Research on Improving the Security Protection Capability of Critical Information Infrastructure based on Knowledge Graph

Yang Zhang, Songsen Yu
(South China Normal University, Guangzhou 510631, China)

Abstract: Critical information infrastructure is an important information system to ensure national social stability and normal economic operation, and it is also the main target of cyber-attacks. This paper proposes to build a threat-management knowledge graph, integrate security management requirement data and cyber threat intelligence data, and associate two types of data by link prediction algorithm, to realize the knowledge reasoning ability of threat management knowledge graph. The research shows that the threat-management knowledge graph can be used to deduce the security events of critical information infrastructure, find the corresponding security management requirements, and effectively improve the security ability of critical information infrastructure.

Key words: Critical information infrastructure; Threat-management knowledge graph; Security protection capability;

Critical information infrastructure is to ensure the national social stability, the normal operation of the economy of the core institutions or information systems, once its security accidents or data leaks, will have a huge impact on the society, therefore, critical information infrastructure is the focus of national cyber security protection object, but also the main target of cyber-attacks. China's protection of critical information infrastructure is mainly based on the Regulation on Security Protection of Critical Information Infrastructure and the People's Republic of China Cyber Security Law (referred to as the Security Law) , which stipulates the security management requirements and baseline indicators for the base unit. However, the existing security management requirements are formulated from the top down, which is difficult to respond to complex and changeable cyber-attacks. In the face of cyber security incidents that have occurred, it is difficult to find corresponding management solutions and responsible entities, which limits the cyber security capabilities of critical information infrastructure.

With the increasing complexity of cyber attack methods, cyber security practitioners, security service manufacturers, research institutions and organizations in various countries build a variety of cyber threat intelligence sharing platforms, such as cyber security forums, vulnerability databases, threat intelligence centers, attack pattern enumeration and security incident reports, which comprehensively describe and specify the characteristics and methods of various cyber-attacks. How to effectively use the cyber threat intelligence data, mining the effective information of the data, and associating the threat intelligence and management requirements is one of the main issues to improve the security protection capability of the customs base.

This paper puts forward the method of integrating cyber threat intelligence data and cyber security management data by constructing threat management knowledge graph, and using the reasoning ability of knowledge graph to realize the association between cyber threat intelligence knowledge and cyber security management knowledge, and then deduce the security management requirements corresponding to cyber security events to improve the security protection ability of the base.

1. Related research

In view of the important position of critical information infrastructure in the political and economic fields, many researches have been carried out at home and abroad to improve its security protection capability. Gao Mangru et al. proposed a method based on knowledge graph and high-dimensional semantic clustering, and derived a four-level index on the basis of the three-level index of the System, so as to improve the granularity of the protection of the base unit. From the perspective of national security, Zhang Yiteng put forward the risks faced by critical information infrastructure and suggestions for coping with them. Li Shengbao and others put forward countermeasures for the protection of the key base based on regional cyberspace governance. Liao Fangyuan et al. reviewed the methods and trends of G-based defense driven by artificial intelligence.

Existing studies have elaborated the protection methods of the kwan base from multiple perspectives, but this protection method is still top-down, relying on expert knowledge and manual inspection. As a result, the security inspection period of critical information infrastructure is long, subjective and unable to adapt to zero-day attacks, and when security incidents occur, it is difficult to locate its management vulnerabilities and responsible subjects.

The method proposed in this paper to improve the security protection capability of critical information infrastructure based on the threat-management knowledge graph is based on the bottom-up idea. By integrating a large amount of cyber threat intelligence and cyber security management requirements, the threat-management knowledge graph is constructed, and the correlation between the two types of knowledge is deduced based on the graph algorithm. To realize the intelligent analysis of the security protection of critical information infrastructure, deduce the security management requirements behind each cyber security event, and then improve the security protection capability of critical information infrastructure.

2. Build a threat-management knowledge graph

Knowledge graph technology can be used to store massive multi-source heterogeneous data, and form a large-scale graph structure through rich edge interconnection, which provides the basis for the subsequent graph calculation. By constructing the threat-management

knowledge graph, the paper integrates the two types of knowledge of cyber threat intelligence and cyber security management, realizes the interconnection operation between the two types of data, and deduces the causal relationship between the two types of knowledge.

(1) Cyber security ontology design

Ontology is the infrastructure of knowledge graph. By defining node categories and logical relations between nodes in the ontology, the framework planning of the graph can be realized, industry consensus can be reached, and the foundation for downstream tasks can be provided. In order to improve the security capability of the base, the external cyber threat intelligence data and cyber security management requirement text can be defined as multiple ontologies respectively, and the ontologies of the two types of knowledge can be associated through the relationship.

6 ontologies and 7 kinds of relations are designed in the threat-management knowledge graph ontology framework as shown in Figure 1, and the data sources are of two types: One is Cyber Threat Intelligence data (CTI). The research constructs CAPEC (Common Attack Pattern Enumeration and Classification), CWE (Common Weakness Enumeration) and CVE (Common Vulnerabilities & Exposures), which describe the patterns of cyber-attacks, Vulnerabilities & Exposures of objects respectively, are closely related to each other and already have rich correlations (associated by hyperlinks in the data sources). The second is cyber security management requirements. The research collecte domestic and foreign cyber security related management documents, and base on the tree structure of the documents, divide them into three ontologies from top to bottom: Doc, Title and Req, and build tree association relations based on the document context, the specific management requirements are leaf nodes. On the basis of this ontology framework, the relation from CAPEC node to Req node is constructed manually, and the association between cyber threat knowledge and management knowledge is realized.

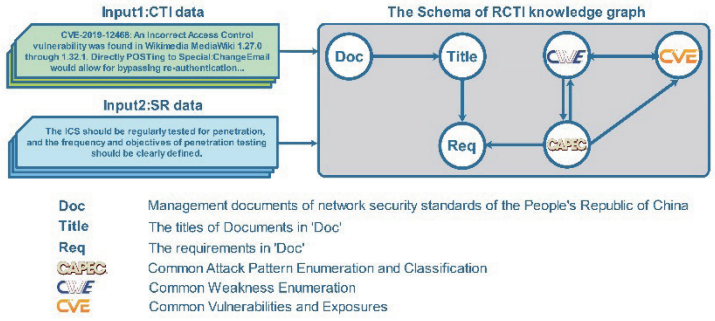

Figure 1. Cyber security knowledge graph ontology architecture

(2) Cyber security named entity identification

To transform the ontology framework into knowledge graph, it is necessary to automatically extract cyber security entities from massive cyber security data and assign them to the ontology framework. For example, the organization name, malware name or vulnerability name in the cyber security document is identified and put into the corresponding CAPEC, CVE, CWE or Req ontology to form the nodes of the knowledge graph.

Cyber security data is mostly stored and shared in the form of natural language, so named entity recognition technology based on natural language processing (NLP) technology can be used to extract cyber security entities. At the same time, for the semi-structured cyber threat intelligence data, the rule-based method can also be used to extract the cyber security entity information. The knowledge graph of owning entities is related by categories to form a preliminary large-scale graph structure.

(3) Cyber security relationship extraction technology

Relationship extraction is a technology that builds association edges for existing nodes in the knowledge graph based on new data, or directly extracts triples (h,r,t) with edges from the source data, where h and t represent the head entity and the tail entity respectively, and r represents the association relationship between the two. The relationship extraction of cyber security relies on a large number of cyber security terms and offensive and defensive knowledge, which has complex semantics and high professionalism, and is difficult to extract. Therefore, end-to-end algorithms based on deep learning are usually adopted.

For example, Li Tao et al. proposed a triplet extraction method based on adversarial active learning technology. Through sequence annotation and multiple iterations of the active learning algorithm, the ability to obtain higher knowledge extraction with fewer manual scalar notes is realized. Wen Qinghua et al. proposed a multi-strategy open relation extraction method to extract triples by comprehensively utilizing entity context, syntactic dependency and entity attribute information. The relational extraction technology constructs a large number of interconnection edges among multiple ontologies, making the structure of the threat-management knowledge graph further dense.

2. The linkage of cyber threat intelligence and security management knowledge is realized based on threat-management knowledge graph

Through entity extraction and relation extraction techniques of cyber security, threat-management knowledge graph has become a multi-source heterogeneous directed graph with a large number of nodes and relationships. However, the correlation between management knowledge (Req node) and cyber threat intelligence (CAPEC node) in the graph is still few, and the relationship between the two types

of data is still sparse, lacking of reasoning ability. Therefore, it is necessary to complete the knowledge graph through link prediction technology, so as to realize the linkage between cyber security management data and cyber threat intelligence data.

It is very difficult to construct link samples from CAPEC to Req. Builders need to master a lot of knowledge and experience of cyber attack and defense, and analyze data such as cyber security incident reports or emergency response reports to extract the causal relationship between cyber attack patterns and management requirements, and form correlation samples. Then, on the basis of a small number of existing relationship samples, the link prediction technology is used to learn the relationship model between the two types of nodes to realize the association of the two types of data.

Link prediction is a technology that deduces unknown edges based on the existing entity, relationship and topological information of knowledge graph. It can be expressed as: (h, r,?) or (? , r, t). A translation model-based approach, a Bert-based approach, or a graph-based neural cyber approach can usually be adopted:

(1) The method based on translation model represents the triples (h, r, t) respectively as vectors in high-dimensional space, and realizes the inference of new relations based on the translation characteristics of the vectors: $h \approx r+t$

(2) The Bert-based approach uses the text description of the triplet as input, represents the text description as a high-dimensional spatial vector through the Bert model or its variants, and deduces the new relation through the classifier or the translation model.

(3) The graph-based neural cyber method uses the topological information of the knowledge graph to obtain the spatial representation of nodes through the multi-layer propagation of node and edge representation vectors, and then realizes the inference of new relations. Its message propagation method can be expressed as:

$H^{l}=\sigma(AH^{l}\text{-}^{1}W^{l-1})$

Where H^{l} is the node representing of layer l, A is the adjacency matrix of the graph, and W^{l-1} is the learnable parameter of layer L-1.

Based on the link prediction technology, the topological information of the threat-management knowledge graph is enhanced, and the cyber threat knowledge and security management knowledge are effectively correlated to form a complete threat-management knowledge graph, as shown in Figure 2.

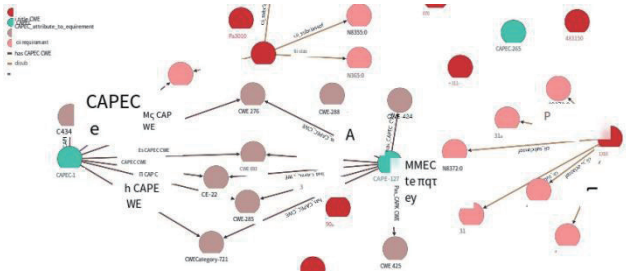

Figure 2. Threat Management Knowledge Graph (part)

The knowledge graph is usually stored, queried and visualized using relational data frameworks such as Neo4j or Orientdb. Figure 2 shows the threat-management knowledge graph built on the Orientdb framework.

4. Improve the security protection capability of critical information infrastructure based on threat - Management Knowledge Graph

Through the understanding of cyber security events on the base and the analysis of cyber attack events, the attack mode and vulnerability can be located in the threat-management knowledge graph, and the security management requirements and disposal methods corresponding to the attack mode can be deduced by link prediction technology, which can improve the security protection capability of critical information infrastructure.

(1) The result of cyber security knowledge inference

Based on the model obtained by the link prediction module, the nodes and relations in the knowledge graph have obtained a unique high-dimensional vector representation. By calculating the translation distance of the vector, the inference from cyber threat intelligence knowledge to management knowledge can be realized. After manual judgment, the correct results are selected as follows: cos(h+r, t)

Table 1 Inference results from CAPEC node to Req node (part)

Security Threat Intelligence	Security management requirements
Flash Injection	Regular audits of input validation should be ensured.
	System input should be restricted according to the content and format specified by the organization.
DNS rebinding	Cyber monitoring audit data standardization, its content must include date, time, subject identification, object identification, type, result, IP address and port.
Postfix, Null Terminate and Backslash	The response to invalid input should not change the order of operations.
	Regular audits should be conducted for input validation errors.
	The correctness of the input data should be verified.
	Develop an information security strategy, appoint a security officer, and coordinate the security management of the entire organization.

Rainbow Table Password Cracking	Make rules for analysis for things like cyber attacks and access breaches, including host scanning, port scanning, DDoS attacks, worms, password guessing, springboard attacks, etc.
	Inadequate password protection: Without proper password management mechanisms, unauthorized users may gain illegal access to confidential information.
	Evaluators on-site check whether passwords are kept in clear text on local systems or portable devices, whether there have been password leaks in the past, and assess the reliability of passwords using brute force techniques in simulated scenarios.
	Unauthorized (misuse, sniffing, spoofing, and social activity) access to user credentials stored in server components of industrial control systems.

Table 1 shows the inference results (in part) from the CAPEC node to the Req node, as shown in Table Flash Injection (CAPEC-182) is a cyberattack method that utilizes Flash for injection attacks. In this attack, the attacker tricks the victim into executing malicious Flash content, executing instructions specified by the attacker or making a Flash call and loading the material provided by the attacker. In view of this cyber-attack mode, the link prediction model deduces two corresponding security management requirements, both of which are input restriction and verification requirements, effectively prohibiting Flash elements during data input, which can effectively prevent the occurrence of its attacks or find and patch the vulnerabilities as soon as possible after the attack, and prove the validity of its reasoning.

(2) Based on threat - management knowledge graph, improve the security protection capability of critical information infrastructure

A cyber security incident is an impact caused by a cyber-attack on a key organization. No matter whether the attack is successful or not, as long as the characteristics of a cyber-attack are detected in the critical information infrastructure system, it is regarded as a cyber security incident. The critical information infrastructure can obtain the attribute information of the cyber-attack, such as attack mode, attack target, vulnerability or attack stage, through the intrusion prevention system, map the obtained attribute to the corresponding CAPEC, CVE or CWE node in the threat-management knowledge graph, and query the associated management requirement node by viewing the topology information of the node. The result of security management requirements corresponding to the cyber security event can be obtained.

Through the analysis of the cyber security incident and its security management requirements, the critical information infrastructure can self-check the implementation of the management item and the responsible subject, fill the loopholes, trace the source, and then realize the protection of the next similar cyber attack, and improve the security protection capability of the critical information infrastructure.

5. Concluding remarks

This paper studies the method of improving the security protection ability of critical information infrastructure based on knowledge graph technology, proposes to construct threat management knowledge graph to integrate cyber security threat intelligence data and cyber security management data, and use link prediction algorithm to associate the two types of knowledge, so as to realize the security reasoning ability of threat management knowledge graph. Based on the cyber security attack mode of critical information infrastructure, the reasoning results are given to find the corresponding security management requirements for cyber security events, which can effectively improve the security protection ability of critical information infrastructure.

References:

[1] Mengru Gao,Fangjun Xie,Hongqin Dong etal. Research on Cyber security evaluation system for critical information infrastructure [J]. Information Cyber Security,2019,No.225(09):111-114.

[2] Tao Li,Erao Guo,Ankang Ju. [J]. Journal of Communications,20,41(10):80-91.

[3] Qinghua Wen,Hongyin Zhu,Lei Hou etal. Multi-strategy Chinese open relation Extraction. Journal of Chinese Information Processing,2023,37(01):88-96.

[4] Siyu Tang,Saifei Li,Lijie Zhang. Analysis of Cyber Security Knowledge graph construction based on Neo4j [J]. Information Security and Communication Security,2022,No.345(08):60-70.

[5] Yingjie Wang,Chengye Zhang,Fengbo Bai etal. Research review on named entity recognition in Chinese [J]. Exploration of Computer Science and Technology,2023,17(02):324-341.

[6] Shengbao Li,Jiao Cheng,Yu Zhao etal. Analysis of Cyber security situation and protection mechanism for regional critical information infrastructure [J]. Cyber Security Technology and Application,2021,No.247(07):27-29.

[7] Fangyuan Liao,Jianfeng Chen,Zhiwang Gan.Review of AI-driven Critical Information Infrastructure Defense research [J]. Computer Engineeri ng,2019,45(07):181-187+193.

[8] Xiaobo Niu,Qun Fang,Xiao Shao. Cyber security emergency response based on threat assessment [J]. Cyber Security Technology and Application,2022, No.263(11):3-4.