Event-Based Threat Intelligence Ontology Model

Peng Wang(✉), Guangxiang Dai, and Lidong Zhai

19 Shucun Road, Haiden District, Beijing, China
wangpeng3@iie.ac.cn

Abstract. Cyber Threat Intelligence (CTI) has become an essential part of contemporary threat detection and response solutions. However, threat intelligence is facing challenges such as lack of unified standards, low efficiency of aggregation, difficulties in widely sharing, and low level of formalization in large-scale applications, which limits its potential in threat detection and response. In response to these challenges, this paper proposes an event-based threat intelligence ontology model based on a thorough analysis of existing threat intelligence standards, aiming to address the urgent need for efficient threat intelligence aggregation and human-machine application. Firstly, the ontology model leverages the semantic characteristics of events to reorganize the elements of threat intelligence, enabling humans to make quicker decisions, simplifying the hierarchical structure for automation processing, while being compatible with existing standards to promote intelligence sharing. Secondly, it combines the skeleton method and Formal Concept Analysis (FCA) method to achieve semi-automated construction, which can improve the efficiency and level of formalization, and aiding in the automated correlation analysis. Finally, we evaluate the proposed ontology and validates its effectiveness with specific instance data, hoping to provide inspiration and reference for other researchers.

Keywords: Threat Detection and Response · Threat Intelligence · Event-based Ontology · Intelligence Aggregation · Correlation Analysis · Intelligence Sharing

1 Introduction

As the cyberspace confrontation becomes increasingly intense, the direction of security operation is changing from passive defense to active defense, which is characterized by continuous threat detection and response, and timely and accurate warning for assets. To achieve this goal, threat intelligence is essential. At present, domestic and foreign security companies have established their own threat intelligence platforms. However, due to the differences in the level of technology, fields and standards of each, coupled with market competitions, these platforms can not achieve large-scale convergence and sharing. To promote the sharing of threat intelligence, the industry has proposed a series of standards and specifications, involving the unified description and exchange of threat intelligence among different entities. The mainstream related standards include CyboX (Cyber Observable eXpression), CAPEC (Common Attack Pattern Enumeration and

M. Yung et al. (Eds.): SciSec 2023, LNCS 14299, pp. 261–282, 2023.
https://doi.org/10.1007/978-3-031-45933-7_16

Classification), OpenIOC (Open Indicator of Compromise), STIX (Structured Threat Information eXression, Structured Threat Information eXression), China's proposed information security technology cyber security threat information format specification (GB/T 36643–2018), TAXII (Trusted Automated eXchange of Indicator Information of Trusted Automated eXchange of Indicator Information), etc. Based on these standards, the academic community has proposed and constructed corresponding threat intelligence sharing models and platforms [1]. These standards have promoted the sharing of threat intelligence to a certain extent, but due to the mechanism factors such as trust barriers and difficulty in allocation of benefits [2] and protection of privacy [3], vendors are not willing to share high-value intelligence, and the actual sharing is not effective. In addition, there are difficulties in producing and aggregating threat intelligence on the basis of these standards. Firstly, although some standard represented by STIX has strong expressive capability, it also has high complexity and can not reach a usable level in terms of automatic extraction accuracy. Secondly, the low level of formalization makes it difficult to conduct automatic correlation analysis, which is not conducive to the deep application of intelligence.

To facilitate the convergence and sharing of threat intelligence on a large scale, some researchers have proposed some ontology-based models and concepts to organize threat intelligence. An ontology is a formal description of important concepts shared in a specific domain, which provides a consistent framework and semantic model for individuals with different backgrounds and purposes by reducing conceptual and terminological ambiguities, thus enabling ubiquitous understanding and communication of information. Therefore, compared with various existing threat intelligence exchange languages, the ontology-based information sharing approach is more responsive to the needs of event information and knowledge organization in the modern cybersecurity domain [4].

Threat intelligence can be divided into human-readable threat intelligence and machine-readable threat intelligence from the perspective of user role. Regardless of the type, it should facilitate the role to understand the intelligence quickly so that it can make a swift decision. Domestic and foreign researchers generally believe that events have natural semantic properties and there are intrinsic connections among events, and building ontology models centered on events can facilitate the analysis of internal factors of events and the reasoning of relationships among events [5]. Ultimately it is conducive to promoting semantic retrieval and knowledge sharing [6].

Based on the analysis above, an event-based threat intelligence ontology model is proposed for threat detection and response scenarios that require efficient threat intelligence aggregation and human-machine co-application. Firstly, the ontology model uses the semantic characteristics of events to reorganize the elements of threat intelligence, which helps human to understand and make decisions quickly, and simplifies the expression hierarchy and improves the degree of structure, which facilitates the automated processing by machines, while it is easy to maintain compatibility with existing standards and promote the sharing of intelligence; secondly, the model combines the skeleton method and Formal Concept Analysis (FCA) method to achieve semi-automated construction, which improves the construction efficiency and formalization level of the model and helps to automate the correlation analysis of intelligence; finally, this paper

evaluates the ontology and verifies the effectiveness of the model with specific instance data, hoping to provide reference for other researchers.

The next chapters of this paper are organized as follows:

Section 2 reviews and summarizes the work related to the construction of threat intelligence ontology. Then Sect. 3 describes the detailed building process of the event-based ontology, and Sect. 4 introduces the application method of this ontology. Finally, Sect. 5 summarizes the full work and proposes the research direction in the future.

2 Related Works

2.1 Ontology Construction Research

From the perspective of the degree of manual involvement, the current ontology construction methods can be categorized into three types: manual construction, semi-automated construction and automatic construction. However, there is no highly effective method for automatic construction, so the first two methods will be mainly introduced.

(1) Manual Construction

At present, the mature manual construction methods include the skeleton method, TOVE (Toronto Virtual Enterprise) method, cyclic five-step method and six-step method. Among these methods, the skeleton method and six-step method have evolutionary and optimal evaluation steps, which help reuse and enhance the value of existing ontologies [7], and are therefore more commonly used. Manual construction methods rely on experts in this field and have high accuracy, but they are also subjective, costly and have poor portability.

Taking the skeleton method for an example, it generally needs to go through the main steps, which include application goal and scope determination, ontology analysis, ontology representation, and ontology evaluation. The first step is mainly used to determine the application area and scope of the ontology, which requires sorting out the specific knowledge of the domain according to the application scenario. Then, a conceptual model is formed through the ontology analysis step, and a normalized application model is created through the ontology representation step. Finally, after evaluation and correction, an ontology is constructed.

Among the steps, the ontology representation step involves how to express the ontology model, and the commonly used modeling meta-language for ontology models mainly includes concept classes, relations, functions, axioms, and instances [4]. Concept classes represent the set of all objects that conform to the concept, including common information such as properties and behaviors; relations refer to the logical or interactive relations between concept classes, such as inclusion relations, usage relations, etc.; functions can be regarded as special interactive relations between classes, where n-1 elements can uniquely determine the n-th element, and are often used in knowledge inference; axioms refer to eternal truth assertions, which are the basis of inference rules in the conceptual system; instances are the concretization of concept classes, which have all the properties and behaviors specified by the concepts and are influenced by conceptual relations. In the step of ontology evaluation, there is no unified evaluation system. Yue Lixin et al. [8]

selected five indicators of completeness, clarity, consistency, scalability and compatibility to evaluate multiple ontologies constructed by different methods at home and abroad, and Boeker et al. [9] proposed three indicators: usability, structure and functionality to evaluate ontologies.

(2) Semi-automated Construction

Semi-automated construction methods are mainly based on manual construction and automate some part of the steps to reduce the construction cost and subjectivity. There are mainly statistical-based methods and deep natural language processing-based methods. Statistical-based construction methods mainly use clustering, word frequency statistics, word co-occurrence analysis and other techniques for ontology element extraction and inter-element relationship mining, which use simple natural language processing techniques and are not ideal for relationship extraction; while deep natural language processing-based construction methods use semantic analysis techniques such as lexical annotation, syntactic analysis, dependency analysis, and semantic annotation, which can more effectively mine the relationship between elements, but these methods are difficult to apply to multiple domains because of the high requirements for training models [10].

In addition to this, there are many scholars working on the semi-automated construction of ontologies based on the formal concept analysis (FCA) approach. Formal concept analysis theory [9], a tool for data analysis and rule extraction from a formal context, improves automation by automating the construction of concept lattices to compensate for the tedious ontology hierarchical structure construction process. At present, this method has been widely applied in many fields. But it suffers from op-erational limitations when targeting multi-source heterogeneous data [10]. Liu Ting et al. [11] proposed a semi-automated construction method of coal mining face ontology CFOCFCA (Coal Face Ontology Construction based on FCA) based on the character-istics of coal mining face. Sun Li et al. [12] proposed a subject word list and FCA-based maritime ontology construction method, which merges structured resources (subject word list) and unstructured resources (text) to construct ontologies, extending the coverage of ontologies.

With the rapid development of Large Language Model (LLM) in recent years, LLM-based ontology construction methods have received widespread attention. For example, Milena T et al. [15] proposed to extract valuable information from unstructured text by automated means, to assist in the construction of knowledge ontologies. In addition, a series of LLM-based information extraction techniques have been proposed [16, 17, 25], which have to some extent facilitated the automated construction of ontologies. However, the LLM-based methods are currently limited by the lack of annotated data in large model training and have not been widely used.

2.2 Ontology Research of Threat Intelligence

Traditional researches of threat intelligence ontology can be divided into generalized ontologies and specialized ontologies. The generalized ontologies model the main concepts in the threat intelligence domain and focus on the representation. For example, Gao Jian et al. [18] constructed a threat intelligence ontology model that can be shared, reused, and extended based on STIX2.0 standard, and used the knowledge graph to

visually represent the important elements in intelligence and relationship between them, which helps intelligence analysts to make analytical decisions, but it's limited by the high complexity of representation and the difficulty of practical implementation. Specialized ontologies model sub-domains of threat intelligence and focus on applications. For example, Christian R et al. [19] proposed the ontology model MALOnt2.0 for capturing malware threat intelligence from heterogeneous data sources while constructing the malware threat knowledge graph MalKG, but it only supports graph query functions and is lack of inference and analysis capabilities. Yeboah-Ofori A et al. [20] proposed a cyber attack ontology to improve security based on cyber supply chain security, but the ontology only supports first-order logical queries in terms of application and doesn't contain inference capabilities. Sánchez-Zas C et al. [21] proposed an ontology for real-time risk management and cyber situational awareness that defines and validates a series of inference rules, but it lacks automatic response capability. Syed R et al. [22] proposed a network security vulnerability ontology that integrates vulnerability information from multiple sources and has a wide coverage. In addition, the authors designed an alerting system based on this ontology, which was evaluated to have a good performance, but the system is not fully automated.

Event ontology is a representation method for event knowledge, and there are different representation models in different research fields, which are generally divided into three categories: representation models based on conceptual hierarchy, logical hierarchy and event hexadecimal. In the third model, the event consists of six elements: action, object, time, environment, assertion and language, and the action element is the core, which can describe the event dynamically. The relationship between event elements can be described in detail and is more commonly used at present. In the research related to event-based threat intelligence ontology, there are no mature and systematic research results at home and abroad. Li Wenxiong et al. [23] studied network attack behavior and attack events from the attack case perspective and constructed a network attack case ontology, but the event elements were incomplete and lacked inference analysis capability. Yazid Merah et al. [24] proposed an ontology for risk detection, considering both security events and threat intelligence, and developed a network risk detection framework based on inference rules, but the application scope is limited to query and retrieval.

In conclusion, the existing threat intelligence ontology models, whether traditional or event-based ontology models, mainly focus on the expression, sharing, reuse and expansion of intelligence, and focus on the application of inference rules for retrieval and query, with little on intelligence correlation analysis and automated response.

3 Ontology Model Construction

In order to remedy these shortcomings and improve the dynamic semantic expression and reasoning ability of threat intelligence, the improved skeleton method is adopted to construct the threat intelligence domain ontology, in which the formal concept analysis method is used to improve the level of automation of ontology construction. The overall process is shown in Fig. 1.

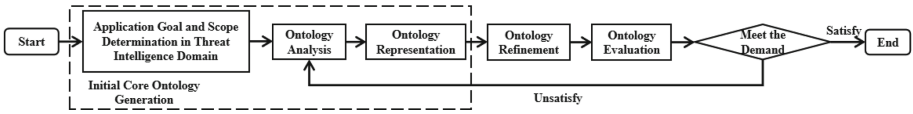

Fig. 1. Flow chart of Event-based Threat Intelligence Ontology Construction

The first three steps are defined as the Initial Core Ontology Generation Module, which is completed by domain experts. The Ontology Refinement step acquires the implicit information in the threat intelligence data automatically, and then combines the expert experience to refine the ontology.

3.1 Initial Core Ontology Generation

Application Goal and Scope Determination. The ontology designed in this paper is oriented toward both security operators for asset-protection-oriented security operations and numerous intelligence providers for facilitating the dissemination of threat intelligence. Specifically, the ontology model is applied to the following three aspects: intelligence aggregation, intelligence correlation analysis and intelligence sharing.

(1) Intelligence aggregation: Before being used for security operations, the threat intelligence ontology first needs to be able to store and manage intelligence, and other forms of intelligence should be easily converted according to the ontology model.
(2) Intelligence correlation analysis: For asset protection scenarios, the ontology should have intelligence correlation analysis capability, such as combining existing intelligence data to analyze threat information related to assets, including attacker information, asset vulnerabilities, countermeasures, etc.
(3) Intelligence sharing: For many intelligence providers, the ontology should have efficient intelligence sharing capability.

Ontology Analysis. In order to meet the application requirements mentioned above, this paper uses the event elements as the main line to organize threat intelligence, and each element of the event is elevated to the top level. With the principle that "each piece of intelligence is a (group of) event", each piece of intelligence should contain the following information shown in Table 2 of Appendix.

Based on the information requirements above, a conceptual model diagram of event-based threat intelligence ontology is developed as follows (Fig. 2).

Ontology Representation. Combining the elements in the conceptual model, and using object-oriented design philosophy, each of the elements above is abstracted into an event class. In this paper, the event-based threat intelligence ontology (ETIO) is defined formally as follows:

Definition 1. ETIO ::= {TECs, ECs, As, Rs, Rules}.

Among them, TECs is the set of top-level event classes, ECs is the set of classes other than top-level event classes for future expansion of the event ontology, As is the

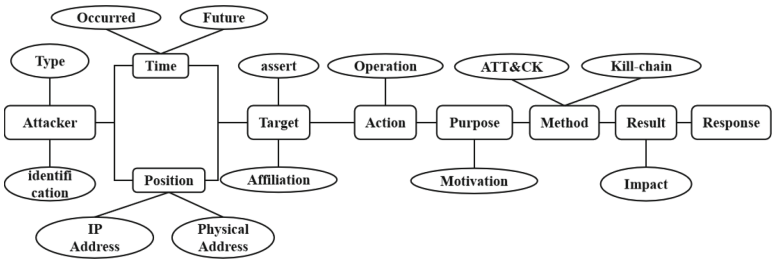

Fig. 2. Conceptual Model Diagram of Event-based Threat Intelligence Ontology

set of attributes of each event class, Rs is the set of relationships between event classes and between events, and Rules is the set of inference rules.

Considering the complexity of the event elements, the TECs are further defined formally in this paper as follows:

Definition 2 TECs: : = {Attacker, Time, Location, Target, Action, Motivation, TTP, Result, CourseOfAction}

Among them, Target describes the attacked object (also represents the defender); Action describes the type of attack in this event; TTP describes the attacker's attack method, including technique and tactics, tools, and process; Result describes the attack result; and CourseOfAction describes the response measures.

Definition 3 As: : = {Name, Type, Country, Time, Network location, Geographocal position, Location regularity, Version, Number, Tactical objective, Straregic target, Influence degree, Reliability, Description}

Definition 4 Rs: : = {R_event, R_event_class}
R_event: : = {Result, Follow, Co_occurrence}
R_event_class: : = {Has, Occurre_in, Include, Use, Aim_at, Cause, Belong_to, Locate_in}

Here, R_event refers to the relationship between events, R_event_class refers to the relationship between event class.

Definition 5. Rules ::= {Rules_as, Rules_other}.

Here, Rules_as denotes the inference rule between attributes of event class and Rules_other indicates other forms of inference rules.

Through the formal definitions above, a semantic foundation based on event elements is laid for threat intelligence, in which the set of relationships and inference rules can

be defined according to the application scenarios of the ontology in a targeted and complementary manner.

Oriented to the application requirements of threat intelligence in the security operation process, each top-level event class in TECs is further refined into subclasses and attributes, and converted from concept ontology to application ontology. The results are presented in Table 5 of the Appendix.

By defining the threat intelligence ontology, a hierarchical description of the threat intelligence knowledge required for guiding security operations and automated response is achieved. Among these elements, the design of subclasses fully conforms to object-oriented thought and is highly reusable, thereby reducing the number of layers and complexity.

3.2 Ontology Refinement

Traditional ontology construction methods rely heavily on human involvement, with a large degree of subjective influence, poor scalability and high construction costs. To tackle these problems, we combine the formal concept analysis method and expert experience to semi-automatically refine the ontology (Fig. 3).

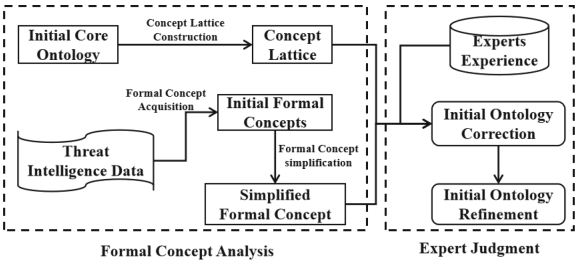

Fig. 3. Flow Chart of Ontology Refinement

Among them, the formal concept analysis part relies on the machine to automate the implementation, which can achieve the effect of human-machine collaboration.

Formal Concept Analysis (FCA). The theory of FCA is based on the mathematization of concepts and their hierarchies, and mathematical means are used to represent objective knowledge, thus weakening the subjective influence of ontology builders. In the FCA approach, the concept lattice is able to describe the hierarchy among concepts in an essentially clear manner and uncover the implicit information in the data [12], as well as facilitate human understanding. Therefore, in this paper, FCA will be used to assist experts in refining the ontology.

Formal Concept Acquisition and Simplification. Since there is a large amount of structured and unstructured data in the threat intelligence domain, which is real-time and may contain implicit information, the FCA approach is considered to automate the mining of implicit information related to event elements in the data. Based on this idea, this paper

extracts the objects and attributes contained in the threat intelligence data by using textual information extraction techniques, and then combines thematic models to simplify the formal concepts.

(1) Formal Concept Acquisition

ChatGPT has become a hot topic due to its powerful text generation capability recently. Wei X et al. [25] proposed ChatIE, which transformed the zero-sample information extraction task into a two-stage framework with multiple rounds of answering questions, and evaluated three tasks of relationship extraction, named entity recognition, and event extraction, and the experimental results showed that on two languages and six datasets, ChatIE achieves rather good results. Inspired by this, this paper considers using ChatGPT to extract objects and attributes from textual data in the threat intelligence domain. Specifically, we select two blogs, one report and one news media report that are highly relevant to cyber security as data sources, and select "cyber attack" as the keyword for information extraction, and finally obtain 117 initial formal concepts. Considering that these formal concepts are not only large in number but also uneven in quality, it is necessary to simplify the formal concepts before providing them to experts for analysis.

(2) Formal Concept Simplification

The LDA (Latent Dirichlet Allocation) topic model is an unsupervised learning algorithm that can efficiently process document data and classify numerous documents into topics according to probability distributions while displaying topic words. Therefore, in this paper, we consider to simplify formal concepts by using topic words in text. We select the same data and use the method based TF-IDF [26] and set the number of topics and the number of subject words to 4 and 50, respectively. Then we use ChatGPT to further sieve out the subject words that are not related to "cyber attack", and finally obtain a subject word list with a length of 32.

We define the simplification rule as follows: for each formal concept, we count the number of topic words (denoted as N), and set the threshold (denoted as keys), if $N \leq keys$, we delete the formal concept, otherwise we keep it. After the experimental analysis, when keys are 2, 3 and 4, the number of simplified formal concepts is 46, 16 and 8 respectively. Finally we choose $keys = 4$, and the set of formal concepts is shown on the github repository[1].

Concept Lattice Construction. Firstly, the initial core ontology is transformed into the formal context K1 by defining the following rules: i) the bottom concept class (or the top concept class if there is no subclass) of all ontologies is selected as the object in the formal context; ii) the attributes in the ontology are selected as the attributes in the formal context. Next, we use the method proposed by Lindig C et al. [27] to automate the construction of the concept lattice L1 corresponding to the formal context K1, which will be provided to the experts for ontology correction.

Expert Judgment

Initial Ontology Correction. Domain experts can be subjective in the process of constructing initial core ontologies, especially in the selection of attributes. In FCA, there

[1] https://github.com/LIGHTdgx/ETIO-Extraction-Results/tree/Results.

are three principles [28]: (i) concepts are described by attributes; (ii) attributes determine the hierarchy of concepts; and (iii) when two concepts have the same attributes, these two concepts are considered to be the same. Based on these principles, ontology correction is performed by domain experts according to the following correction principle: for the same attribute of different objects, choose to keep, delete or further divide it; for objects with the same attribute, choose to merge objects or add new attributes to distinguish them (Fig. 4).

Fig. 4. Concept Lattice L1

In the *Concept Lattice Construction* subsection, we constructed the concept lattice L1 corresponding to the initial core ontology, but we also observed that the attacker and the affiliation are grouped into the same formal concept due to the fact that these two concept classes in the Initial Core Ontology have the same attributes. According to the correction principle, we chose to delete attribute "country", add new attribute "experience", and further divide some attributes to make the distinction. Similarly, through a series of corrections, we obtain the corrected formal background K2 and the corresponding concept lattice L2. To make the figure clearer, we number the following attributes "Name of Attacker, Type of Attacker, Name of Asset, Type of Asset, Name of Affiliation, Type of Affiliation, Type of Time, Number of Observation, Name of Tool, Number of Attack Method, Deployment Difficulty, Number of CourseOfAction, Geographical location, Network location, Tactical Objective, Strategic Target, Degree of Influence" in order from "A" to "Q".

The corrected concept lattice L2 shown in Fig. 5 is more consistent with the definition of the initial ontology. Additionally, an implicit message can be inferred: the ATT/CK becomes a sub-concept of the kill chain and Vulnerability. Through the mapping rule proposed by Wei Lian et al. [29] between ontology and concept lattice, we mapped the modified concept lattice as ontology, and according to the division of abstraction levels in the network attack model, we chose to move ATT/CK into the subclass of the Kill chain.

Initial Ontology Refinement. In the subsection of *Formal Concept Acquisition and Simplification*, we obtained a simplified set of formal concepts, which were then analyzed by domain experts and constructed a new concept lattice to obtain the implied information, and finally the ontology was refined.

From the perspective of "object", we consider "AI security" as a generalization. From the perspective of "attribute", the data reflects the different stages of AI technology that can be used to automate cyber attack and defense, so we choose "stage", "deployment difficulty", and "automation" to summarize it. Finally, we obtain a new formal concept:

{“object”: “AI security”, “attribute”: “stage, deployment difficulty, automation”}, which is added to the formal context K2 to build the new concept Lattice L3.

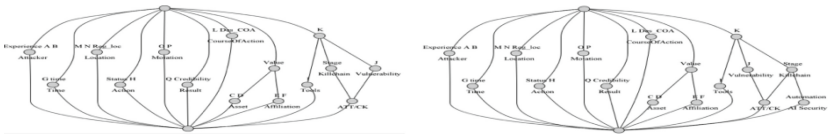

Fig. 5. Concept Lattice L2 and L3

We can infer that AI security is a sub-concept of the Kill chain along with the ATT/CK, which means that a cyber-attacker can combine AI techniques with attack techniques in ATT/CK to improve attack efficiency, but the top-level attack steps are still based on the Kill chain.

Finally, the refined concept lattice has been mapped into ontology according to the same mapping rules and so far we have obtained the corrected and refined ontology.

3.3 Ontology Evaluation

We use the “usability” and “structure” and "functionality" metrics mentioned in [9] to evaluate the ontology, where usability is concerned with the specific use of the ontology, i.e., whether other people can use the ontology without ambiguity. After the detailed description of the ontology construction process above, it is convenient for others to understand and meet the usability requirements. Structure is concerned with the formal structure of the ontology. Since the ontology defined in this paper is an event-based threat intelligence domain ontology, it is necessary to evaluate whether it encompassed all the information in the threat intelligence dimension and the event dimension, and the following will refer to the ontologies related to threat intelligence and security events for evaluation. Functionality is concerned with ontology applications, and we will discuss it in the next chapter.

Threat Intelligence-Related Ontology. Malware ontology [19] (MALOnt2.0), TAL ontology [30] (Threat-Agent-Library) and threat intelligence ontology proposed by Gao et al. [18] were selected for comparison, using the main components in the Information Security Technology Cybersecurity Threat Information Format Specification [31] (GB/T 36643–2018) as evaluation criteria. The results are shown in Table 3 of Appendix. From the table, it’s evident that the ontology defined in this paper contains most of the elements in the threat intelligence domain. Compared to the malware ontology which mainly describes the attack mode and attack behavior, the TAL ontology which focuses on describing information related to the threat subject, and the ontology proposed by Gao et al. which aims to describe the specific attack mode and attack indicators.

Security Event-Related Ontology. The computer security event ontology [32], the intrusion detection ontology [33], and the network attack case ontology [23] were selected for comparison, and the results are shown in Table 4 of Appendix. From the table, we can see that the ontology defined in this paper has rich elements in the event dimension,

which basically covers all elements of cyber security events, and the practitioners of security using this ontology can select different elements for correlation analysis.

Combining the analysis results in Table 3 and Table 4, it can be concluded that the ontology defined in this paper has a wide coverage and a strong richness.

4 Ontology Applications

Aiming at threat detection and response for asset protection, the ontology is applied from three perspectives: Intelligence Aggregation, Correlation Analysis, and Intelligence Sharing for multi-source heterogeneous intelligence data, and the flow chart of ontology application is as follows. In the step of intelligence aggregation, for structured data, it can be directly transformed into instance data through field mapping, and for unstructured data, the corresponding field values can be obtained from the text through information extraction technologies (Fig. 6).

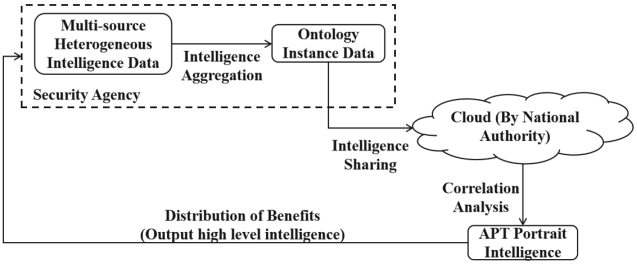

Fig. 6. Flow Chart of Ontology Application

4.1 Ontology Mapping

Ontology mapping refers to the extraction and transformation of information from data described by other standards. For reasons of corporate interests and privacy protection, most of the structured data in the field of threat intelligence is under control of domestic security vendors. Thus, we have chosen unstructured data as our data source. Next, we will take an APT report of AridViper [34] as an example. The data will be mapped according to the ontology of this paper by ChatGPT and the result will be presented as a conceptual diagram. The attributes of ATT/CK and CourseOfAction are referred to the technical and tactical knowledge base proposed by MITRE [35], and the attributes of attacker experience are referred to the "sophistication" field in the threat body component of the standard [31] (Fig. 7).

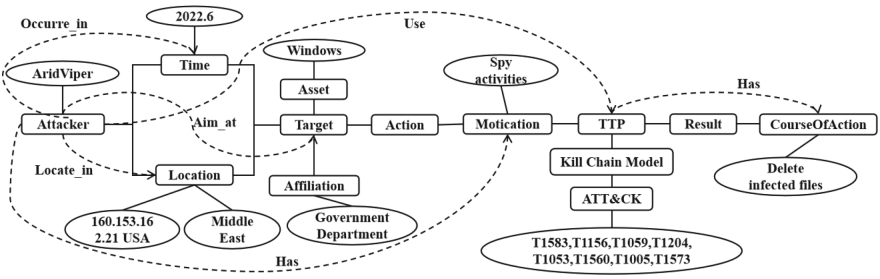

Fig. 7. An Instance Data for APT Report

4.2 Ontology-Based Intelligence Correlation Analysis

With the rapid development of network security technology, in practical application scenarios, the same attack organization will change its attack characteristics with the improvement of economic strength and technical capability. Therefore, we consider different types of attacks for the same attack organization to conduct correlation analysis, where the different types mainly refer to the attack target and the attack method. Decision implication [36] is a method that could be used for automated association analysis. In this paper, we will use the decision implication method to obtain the decision implication about APT organization by using the attribute values of elements in APT attack event as input, and then construct a portrait of APT organization to enable threat detection and response.

Instantiation of Inference Rules. To make the correlation analysis based on different classes of attack events more general, we instantiate the inference rules.

Definition 6. Rules_as ::= $\{f(\text{K\_as}) \rightarrow \text{d1}, g(\text{d1}) \rightarrow \text{d2}, h(\text{d2}) \rightarrow \text{d3}\}$.

Definition 7. Rules_as_augment : If$f(\text{A}) \rightarrow \text{B}, \text{A} \in \text{A1}, \text{B1} \in \text{B}$, then$f(\text{A1}) \rightarrow \text{B1}$.

Definition 8. Rules_as_combine : If$f(\text{A}) \rightarrow \text{B}, f(\text{A1}) \rightarrow \text{f(B1)}$, then$f(\text{A} \cup \text{A1}) \rightarrow \text{B} \cup \text{B1}$.

Here, function f is the inference rule corresponding to the decision implication method, and its mapping logic is described in [36]. K_as is the formal background set about the attribute set, respectively. Functions g and h are two inference rules based on decision implication with the mapping logic referring to the descriptions in Definition 7 and 8, and d1, d2 and d3 denote different decision implication. In addition, the literature [37] proves the soundness, completeness and non-redundancy of the latter two inference rules. In the following, we will apply these three inference rules to analyze the instance data.

Instance Analysis. Five representative event elements are selected: the attack time, attack motive, attack target, attack method and response measures, so that different attack events can be represented based on the attribute values of these elements. We selected eight reports on the "white elephant" APT organization as the data source and used

ChatGPT to extract entities. We established qualifiers for attributes to conduct multiple rounds of data extraction and processed the attributes with empty fields to facilitate correlation analysis. The extraction results are shown in github repository[1] (Table 1).

Table 1. Association Formal Background R1

	①	②	③	④	⑤	⑥	⑦	⑧	⑨	⑩	⑪	⑫	⑬	⑭	⑮	⑯	⑰
Event1			√	√	√	√		√	√	√	√	√		√	√	√	√
Event2	√		√	√	√	√	√	√	√	√	√		√		√	√	
Event3				√				√	√						√	√	
Event4	√		√	√	√	√	√	√	√	√			√		√	√	
Event5			√	√	√	√						√			√	√	
Event6	√	√	√	√	√	√	√	√	√					√	√	√	
Event7	√	√	√	√		√	√	√	√						√	√	√
Event8			√	√	√	√	√	√	√	√					√	√	

Aiming at threat detection and response, we select the attributes of attack method as the conditional attribute, and the attributes of attack target and response measures as the decision attribute to obtain the association formal background R1. For ease of presentation, we remove the attribute values unique to each event, and number the remaining attribute values "Reconnaissance, Resource Development, Initial Access, Execution, Persistence, Defense Evasion, Discovery, Collection, Command and Control, Exfiltration, CVE-2014–4114, CVE-2015–1641, MetaSploit, LINK, Data, Organization, Improving email precautions" in order from ① to ⑰.

Based on the first inference rule, the decision implication d1 is obtained and we demonstrate one of them:

{Initial Access, Execution, Defense Evasion, Collection, Command and Control} → {Data, Organization, Improving email precautions}

Based on the third inference rule, we get the combined decision implication d2.

{Reconnaissance, Resource Development, Initial Access, Execution, Persistence, Defense Evasion, Discovery, Collection, Command and Control,Exfiltration,CVE-2014–4114,CVE-2015–1641,MetaSploit} → {LINK, Data, Organization, Improving email precautions}

Based on the second inference rule, it is impossible to continue to augment the conditional attributes at this point. Finally, we represent decision implication d2 based on the ontology structure to obtain a portrait of the attack organization consisting of three elements: attack target, attack method and response measures (Fig. 8).

Evaluation Analysis. Finally, we evaluate the functionality of the ontology from three aspects: time performance of ontology mapping, completeness of ontology mapping and coverage of the portrait. The completeness represents the proportion of non-missing

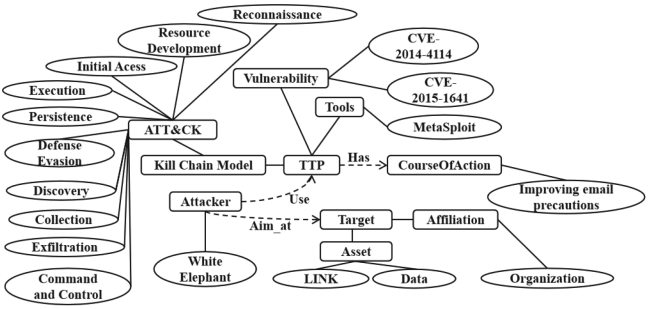

Fig. 8. Portrait of "White Elephant" APT Organization

fields in the mapping result (i.e. instance data) while the coverage represents the proportion of instance data transformed into a portrait. The formal definitions are as follows:

We denote N_i (i = 1,2,…,8) as the instance data extracted from the ith APT report, $|N_i|$ as the number of non-empty fields it contains, N as the number of mapping template fields (N = 11), e as the number of leaf nodes in the generated APT organizational portrait graph, M as ontology mapping completeness, and P as the portrait graph coverage.

$$M = \frac{1}{N} \sum_{i} \frac{|N_i|}{N}$$

$$P = \frac{e}{\sigma(\sum_i |N_i|)}$$

Here, σ denotes the de-duplication operation. After calculation, the completeness of ontology mapping is 84.1%, and the coverage of portrait is 26.6%. The mapping process took about 139 s. It can be seen that the mapping method used in this paper has good completeness and time performance. However, due to the small amount of data used in this study (only 8 APT reports), the coverage of the portrait is relatively low.

4.3 Ontology-Based Intelligence Sharing

To address the problems of trust barriers, benefit distribution [1] and fake intelligence provision in the current threat intelligence sharing model, a threat intelligence sharing mechanism is proposed based on the ontology model defined in this paper. The mechanism is targeted at security agencies and a CTI cloud served by national authorities, and the main steps are as follows.

(1) Security devices in different agencies only share raw intelligence information and do not involve sensitive intelligence (e.g., vendor product vulnerability information) or private data, which could reduce not only the sensitivity of intelligence but also the problem of sharing trust barriers.
(2) The national authority plays the role as the CTI cloud and is responsible for collecting and pre-processing the raw intelligence provided by each security device into an event-based threat intelligence ontology structure.

(3) The CTI cloud performs correlation analysis of the collected intelligence based on the ontology structure to generate higher-order intelligence (such as the portrait intelligence of APT organization described in **Instance Analysis** Subsect. 4.2). As the correlation analysis methods are continuously supplemented, the types of the higher-order intelligence generated by the CTI cloud will be enriched.
(4) The contribution degree of each original intelligence is calculated according to the contribution in the correlation analysis process of generated higher-order intelligence, while a penalty mechanism is established by combining the negative feedback from the intelligence user. Finally, the total contribution of the security organization is updated by the cloud in real time.
(5) The degree of contribution is used as the main reference for benefit distribution, which can take on various forms, such as awarding, certifications and licenses for enjoying higher-order intelligence subscription services, etc.

5 Summary and Outlook

In this paper, an event-based threat intelligence ontology model is proposed to address the urgent need for efficient aggregation of threat intelligence and efficient human-machine applications in threat detection and response scenarios. The semantic characteristics of events are used to reconstruct threat intelligence, which simplifies the expression hierarchy and improves the structure compared with existing standards. Secondly, we propose a semi-automated construction method based on improved skeleton method, which improves the model construction efficiency and formalization level. Then we introduce the application method of the ontology through an example, which can improve the efficiency aggregation and automated analysis level of threat intelligence to a certain extent, and promote threat intelligence sharing on a large scale. There are also some problems, such as the efficiency of human-computer combination in ontology refinement still needs to be further improved. In addition, the quality of the instance data obtained by mapping of unstructured data is affected by the data source and the prompt words used in the extraction process. Moreover, threat intelligence is time-sensitive but the ontology model cannot effectively represent the dynamic evolution of knowledge. The invalid intelligence is still required for manual filtering.

Our future work will focus on the following aspects:

(1) Research more effective automated/semi-automated ontology construction methods to improve the efficiency of ontology construction;
(2) Further improve the set of relations and attributes in the model and develop more effective inference rules to extend the application scope of the ontology;
(3) Research more effective information extraction methods to acquire instance data with higher quality, such as finetuning the GPT model;
(4) Research sharing techniques and mechanisms based on this ontology model, especially for the calculation of contribution degree;
(5) Research more effective evaluation criteria for ontology;
(6) Research threat intelligence correlation analysis method that deals with incomplete instance data referring to the work of Ning Hu et al. [38];
(7) Research dynamic knowledge representation methods based on spatio-temporal information for threat intelligence referring to the work of Jia Y et al. [39].

Appendix

Table 2. Event-based Threat Intelligence Concept Ontology

Event Elements	Element Description	Content Example
Human elements	Attacker's identity information (organization and country)	APT 29
Time elements	① Time of single attack ② Time rule of multiple attacks ③ Predict the occurrence time and probability of attacks	① 2022/5/12 22:00:23 GMT + 08:00 ② Launch DOS attacks frequently within 30 days ③ The probability of attacking a certain type of asset within 30 days is 70%
Location elements	① Attacker's network address (address pool) ② Attacker's Physical address	① Network address: 10.10.10.10 ② Physical address: XX country XX Province XX city
Object elements	① Target assets ② Asset owner information	① Windows 7 PC host ② XX Company
Motivational elements	Purpose of attack	Destruction, data theft, remote control
Movement elements	Behavior type	Normal access, determined attack, suspicious access
Methodological elements	Means of attack	Use the Blue of Eternity vulnerability to launch blackmail attack
Result elements	① Impact degree ② Credibility	① Serious impact, slight impact, no impact ② Reliability expressed by probability
Response elements	Preventive measures and disposal suggestions	Update the patch or upgrade to a higher version of Windows system as soon as possible

Table 3. Event-based Threat Intelligence Concept Ontology

	Our Model	MALOnt	TAL	Gao's Model
Type of Threat	√	√	√	√
Experience of Threat	√		√	
Time	√	√		
Impacted Assets	√			√
Security Event Status	√			
Motivation of Threat	√		√	
Attack Behavior	√	√		√
Impact Assessment	√			
Credibility	√			
Response Measures	√			√
Observable Data	√	√		√
Attack Stage	√			√
Attack Method	√	√		√
Information Source		√		
Attack Resources			√	
Vulnerability	√	√		√

Table 4. Event-based Threat Intelligence Concept Ontology

	Our Model	[26]	[27]	[17]
Time	√			
Location	√			
Attacker	√	√		√
Victim	√	√		√
Method	√	√	√	√
Result	√	√	√	√
Measure	√			√
Motivation	√	√		
Behavior	√	√		

Table 5. Event-based Threat Intelligence Application Ontology

Event Category	Subcategory	Attribute	Description
Attacker	/	name	Name or code information of the attacker
	/	type	Individuals and organizations
	/	Country	Country of attacker
Time	/	type	Single attack occurrence time, multiple attack occurrence time pattern, attack time prediction
	/	time	Single attack occurrence time, predicted time
Location	/	network location	The IP address used by the attacker
	/	geographical location	Geographic location of the attacker
	/	location regularity	Used to describe non-independent events(e.g., multiple attacks of the same kind launched by the same attacker) location patterns, such as multipoint concurrency
Target	Asset	name	Specific name of the asset, such as Windows 7 PC
		type	Hardware, link, service, data
		version	Version information corresponding to the target asset
	Affiliation	type	Individuals and organizations
		name	The name of Individuals and organizations
		Country	The country of Individuals and organizations
Action	/	type	Normal access, determined attack, suspicious access
	Observation	number	Numbering of observable behaviors described in STIX
Motivation	/	tactical objective	Attack objectives achieved at the tactical level
	/	strategic objective	Attack objectives achieved at the strategic level
TTP	Kill chain	stage	The seven stages described by the Killchain model

(*continued*)

Table 5. (*continued*)

Event Category	Subcategory	Attribute	Description
	ATT/CK	number	The number of the attack means described in ATT/CK
	Vulnerability	number	The CVE number of the used vulnerability
	Tool	name	The name of the used tool
Result	/	influence degree	The severity of the attack and the degree of caused impact, 0 - no 1 - minor 2 - moderate 3 - severe
	/	credibility	The credibility of this attack intelligence information, a continuous value between 0 and 1, 0 represents the lowest credibility, 1 represents the highest credibility
CourseOfAction	/	measure	Response codes taken
		description	A description of the response, such as patching

References

1. Karatisoglou, M., Farao, A., Bolgouras, V., Xenakis, C.: BRIDGE: BRIDGing the gap bEtween CTI production and consumption. In: 2022 14th International Conference on Communications (COMM), 16 June 2022, pp. 1–6. IEEE (2022)
2. Lin, Y., Liu, P., Wang, H., et al.: Overview of threat intelligence sharing and exchange in cybersecurity. J. Comput. Res. Dev. **57**(10), 2052 (2020)
3. Sarhan, M., Layeghy, S., Moustafa, N., Portmann, M.: Cyber threat intelligence sharing scheme based on federated learning for network intrusion detection. J. Netw. Syst. Manag. **31**(1), 3 (2023)
4. Chen, J.F., Fan, H.B.: Ontological threat intelligence sharing in cyberspace security. Commun. Technol. **51**(1), 177–183 (2018)
5. Liu, X.F., Fu, J.G., et al.: A comparative study of event-centric ontology models. J. Libr. Inf. Sci. **6**(02), 52–60 (2021)
6. Liu, Q.: Research on Ontology Construction and Application Based on Emergencies-Take the Covid-19 epidemic as an example. Shanxi University, Shanxi (2021)
7. Liu, S., Liu, X., Liu, X.: Overview of event ontology representation model and construction. J. Beijing Inf. Sci. Technol. Univ. **33**(2), 35–40 (2018)
8. Yue, L., Liu, W.: A comparative study of domestic and foreign domain ontology construction methods. Intell. Theory Pract. **39**(8), 119–125 (2016)
9. Astrid, D.R., Martin, B., Ludger, J., et al.: Evaluating the good ontology design guideline (GoodOD) with the ontology quality requirements and evaluation method and metrics (OQuaRE). Plos One **9**(8), e104463 (2014)
10. Ren, F.L., Shen, J.K., et al.: A review for domain ontology construction from text. Chin. J. Comput. **42**(3), 654–676 (2019)
11. Ganter, B., Wille, R.: Formal Concept Analysis. Springer, Berlin (1999)

12. Han, D.J., Gan, T., et al.: Research of ontology construction method based on formal concept analysis. Comput. Eng. **42**(02), 300–306 (2016)
13. Liu, T.: Research on Dynamic Ontology Construction and Reasoning Rules of Minning Face. Taiyuan University of Science and Technology, Taiyuan (2017)
14. Sun, L.: Research on Maritime Ontology Construction Based on Thesaurus and FCA. Dalian Maritime University, Dalian (2010)
15. Trajanoska, M., Stojanov, R., Trajanov, D.: Enhancing Knowledge Graph Construction Using Large Language Models. arXiv preprint arXiv:2305.04676 (2023)
16. Wang, S., Sun, X., Li, X., et al.: Gpt-Ner: named entity recognition via large language models. arXiv preprint arXiv:2304.10428 (2023)
17. Gao, J., Zhao, H., Yu, C., et al.: Exploring the feasibility of chatgpt for event extraction. arXiv preprint arXiv:2303.03836 (2023)
18. Gao, J., Wang, A.: Research on ontology-based network threat intelligence analysis technology. Comput. Eng. Appl. **56**(11), 112–117 (2020)
19. Christian, R., Dutta, S., Park, Y., et al.: An ontology-driven knowledge graph for android malware. In: Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security, pp. 2435–2437 (2021)
20. Yeboah-Ofori, A., Ismail, U.M., Swidurski, T., et al.: Cyberattack ontology: a knowledge representation for cyber supply chain security. In: 2021 International Conference on Computing, Computational Modelling and Applications (ICCMA), pp. 65–70. IEEE (2021)
21. Sánchez-Zas, C., Villagrá, V.A., Vega-Barbas, M., et al.: Ontology-based approach to real-time risk management and cyber-situational awareness. Futur. Gener. Comput. Syst. **141**, 462–472 (2023)
22. Syed, R.: Cybersecurity vulnerability management: a conceptual ontology and cyber intelligence alert system. Inf. Manag. **57**(6), 103334 (2020)
23. Li, W.X., Wu, D.Y., et al.: Research on cyber attack case base model based on onotology. Comput. Sci. **41**(10), 5 (2014)
24. Merah, Y., Kenaza, T.: Ontology-based cyber risk monitoring using cyber threat intelligence. In: Proceedings of the 16th International Conference on Availability, Reliability and Security, pp. 1–8 (2021)
25. Wei, X., Cui, X., Cheng, N., et al.: Zero-shot information extraction via chatting with ChatGPT. arXiv preprint arXiv:2302.10205 (2023)
26. Ge, B., Zheng, W., Yang, G.M., et al.: Microblog topic mining based on a combined TF-IDF and LDA topic model. In: Automatic Control, Mechatronics and Industrial Engineering, pp. 291–296. CRC Press (2019)
27. Lindig, C.: Fast concept analysis. In: Working with Conceptual Structures-Contributions to ICCS 2000, pp. 152–161 (2000)
28. Qian, J.: Research on Approaches of FCA-based Ontology Building and Mapping. National University of Defense Technology, Changsha (2016)
29. Wei, L., Li, D.M., et al.: Research on heterogeneous resource ontology construction based on FCA and Word2vec. Inf. Sci. **35**(3), 69–75 (2017)
30. Mavroeidis, V., Hohimer, R., Casey, T., et al.: Threat actor type inference and characterization within cyber threat intelligence. In:2021 13th International Conference on Cyber Conflict (CyCon), pp. 327–352. IEEE (2021)
31. GB/T 36643–2018. Information security technology—Cyber security threat information format (2018)
32. Howard, J.D., Longstaff, T.A.: A common language for computer security incidents. Sandia National Lab.(SNL-NM), Albuquerque, NM (United States); Sandia National Lab.(SNL-CA), Livermore, CA (United States) (1998)

33. Undercofer, J., Joshi, A., Finin, T., et al.: A target-centric ontology for intrusion detection. In: Workshop on Ontologies in Distributed Systems, held at The 18th International Joint Conference on Artificial Intelligence (2003)
34. The Phantom that Wanders the Middle East - Analysis of Recent Attack Activity by APT Group AridViper. https://www.uu11.com/keji/690217.html. Accessed 26 NOv 2022
35. ATT&CK Matrix for Enterprise. https://attack.mitre.org/. Accessed 25 Oct 2022
36. Zhang, S.X.: Research on Knowledge Representation and Reasoning Based on Decision Implication. Shanxi University, Taiyuan (2021)
37. Yanhui, Z., Deyu, L., Kaishe, Q.: Decision implications: a logical point of view. Int. J. Mach. Learn. Cybern. **5**, 509–516 (2014)
38. Ning, H., Tian, Z., Hui, L., Xiaojiang, D., Guizani, M.: A multiple-kernel clustering based intrusion detection scheme for 5G and IoT networks. Int. J. Mach. Learn. Cybern. **12**(11), 3129–3144 (2021). https://doi.org/10.1007/s13042-020-01253-w
39. Jia, Y., Gu, Z., Li, A.: MDATA: a new knowledge representation model. Springer, Heidelberg (2021). https://doi.org/10.1007/978-3-030-71590-8