day attack graph and discover the possible 0-day attack of the target system, according to the above methodology. The experimental results revealed that the suggested strategy might increase the accuracy of 0-day attack prediction with the aid of KG. Furthermore, employing the path ranking algorithm can aid in tracing the causes of predicted outcomes in order to increase the explanatory ability to forecast.

3.2. Threat Hunting

For the task of cyber threat hunting, a system was developed by Gao et al. [15], aiming at facilitating log-based cyber threat hunting by leveraging vast external threat knowledge provided by OSCTI [16]. The system is composed of two subsystems, i.e., a pipeline IE model for building threat behavior KG and a querying subsystem based on system auditing, which can collect audit logging data across hosts. This also includes a threat behavior query language (TBQL) and a query synthesis technique that automatically synthesizes a TBQL query based on the threat behavior KG with event sequence information, which could be used to discover matched system auditing records. Nevertheless, one of the limitations of this system is that it does not consider the attacks that are not detected by the system auditing. Furthermore, existing methods frequently have significant limitations in terms of the interpretability, quantity, and relevancy of the warnings issued.

3.3. Intrusion Detection

Besides the intrusion detection methods mentioned by [17], the CSKG could also be constructive in detecting intrusion. Kiesling et al. [10] gave a query-based case to demonstrate how network intrusion detection system (NIDS) alerts may be linked to the SEPSES CSKG to gain a better knowledge of possible threats and current assaults. Chen et al. [18] proposed a detection method of a DDoS attack based upon a domain KG, which is mainly aimed at DDoS attacks on TCP traffic. The KG is used to express the communication process of TCP traffic between two hosts. Together with calculating the value for one-way transmission propensity, a threshold was set to determine whether the source host is the initiator of a DDoS attack. To comprehensively describe the DDoS attack, Liu et al. [19] constructed a DDoS attack malicious behavior knowledge base, which contains two parts: a malicious traffic detection database and a network security knowledge base. The front one is responsible for detecting and classifying malicious traffic generated by DDoS attacks. The network security knowledge base, which includes the network topology graph, malicious behavior traceability graph, malicious behavior feature graph, and traffic behavior KG, is at the heart of the malicious behavior knowledge base of DDoS attacks. Data structure processing, malicious behavior KG creation, behavior reasoning, and feedback are all handled by the network security knowledge base. In 2021, Garrido et al. [20] applied a machine learning method to KGs to identify unusual behaviors in industrial automation systems integrating IT and OT elements. Using a readily available ontology [21], this builds a KG by combining three major sources of knowledge: automation system information, application-level observations (e.g., data access events), and network observations (e.g., connections between hosts). Inspired by KG completion methods, this adopts a graph embedding algorithm to rate the likelihood of triple assertions emerging from observed security events. Experimentally, the suggested method produces intuitively well-calibrated and interpretable alarms in a variety of contexts, pointing to the relational machine learning potential benefits on KG for the task of intrusion detection. Although the results are generated on a reduced-scale prototype and without the help of CTI, the present research explores, for the first time, the synergistic combination of KG and industrial control systems.

4. Attack Investigation

4.1. Attack Path Analysis

As mentioned above, the CyGraph could query out potential attack paths based on the network environment. Similar to CyGraph, Neol et al. [22] illustrated a graph-based strategy with a unique model of attack graph that mixes a complicated mix of network data, such as the network topology, firewall strategies, vulnerability intelligence, attack patterns, and threat alerts, via cybersecurity data standardized languages. Furthermore, the researchers, created a model that predicts possible attack paths based on network events (intrusion alerts, sensor logs, etc.). Correlating observed attack events with prospective attack pathways provides the optimum reaction choices, particularly for safeguarding important assets, and enhances situational awareness, such as missed attacking steps inferring and false positive minimizing. The attack graph that results is stored in the Neo4j database [23] for query and visualization. Despite its efficiency of query and visualization for potential attack paths, the proposed KG still faces several disadvantages. First, this did not show people how to use KG to infer new knowledge. Secondly, the corresponding datasets, such as the input format of alerts data and firewall policies, were not demonstrated clearly. Finally, the OSCTI (the knowledge provided by METRE) is not reflected in the architecture.

To extend the information on the attack path, Ye et al. [24] designed a cyber attack KG with four types of entities, including software, hardware, vulnerabilities, and attack entity. With the help of four kinds of attributions of attack entity (i.e., attack conditions, attack methods, success rate, and earnings), this used KG to generate an attack path and improve the assessment of vulnerability rather than rely on the CVSS score [25]. Thanks to the knowledge representation and information management ability of KG, the attack path could update local information based on multiple sources. To further increase efficiency, a graph-based strategy for determining the ideal penetration path is proposed, taking into account insider and unknown attacks. Wang et al. [26] defined a two-layer threat penetration graph (TLTPG), where the upper layer is a penetration graph of the network environment, and the lower layer is a penetration graph between any two hosts. The KG was used to describe the attack-related resources (e.g., software, vulnerability, ports in use, and privilege of a successful attack) of each host, which would be of great benefit not only to generate the penetration path between hosts but also to integrate collected information of 0-day vulnerability attack for unknown attack prediction. In the power networks, Chen et al. [27] generated an expansion attack graph for obtaining the maximum probability vulnerability path and providing the success rate of attacking the power grid and the loss.

4.2. Attack Attribution

This step is known as attack attribution. The attack source, intermediate medium, and corresponding attack path can all be determined using attack attribution technologies, allowing for more tailored protection and countermeasure techniques to achieve active defense. As can be seen, the attribution of attacks is a crucial step in the transition from passive to active defense. Based on an ontology with six dimensions, namely host asset, vulnerability, attack threat, evidence, location, strategy, and the relationships between them, Zhu et al. [28] constructed a CSKG for the network of space-ground integration information. In addition, each dimension has several unique attributes and data sources. The research proposed an automated cyberattack attributing framework. Attack attribution could be performed from several aspects by querying the established CSKG. As an example, given in the research, based on the host asset level, security employees could query the KG to find out the vulnerable host asset suspected of being attacked, associated vulnerability, and attribution strategy in sequence. Then, by implementing the corresponding attribution strategy, the attacker’s evidence and locations could be identified, and the attacked host asset can be determined. Xue et al. [29] analyzed the existing provenance graph construction technology based on causation in NSFOCUS blogs. The study introduced the provenance graph construction from three dimensions, including the terminal dimension, the perspective of Syslog and application log correlation, and the association of network and terminal. The terminal perspective method focused on the relationships between processes, files, and filenames in a single isolated host and ignored the application log, which was replenished by the second dimension. Moreover, the third level method extended the provenance graph from a single-host to a multi-host network, which could be further enhanced through causal analysis to gett a complete attack process. However, this did not consider the semantic context and OSCTI provided by CSKG.

4.3. Consequence Prediction

Common software weaknesses, such as poor input validation and integer overflow, can directly or indirectly impair system security, resulting in negative effects such as denial-of-service (DOS) and unauthorized code execution. Understanding the consequence of weakness becomes significant to assessing the risk of a system and to take prompt response. In 2018, Han et al. [30] built a KG based on common weakness enumeration (CWE) [31], which provides detailed information regarding vulnerabilities such as the textual descriptions, relationships between software weaknesses, and common effects. The available CWE data do not allow sophisticated reasoning tasks, such as missing links prediction and the prediction of common consequences. This developed a description-embodied, translation-based knowledge representation learning approach for embedding both the weaknesses and their relationships into a semantic vector space. Following the vector embedding generation, extensive experiments were conducted to estimate the performance of KG in knowledge acquisition and inference tasks. CSKG could be exploited for three different reasoning tasks: link prediction task based on CWE, CWE triple categorization task, and threat consequence prediction task. Datta et al. [32] transferred the consequence prediction problem to the classification task by introducing a dataset and building machine learning models and natural language processing (NLP) models. The cyberattack dataset includes 93 different assaults and their descriptions, which are annotated with technical and nontechnical consequences. The goal is to provide security researchers with tools that make it simpler to convey the consequences of an attack to diverse stakeholders who may have little to no cybersecurity experience. Furthermore, the suggested technique can lessen researchers' cognitive strain by automatically forecasting the consequences when new attacks are identified.

4.4. Attack Analysis

Besides the above application, Qi et al. [33] built a CSKG, which includes two subgraphs: CSKG and scene KG, for attack analysis. The CSKG is the core graph representing the knowledge about vulnerabilities, attacks, assets, and the relationships among them, which can be collected from various websites publishing vulnerability and attack analysis intelligence and can be updated gradually. Scene KG is an extended graph constructed using node and connectivity information of the network involved in a specific attack. The data gathering system and detection system provide the input data for the whole analytical framework. Composite attack chains are generated from several single attacks using the CSKG, attack rule base, and spatiotemporal property restrictions. With so many alerts to analyze, cyber investigators frequently suffer from alert fatigue, leading them to disregard a high number of alerts and overlook real attacks. It has been shown that distinct attacks, independent of the vulnerabilities exploited or payloads performed, may use identical abstract tactics. Alsaheel et al. [34] presented ATLAS, a methodology that generates an end-to-end attack scenario from ready-made audit logs using a causal graph. ATLAS employs a revolutionary mix of NLP, causality analysis, and machine learning approaches to construct a sequence-based framework that extracts essential patterns of attack and nonattack behaviors from such a causal graph. Given a security event, an attack symptom node in the causal graph is generated at the inference phase. ATLAS then builds a collection of possible sequences linked with the symptom node, employs the sequence-based framework to identify nodes that contribute to the attack, and combines the discovered attack nodes to generate an attack story.

6G-oriented network intelligence requires knowledge from both within and outside the network. Therefore, Wang et al. [35] proposed a method for building cyberattack KGs based on CWE as well as CAPEC [36], which is deployed in Neo4j (a graph database). This only introduced two query-based application scenarios in detecting and responding to DDoS flood attacks and multi-stage attacks based on Neo4j's query and display functionality, rather than based on the reasoning function of KG. This just focused on the analysis and application with CVE and CAPEC, which did not cover the complete knowledge of cybersecurity.

5. Intelligent Security Operation

5.1. Intelligent Operation

An AI-driven security operations framework was presented by Zhang et al. [37]. CSKG could support dynamic query and aggregation analysis of security data, improving the integrity of security data operation analysis. The KG is a unified data view, which can support the realization of multi-level technical capabilities such as subsequent risk perception, causal cognition, and robust decision-making. Some challenges were discussed from the aspects of data, models, and semantic context. However, this did not demonstrate the specific method.

5.2. Security Alert or Event Correlation Analysis

Given the ever-changing threat landscape, cybersecurity researchers overseeing the Security Operation Center (SOC) are frequently overburdened with various security events while also attempting to stay up with the most recent threats in the field. Effectively analyzing vast amounts of various alerts or event data brings opportunities to detect concerns before they become problems and to avoid further cyberattacks. Conventional techniques frequently store the many aspects of security information in distinct databases, which results in the absence of synergies between the multiple dimensions. As illustrated by Xue [38], the main challenge faced by the application of cyber CSKG is that there is no direct connection between the KG based on abstract attack knowledge such as STIX 2.0 and the system and network logs that contain the behavior information. It is a semantic gap between them. For complicated cyberattacks, it is difficult to incorporate all context information quickly to initiate real-time and accurate analysis. To create the attack scene, traditional rule-based association analysis relies on expert knowledge, which lacks the capacity to reason automatically.

To address the aforementioned issue, Wang et al. [39] presented an integrated correlation analysis approach to a cybersecurity event. The approach included the vulnerability KG, threat intelligence KG, the network infrastructure KG, and intrusion alert KG into the CSKG, as well as documented the data sources for each dimension. Following alert normalization and alert fusion, the alert verification was conducted by judging whether the vulnerabilities of one alert are in the host vulnerability set. Furthermore, the attack thread correlation analysis process relies on the existing alerts to query the associated alerts, CVE items, and CAPEC items, which could be conducive to predicting the real purpose of attackers. In the researchers' thesis [40], rebuilding the scene of a series of alerts based on KG was introduced in detail. The researchers conducts an experiment on the DARPA 2000 dataset to assess the performance of the proposed framework by comparing the number of remaining alerts after correlation analysis. This showed an example of the use of KG for correlation analysis. Qi et al. [41] believed that cyberattacks involve various attack phases that are related to IDS alarms. Based on this thought, an association analysis model developed on cybersecurity attack events KG is presented to display a cyberattack scenario in a special air–ground integrated network graphically. The CSKG includes five tuples: attacks, alarms, events, relations, and the rules. The association analysis was used by calculating the coincidence degree between the gathered events sequence and the attacked events sequence in the KG.

However, due to the absence of a thorough knowledge of the integrated space–ground network as well as the limits of the present experimental settings, this relied solely on simulation tests to validate the viability of the aforementioned approach. Manual analysis of logs often does not scale well and frequently results in a lack of knowledge and insufficient transparency about concerns. To address this issue, Ekelhart et al. [42] introduced a flexible framework for the automated construction of KGs from arbitrary raw log messages. The method closes a key gap and offers up a variety of data sources for KG construction by making the log data suitable for semantic analysis.

6. Intelligent Decision-Making

6.1. Generation of Attack Strategy

Analyzing the attack strategy from the attacker's point of view can assist in identifying current security issues and can give targeted protective recommendations. Compared with the query-based method of CyGraph, a cyberattack method recommendation model that relied upon KG was proposed by Ou et al. [43]. It contains a six-tuple KG construction schema based on four open databases (i.e., CVE, CWE, MSF, CAPEC), the collaborative filtering recommendation that describes difference relationships between nodes by meta-path, a generator of recommendation list with calculating the correlation score of each path with node vector. In the second part, a recommendation algorithm for cyberattack entities is proposed by combining the method of machine learning feature extraction and the method of constructing a heterogeneous information network meta-path. Based on this KG, intelligent searching and recommendations of knowledge related to new threat intelligence can be achieved. Compared with the traditional content-based search recommendation method, this method is more accurate at predicting the weakness of vulnerabilities and can realize the prediction and recommendation of attack patterns based on the natural language description of vulnerabilities. Likewise, from the perspective of an attacker, Chen et al. [44] proposed a method for generating a knowledge-driven attack strategy in order to exploit various vulnerabilities in an industrial control network. The method consists of a vulnerability exploitation KG, an industrial control network graphical representation, and the reasoning rules rely on KG. It is a common idea among cybersecurity experts that look for attack paths at the device level depending on the attack process. The attack strategy formulation process can be partitioned into two steps: The first step is analyzing the current device-level nodes' several vulnerabilities and correlating them to the consequences and pre-conditions of exploitation. To devise a global attack strategy graphically, the research linked the device-level nodes depending on firewall access rules as well as other protection devices after formulating the sequence in which all device vulnerabilities are exploited. Currently, this proposed KG is applied to analyze numerous vulnerabilities on a small-scale industrial control network to generate attack paths. With the expansion of KG, such as supplementing with other threat intelligence, more and more attack strategies need to be generated, especially the most cost-effective attack strategy.

6.2. Security Policy Validation

Vassilev et al. [45] proposed a four-layer (i.e., ontological level, heuristic level, workflow level, and process-level) architecture for CTI analysis, logical analysis, and validation of security policies. The architecture has been verified using a collection of scenarios depicting digital banking's most frequent security risks, and a prototype of an event-driven engine for traversing intelligence graphs has been constructed. However, this framework was created particularly for use in digital banking and did not include any previously used datasets.

7. Vulnerability Management and Prediction

KG technologies provide an exciting opportunity to advance the knowledge of managing considerable vulnerability data by presenting them in a structured ontological format. Cybersecurity vulnerability ontology (CVO), a formal, conceptual knowledge representation model in the vulnerability management area, was created by Syed et al. [46]. Additionally, they utilized the CVO to develop a system of cyber intelligence alert (CIA) that sends out threat alerts regarding potential vulnerabilities and countermeasures. At the practical level, its components contain the social media intelligence extractor-tagger (SMIET), the vulnerability repository and mapper, CVO, RDF converter, the cyber intelligence ontology (CIO), and the engine of cyber alerts rules. Finally, this gave the evaluation approaches, corresponding results, and examples in practice. Based on the industrial internet security vulnerabilities, an industrial CSKG was built and stored in Neo4j by Tao et al. [47] in order to analyze, query, and visualize from the temporal, spatial, and correlation dimensions.

When confronted with actual intrusions, CyGraph correlates intrusion alerts to published vulnerability pathways and recommends the appropriate courses of action for reacting to attacks. CyGraph creates a predicted model of likely attack pathways and major vulnerabilities based on queries. As previously stated, by constructing a knowledge representation learning approach (translation-based, description-embodied), a CSKG based on CWE might be utilized to infer incomplete relationships and common effects. To find hidden relationships among weaknesses, Qin et al. [48] proposed a query-based model for analyzing and reasoning new knowledge automatically. The reasoning flow of the sample CWE Chain was demonstrated based on a vulnerability KG (VulKG), which covers the vulnerability data from NVD, CVE, CWE, and CPE. However, the example could merely partially take over the place of the analysis and labeling work of security specialists under some specific scenarios, where the operator needs to know the query target previously. For effectively managing the sparse or inaccurate malware threat information, a malware KG called MalKG was established by Rastogi et al. [49], which is the first open-source automated malware threat intelligence KG. Additionally, there are approximately 40 thousand triples in the provided MalKG dataset (i.e., MT40K), which include 27,354 unique entities and 34 relationships. The study also manually curated a benchmark KG dataset called MT3K, with 5741 unique entities and 22 relationships, forming 3027 triples. It demonstrated the prediction capabilities of MalKG using two use cases in predicting new information. One of the application scenarios is predicting and sorting all the potential vulnerabilities or CVEs of the malware-impacted software system by comprehensive utilization of information from the network environment, malware, and KG. A vulnerability exploitation KG was built by combining and extracting multi-dimensional domain knowledge, as has already been noted in [44]. Attack strategies depending on KG enhance the performance in comprehensive vulnerability exploitation and flexible response by analyzing each device-level node. Based on an industrial network example, the feasibility of the method was investigated. Similarly, in Wang's study [50], chain reasoning and confidence calculation were also used to support vulnerability detection and finding latent relationships between CWEs. Similarity matching based on a source code level graph is used for judging the similarity between the target node and the node in the vulnerability database, which provides new insights into vulnerability mining. Wang et al. [51] extended the relationships in the vulnerability KG by identifying the alternative vulnerability with similar consequences.

8. Malware Attribution and Analysis

Najafi et al. [52] devised MalRank, a graph-based malware rank inference model aimed to predict a node's maliciousness by its associations with the other entities in the KG, such as common IP ranges or dns servers. This presented a KG that builds global relationships among entities detected in proxy and IDS logs, enhanced with related CTI and open-source intelligence (OSINT). The researchers formulate threat detection in the security information and event management

(SIEM) environment as a large-scale graph inference problem. MalRank maintains a high detection rate, beating its predecessor, belief propagation, both in terms of accuracy and efficiency, according to a series of trials using real-world data acquired from a worldwide organization's SIEM. It was also demonstrated that this method is useful in detecting previously discovered hostile entities, including IP addresses and malicious domain names. Besides the application scenarios reported earlier, MalKG [53] could also be implemented in the malware attribution scenario [49]. For example, given a newly discovered malware attack on one system, the analyst needs to build a fingerprint of the malware's origin by assembling sufficient features, campaign, and others. The goal of MalKG is to automate the prediction of these features associated with a given malware; for instance, the newly discovered malware may share similarities with a disclosed malware linked to a certain APT group. As reported in the white paper [54], the profiling and automatic attribution of APT attacker gangs can be realized through the extraction of key elements of threat intelligence and dynamic behavioral reasoning. The key solution lies in establishing a unified language to describe the behavior and characteristics of different APT organizations, as well as in building a knowledge base about APT organizations. However, the white paper did not disclose the details of the related research.

9. Connection to the Physical System

CSKG uses big data analysis and graph mining technology to deeply analyze the coupling relationship between the information layer and the physical layer in the modern industrial control system and to realize the intelligence of "decision making, risk prediction, accident analysis, attack identification" and other capabilities assisted and automated processing.

To illustrate various cybersecurity analytic capabilities in the MITRE's situational awareness system, CyGraph [2][22], they presented a simple network architecture for the case study. The architecture shows the fundamental connectivity among hosts, firewalls, routers, and switches. The internal network is divided into three security domains (mission client workstations, DMZ, and the data center). The internal network is protected from the outside by the external network firewall, while the vital data-center servers are protected by the internal firewall. The KG was created using information from the topology of the network, firewall rules, and vulnerability scan findings. To verify the effectiveness of the proposed method, a typical internal network architecture model with six types of elements is introduced [24]. In this architecture, the firewall isolates the Internet from the intranet router. The FTP server, host1, and host2 are directly connected to the router. Host1 and host2 can access the FTP server. The database server is connected to the FTP server to receive and respond to requests from the FTP server. For the purpose of generating penetration paths, the essay [26] designed an illustrative network example. The network contains a host on the Internet, a DMZ area, and three subnets. There is a web server in the DMZ area. Subnet1 has two devices (i.e., one Pad and a host), which can be connected to the Internet. Subnet2 has two hosts and cannot connect to the Internet. Subnet3 includes three servers, including a print server, file server, and data server. The attacker is a host on the Internet. It also considered the potential connection between subnet1 and subnet2 via USB. Similar to the architecture above, an experimental network environment was designed in the research [14]. The experimental network contains two subnets protected by two firewalls separately and one DMZ also with a firewall. The two subnets connect to the Internet through the DMZ. Each part of the network has different assets, such as an email server and a web server; in the DMZ, two hosts and a file server are part of subnet1, and an application server is connected to subnet2. Building a suitable experiment network environment could be beneficial to demonstrate the effect of approaches and reproduce the attack and defensive process.

Despite the various traditional network architectures, it is also important to research the security of industrial control systems with a suitable experimental network. In the research of [20], a hardware prototype was described for evaluation based on the architecture of current industrial systems merging IT and OT aspects. A Siemens S7-1500 PLC is used for the automation side, and it is linked to peripherals through an industrial network. A conveyor belt driving subsystem, a human–machine interface (HMI), an industrial camera, and a distributed I/O subsystem with modules interfacing with different sensors for object location and other measurements are among these peripherals. Through an OPC-UA server, the PLC provides the values recorded by these sensors and the system state information. Therefore, the PLC connects to two edge computing servers. Thereafter, the network with main traffic flows was also displayed. A KG-based security assessment technique for power IoT terminals is provided [12] in order to perceive and measure enormous power IoT terminals' security risks and threats in real-time. However, this did not describe a suitable network for evaluation. As analyzed previously, Chen et al. [44] used the domain KG to produce attack strategies by analyzing several vulnerabilities in the industrial control system. The topology of the target network is composed of the Internet, two firewalls, one router, an enterprise network, and an industrial ethernet. One firewall is used to protect all assets of the local network. The other one is situated between the enterprise network and industrial ethernet. The route is between the first firewall and the enterprise network, followed by the second firewall and industrial ethernet. The assets of the enterprise network include a web server, admin host, and printer. Some peripherals, such as an HMI, a data server, a workstation, and three PLCs with different end-effector devices (e.g., valve, flowmeter), are connected to the industrial ethernet. The attacker is a certain host from the Internet, and the PLCs are the attack target.

Based on the above analysis and related research, this sorts out a general experimental network architecture to demonstrate the effect of potential security investigation approaches. This general network mainly contains four parts, including DMZ, a subnet connected to the Internet via a router, a subnet connected to DMZ, and an industrial control network connected to DMZ. Each subnet is isolated by a firewall, and the attackers usually start their offensive action from the Internet. A researcher could utilize it to adapt to a complex network by modifying or adding some devices, extending the subnets, or changing the connection mechanism. The network topology expresses the network environment. In addition, it should also include the software and hardware installed on each node, security protection measures, and existing vulnerabilities.

10. Two New Applications

10.1. Social Engineering

Social engineering, in short, is a sort of cyberattack where the attacker takes advantage of human vulnerability by engaging in social interactions to break cybersecurity [55]. Cyberspace security has been seriously compromised by social engineering. A social engineering ontology in cybersecurity is developed by Wang et al. [56] towards protecting social engineering cyberattacks, as well as a method for evaluating it by its applications. The ontology defines eleven essential entity concepts, as well as 22 types of relationships, which significantly comprise or impact the social engineering area. It offers a structured and explicit knowledge architecture for understanding, analyzing, reusing, and sharing social engineering field knowledge. The KG was also created using fifteen social engineering attack events and scenarios, and it was comprehensively assessed using seven application examples (in 6 analysis patterns) based on query methods.

10.2. Combating Fake Intelligence

Internet users today receive a considerable volume of fake cybersecurity intelligence. In order to get rid of this kind of information, Mitra et al. [57] built a system that captures provenance information and displays it together with the CTI captured. Together with enhancing the exiting CSKG model to combine intelligence provenance, this fused provenance graphs with CSKG. The reasoning capabilities of CSKG enforce rules that aid in the preservation of reliable information while discarding the rest. Moreover, classes that capture provenance can be added to the CSKG schema, providing people with more information about the data’s source. However, the details and datasets of this novel KG were not given.

Apart from that, Xiao et al. [58] developed a KG embedding method to predict software security entity within-type and across-type interactions. Analysts can expand their understanding of software security by discovering such missing connections between existing entities. However, this CSKG is not open-source, so researchers could not read the details of it. In addition, the mentioned white paper [54] reported several other application scenarios of CSKG technology as well as its two classical reasoning methods. Despite the fact that there was some limitation in stating the adequate details, the application scenarios, such as ATT&CK threat modeling, APT threat hunting, intelligent security operation, cyberspace surveying and mapping, supply chain security, and cyber-physical system protection, were outlined and forecasted by the white paper. There are two broad categories of reasoning technologies based on CSKG: relational reasoning based on graph representation learning and multi-relational reasoning methods based on neural networks.

References

1. Zenglin, X.; Yongpan, S.; Lirong, H.; Yafang, W. Review on knowledge graph techniques. J. Univ. Electron. Sci. Technol. China 2016, 45, 589–606.
2. Noel, S.; Harley, E.; Tam, K.H.; Limiero, M.; Share, M. CyGraph: Graph-based analytics and visualization for cybersecurity. In Handbook of Statistics; Elsevier: Amsterdam, The Netherlands, 2016; Volume 35, pp. 117–167.
3. Chen, X. Design and Implementation of Network Attack Situation Detection System Based on Knowledge Graph. Master’s Thesis, Beijing University of Posts and Telecommunications, Beijing, China, 2020.
4. Wang, Y. Research and Implementation of NSSA Technology Based on Knowledge Graph. Master’s Thesis, University of Electronic Science and Technology of China, Chengdu, China, 2020.
5. Wang, B.; Wu, L.; Hu, X.; He, Y. Satellite cyber situational understanding based on knowledge reasoning. Syst. Eng. Electron. 2022, 44, 1562–1571.

6. Wu, S.; Zhang, Y.; Cao, W. Network security assessment using a semantic reasoning and graph based approach. Comput. Electr. Eng. 2017, 64, 96–109.

7. Iannacone, M.; Bohn, S.; Nakamura, G.; Gerth, J.; Huffer, K.; Bridges, R.; Ferragut, E.; Goodall, J. Developing an ontology for cyber security knowledge graphs. In Proceedings of the 10th Annual Cyber and Information Security Research Conference, Oak Ridge, TN, USA, 7–9 April 2015; pp. 1–4.

8. Syed, Z.; Padia, A.; Finin, T.; Mathews, L.; Joshi, A. UCO: A unified cybersecurity ontology. In Proceedings of the Workshops at the Thirtieth AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–13 February 2016.

9. Philpot, M. Cyber Intelligence Ontology. 2015. Available online: https://github.com/daedafusion/cyber-ontology (accessed on 4 April 2022).

10. Kiesling, E.; Ekelhart, A.; Kurniawan, K.; Ekaputra, F. The SEPSES knowledge graph: An integrated resource for cybersecurity. In Proceedings of the International Semantic Web Conference, Auckland, New Zealand, 26–30 October 2019; Springer: New York, NY, USA, 2019; pp. 198–214.

11. SPARQL. Virtuoso SPARQL Query Editor. 2022. Available online: https://w3id.org/sepses/sparql (accessed on 4 April 2022).

12. Pang, T.; Song, Y.; Shen, Q. Research on security threat assessment for power iot terminal based on knowledge graph. In Proceedings of the 2021 IEEE 5th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Xi’an, China, 15–17 October 2021; Volume 5, pp. 1717–1721.

13. Narayanan, S.N.; Ganesan, A.; Joshi, K.; Oates, T.; Joshi, A.; Finin, T. Early detection of cybersecurity threats using collaborative cognition. In Proceedings of the 2018 IEEE 4th International Conference on Collaboration and Internet Computing (CIC), Philadelphia, PA, USA, 18–20 October 2018; pp. 354–363.

14. Sun, C.; Hu, H.; Yang, Y.; Zhang, H. Prediction method of 0 day attack path based on cyber defense knowledge graph. Chin. J. Netw. Inf. Secur. 2022, 8, 151–166.

15. Gao, P.; Shao, F.; Liu, X.; Xiao, X.; Qin, Z.; Xu, F.; Mittal, P.; Kulkarni, S.R.; Song, D. Enabling efficient cyber threat hunting with cyber threat intelligence. In Proceedings of the 2021 IEEE 37th International Conference on Data Engineering (ICDE), Chania, Greece, 19–22 April 2021; pp. 193–204.

16. SENKI. Open Source Threat Intelligence Feeds. 2020. Available online: https://www.senki.org/operators-security-toolkit/open-source-threat-intelligence-feeds/ (accessed on 4 April 2022).

17. Jian, S.; Lu, Z.; Du, D.; Jiang, B.; Li, B. Overview of network intrusion detection technology. J. Inf. Secur. 2020, 5, 96–122.

18. Chen, J. DDoS attack detection based on knowledge graph. J. Inf. Secur. Res. 2020, 6, 91–96.

19. Feiyang, L.; Kun, L.; Fei, S.; Chunhua, Z. Distributed DDoS attacks malicious behavior knowledge base construction. Telecommun. Sci. 2021, 37, 17–32.

20. Garrido, J.S.; Dold, D.; Frank, J. Machine learning on knowledge graphs for context-aware security monitoring. In Proceedings of the 2021 IEEE International Conference on Cyber Security and Resilience (CSR), Rhodes, Greece, 26–28 July 2021; pp. 55–60.

21. Kovalenko, O.; Wimmer, M.; Sabou, M.; Lüder, A.; Ekaputra, F.J.; Biffl, S. Modeling automationml: Semantic web technologies vs. In model-driven engineering. In Proceedings of the 2015 IEEE 20th Conference on Emerging Technologies & Factory Automation (ETFA), Luxembourg, 8–11 September 2015; pp. 1–4.

22. Noel, S.; Harley, E.; Tam, K.H.; Gyor, G. Big-Data Architecture for Cyber Attack Graphs Representing Security Relationships in Nosql Graph Databases. 2015. Available online: https://csis.gmu.edu/noel/pubs/2015_IEEE_HST.pdf (accessed on 4 April 2022).

23. Vukotic, A.; Watt, N.; Abedrabbo, T.; Fox, D.; Partner, J. Neo4j in Action; Manning Publications Co.: Shelter Island, NY, USA, 2015; Volume 22.

24. Ye, Z.; Guo, Y.; Li, T.; Ju, A.K. Extended attack graph generation method based on knowledge graph. Comput. Sci. 2019, 46, 165–173.

25. Ruohonen, J. A look at the time delays in cvss vulnerability scoring. Appl. Comput. Inform. 2017, 15, 129–135.

26. Shuo, W.; Jianhua, W.; Guangming, T.; Qingqi, P.; Yuchen, Z.; Xiaohu, L. Intelligent and efficient method for optimal penetration path generation. J. Comput. Res. Dev. 2019, 56, 929.

27. Chen, Z.; Dong, N.; Zhong, S.; Hou, B.; Chang, J. Research on the power network security vulnerability expansion attack graph based on knowledge map. Inf. Technol. 2022, 46, 30–35.

28. Zhu, Z.; Jiang, R.; Jia, Y.; Xu, J.; Li, A. Cyber security knowledge graph based cyber attack attribution framework for space-ground integration information network. In Proceedings of the 2018 IEEE 18th International Conference on

Communication Technology (ICCT), Chongqing, China, 8–11 October 2018; pp. 870–874.

29. Xue, J. Attack Attribution: Provenance Graph Construction Technology Based on Causation. 2020. Available online: http://blog.nsfocus.net/attack-investigation-0907/ (accessed on 4 April 2022).

30. Han, Z.; Li, X.; Liu, H.; Xing, Z.; Feng, Z. Deepweak: Reasoning common software weaknesses via knowledge graph embedding. In Proceedings of the 2018 IEEE 25th International Conference on Software Analysis, Evolution and Reengineering (SANER), Campobasso, Italy, 20–23 March 2018; pp. 456–466.

31. Mitre. Common Weakness Enumeration. 2022. Available online: https://cwe.mitre.org/ (accessed on 4 April 2022).

32. Datta, P.; Lodinger, N.; Namin, A.S.; Jones, K.S. Cyber-attack consequence prediction. arXiv 2020, arXiv:2012.00648.

33. Qi, Y.; Jiang, R.; Jia, Y.; Li, A. Attack analysis framework for cyber-attack and defense test platform. Electronics 2020, 9, 1413.

34. Alsaheel, A.; Nan, Y.; Ma, S.; Yu, L.; Walkup, G.; Celik, Z.B.; Zhang, X.; Xu, D. ATLAS: A sequence-based learning approach for attack investigation. In Proceedings of the 30th USENIX Security Symposium (USENIX Security 21), Vancouver, BC, Canada, 11–13 August 2021; pp. 3005–3022.

35. Wang, W.; Zhou, H.; Li, K.; Tu, Z.; Liu, F. Cyber-attack behavior knowledge graph based on CAPEC and CWE towards 6G. In Proceedings of the International Symposium on Mobile Internet Security, Jeju Island, Korea, 7–9 October 2021; Springer: New York, NY, USA, 2021; pp. 352–364.

36. MITRE. Common Attack Pattern Enumeration and Classification. 2022. Available online: https://capec.mitre.org/ (accessed on 3 April 2022).

37. Runzi, Z.; Wenmao, L. An intelligent security operation technology system framework AISecOps. Front. Data Domputing 2021, 3, 32–47.

38. Xue, J. Attack Reasoning: Dilemma of Application of Security Knowledge Graph. 2020. Available online: http://blog.nsfocus.net/stucco-cyber/ (accessed on 4 April 2022).

39. Wang, W.; Jiang, R.; Jia, Y.; Li, A.; Chen, Y. KGBIAC: Knowledge graph based intelligent alert correlation framework. In Proceedings of the International Symposium on Cyberspace Safety and Security, Xi’an, China, 23–25 October 2017; Springer: New York, NY, USA, 2017; pp. 523–530.

40. Wang, W. Research for Algorithm of Distributed Security Event Correlation Based on Knowledge Graph. Master’s Thesis, National University of Defense Technology, Changsha, China, 2018.

41. Qi, Y.; Jiang, R.; Jia, Y.; Li, R.; Li, A. Association analysis algorithm based on knowledge graph for space-ground integrated network. In Proceedings of the 2018 IEEE 18th International Conference on Communication Technology (ICCT), Chongqing, China, 8–11 October 2018; pp. 222–226.

42. Ekelhart, A.; Ekaputra, F.J.; Kiesling, E. Automated Knowledge Graph Construction from Raw Log Data. 2020. Available online: http://ceur-ws.org/Vol-2721/paper552.pdf (accessed on 4 April 2022).

43. Ou, Y.; Zhou, T.; Zhu, J. Recommendation of cyber attack method based on knowledge graph. In Proceedings of the 2020 IEEE International Conference on Computer Engineering and Intelligent Control (ICCEIC), Chongqing, China, 6–8 November 2020; pp. 60–65.

44. Chen, X.; Shen, W.; Yang, G. Automatic generation of attack strategy for multiple vulnerabilities based on domain knowledge graph. In Proceedings of the IECON 2021–47th IEEE Annual Conference of the IEEE Industrial Electronics Society, Toronto, ON, Canada, 13–16 October 2021; pp. 1–6.

45. Vassilev, V.; Sowinski-Mydlarz, V.; Gasiorowski, P.; Ouazzane, K.; Phipps, A. Intelligence graphs for threat intelligence and security policy validation of cyber systems. In Proceedings of the International Conference on Artificial Intelligence and Applications, Suzhou, China, 15–17 October 2021; Springer: New York, NY, USA, 2021; pp. 125–139.

46. Syed, R. Cybersecurity vulnerability management: A conceptual ontology and cyber intelligence alert system. Inf. Manag. 2020, 57, 103334.

47. Tao, Y.; Jia, X.; Wu, Y. A research method of industrial Internet security vulnerabilities based on knowledge map. J. Inf. Technol. Netw. Secur. 2020, 39, 6–13.

48. Qin, S.; Chow, K. Automatic analysis and reasoning based on vulnerability knowledge graph. In Cyberspace Data and Intelligence, and Cyber-Living, Syndrome, and Health; Springer: New York, NY, USA, 2019; pp. 3–19.

49. Rastogi, N.; Dutta, S.; Christian, R.; Gridley, J.; Zaki, M.; Gittens, A.; Aggarwal, C. Predicting malware threat intelligence using KGs. arXiv 2021, arXiv:2102.05571.

50. Wang, L. Research on Software Security Vulnerability Mining Technology Based on Knowledge Graph. Master’s Thesis, Xi’an Technological University, Xi’an, China, 2021.

51. Wang, L. Research on Construction of Vulnerability Knowledge Graph and Vulnerability Situation Awareness. Master’s Thesis, University of Chinese Academy of Sciences, Beijing, China, 2020.

52. Najafi, P.; Mühle, A.; Pünter, W.; Cheng, F.; Meinel, C. MalRank: A measure of maliciousness in SIEM-based knowledge graphs. In Proceedings of the 35th Annual Computer Security Applications Conference, San Juan, PR, USA, 9–13 December 2019; pp. 417–429.

53. Dutta, S.; Rastogi, N.; Yee, D.; Gu, C.; Ma, Q. Malware knowledge graph generation. arXiv 2021, arXiv:2102.05583.

54. NEFOCUS. Security Knowledge Graph Technology White Paper. 2022. Available online: https://www.nsfocus.com.cn/html/2022/92_0105/166.html (accessed on 4 April 2022).

55. Wang, Z.; Sun, L.; Zhu, H. Defining social engineering in cybersecurity. IEEE Access 2020, 8, 85094–85115.

56. Wang, Z.; Zhu, H.; Liu, P.; Sun, L. Social engineering in cybersecurity: A domain ontology and knowledge graph application examples. Cybersecurity 2021, 4, 1–21.

57. Mitra, S.; Piplai, A.; Mittal, S.; Joshi, A. Combating fake cyber threat intelligence using provenance in cybersecurity knowledge graphs. In Proceedings of the 2021 IEEE International Conference on Big Data (Big Data), Orlando, FL, USA, 15–18 December 2021; pp. 3316–3323.

58. Xiao, H.; Xing, Z.; Li, X.; Guo, H. Embedding and predicting software security entity relationships: A knowledge graph based approach. In Proceedings of the International Conference on Neural Information Processing, Sydney, Australia, 12–15 December 2019; Springer: New York, NY, USA, 2019; pp. 50–63.

Retrieved from https://encyclopedia.pub/entry/history/show/63196