Association Analysis Algorithm Based on Knowledge Graph for SPACE-Ground Integrated Network

Yulu Qi*, Rong Jiang*, Yan Jia*, Runheng Li†, Aiping Li*
*College of Computer, National University of Defense Technology, Changsha, 410073, China
†Sichuan Yi Lan situation Technology Co., Ltd., Chengdu, 610093, China
e-mail: qiyulu1103@163.com, jiangrong@nudt.edu.cn, jiayanjy@vip.sina.com

***Abstract*—The Space-Ground Integrated Network (SGIN) plays an important role for the future development of the country. The cyber-attacks against it are the focus of the research. In this paper, an association analysis algorithm based on knowledge graph of cyber security attack events is proposed to present the attack scenario for the Space-Ground Integrated Network. The construction of knowledge graph and association analysis can show the scene of cyber-attacks in the form of graphs. During the build process, the construction of an event ontology is an important part of it. Event ontology is used to represent various relationships in the network attack procedure. At last, we present a space-ground integration network security analysis system based on the knowledge graph of cyber security attack events, and uses the association analysis algorithm to analyze the attack scenario.**

Keywords-knowledge graph; event ontology; association analysis

I. Introduction

The space-ground integration information network consists of a space-based backbone network, a space-based access network and a ground node network. Satellite network is interconnected with terrestrial and mobile communication networks. The security equipment monitoring and processing modules are deployed on the space-based backbone network, space-based access network and ground node network respectively, and the collected data are collected separately through the heterogeneous network security Internet access aggregation component and the ground network security Internet security convergence component. Incoming collection subsystem, and then by the integrated network security analysis system of heaven and earth to obtain the analysis results. The network framework is shown in Figure 1.

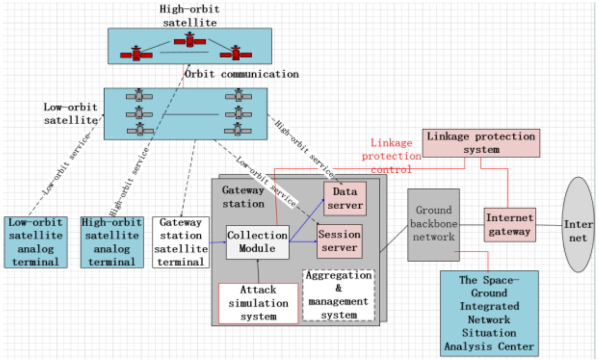

Figure 1. The space-ground integration information network framework

At present, widely used knowledge graph are mostly oriented to natural language. In the network security domain knowledge graph, network security attack events are not discrete and unrelated. One attack event is a combination of multiple attack steps. Attacking steps are also divided into causality, inheritance, selection relationship, etc. Therefore, this article adds event ontologies in the construction of knowledge graph to describe the steps of cyber-attacks. At the same time, the attack and alarm ontologies are added to the knowledge graph, and the relationship between attacks, events, and alarms is defined.

The idea of association analysis is that there are multiple attack steps for an attack. These attack steps are associated with alarms. Certainly, there are many combinations of these attack steps. For example, a type of attack use several different attack weapons and attack steps, but they are related to the same alarm. Based on this idea, a complete cyber-attack knowledge base is needed. As long as we know the attack steps, we can get the complete cyber-attack scenarios.

The rest of this paper is organized as follows: Section 2 discusses related work, and Section 3 presents the SGIN security analysis system architecture design. Section 4 provides simulation experiment and results, and Section 5 provides a conclusion and suggestions for future work.

II. Related Work

For the study of the space-ground integration network, although our country started late compared with the western countries, the relevant research and discussion on the integrated information network of heaven and earth have also continued for more than a decade. In 2006, Academician Shen [1] first proposed the concept and overall concept of China's integration of space and space. In 2013, Liu [2] and others proposed several thoughts and suggestions on the space-ground integration information at the 28th National Conference on Communications and Information Technology. In 2015, Academician Zhang [3] published <<Thinking about the Construction of China's Integrated World and Information Network>>, which clearly explained the positioning and boundary of the space-ground integration information network, and put forward the idea of the basic network architecture and suggestions for construction work.

In addition to strategic guidance, communication security is especially important in the space-ground integration information network. In 2016, Yang [4] et al. proposed a jammer-based secure communications scheme to confront of the situation that the eavesdropper could overheard the signals from user and the jammer.

978-1-5386-7635-6/18/$31.00 ©2018 IEEE

Professor Liu of Harbin Institute of Technology proposed that the Event Evolution Graph affair graph is structurally a directed ring diagram. Nodes represent events and directed edges represent incidental and causal relationships between events. Essentially, it is a Logic Knowledge Base which describes the evolution rules and patterns among events. At present, it is mainly applied to event prediction, common sense reasoning, and consumer intent mining and so on.

The typical topological structure of the Event Evolution Graph is shown in Table I:

TABLE I. TOPOLOGY STRUCTURE OF EVENT EVOLUTION GRAPH

Chain	Represents the sequence of events
Tree	Represents the choice of event
Ring	Represents the cycle of events

The difference between the Event Evolution Graph and the Knowledge Graph is shown in Table II.

TABLE II. DIFFERENCE BETWEEN THE TWO GRAPHS

	Event Evolution Graph	Knowledge Graph
Research object	Predicate events and relationships	Nominal entities and relationships
Organization form	Directed graph	Directed graph
Knowledge representation	Logic relations and probability transfer information	Entity attributes and relationships
Knowledge determinism	Most relationships are uncertain	Most relationships are determined

In the six-element model of events proposed by Liu [5] et al. in 2013, the definition of an event is as follows: Event refers to a certain time and environment, which is participated by a number of roles and shows a number of action features. The event can be formalized by a six-tuple:

$$\text{Event} = (A, O, T, P, S, L)$$

Among them, A represents a set of actions in an event and describes the process of the event. When an event occurs, there is at least one execution action, D represents each object participating in the event, including the subject and the object; T represents the time element, include start time and end time, and divided into absolute time and relative time; P is the position of the event; S is the state set of the object during the event (including preconditions, medium break words, and post-results). The setting condition is the precondition constraint that triggers the event, the intermediate continuation refers to the condition satisfied by each element in the middle of the event, and the post-result refers to the change caused by the event or the change of state of each element. L is the event. The language performance mainly includes the core words of events and the core words of events. Linguistic expression is the linguistic expression of an event and can represent the positional relationship or the inherent collocation pattern between non-core words and core words in a sentence.

The mechanisms and methods for describing events [6] mainly include the following:

1. An event ontology representation model based on the traditional ontology concept hierarchy, such as: Event Ontology [7], Linking Open Description of Event [8], etc.

2. Logical method-based event representation models, such as: Order-Sorted Logic Event Ontology [9], History Event Ontology [10], etc.

3. Event-based representation model based on six elements of events. With the development of event research, many event reasoning mechanisms have been proposed. The most influential ones are Event Calculus [11, 12] and Situation Calculus [13]. Event Calculus is a logic language for representing and reasoning actions and their interrelationships. In Situation Calculus, basic elements such as objects and scenes are added to Event Calculus. The former is an action-related arbitrary element, and the latter is used to represent the origin of the action.

III. THE SGIN SECURITY ANALYSIS SYSTEM ARCHITECTURE DESIGN

A. Network Security Knowledge Graph Construction

This paper presents a method for generating attack scenarios based on association analysis. This analysis method is based on the knowledge graph. Therefore, we must first construct the network security knowledge graph.

1. Network Security Knowledge Ontology Construction

The network security knowledge graph used in this article contains 5 tuples: Attack, Event, Alarm, Relation and Rule. The definitions are as follows:

K denotes knowledge graph [14], K = < concept， instance， relation， properties， rule >, where:

• Concept = {concept$_i$ | i=1,…,n}. The concept is a set of the abstract ontology, such as attack, event, and alarm. Attack is a concept. It is a collective term for all network attacks on the Internet, for example, control classes attacks, causative attacks and so on.

• Instance={instance$_i$ | i=1,…,m}. The instance is a set of concrete examples, event is a concept, and the instance of event is a specific implementation step of the attack and is associated with the attack. For example, the steps of a compound attack are penetration, latent attack, and launching an attack.

•Properties = {<instance$_i$, Pro$_{ij}$, value$_j$>}. The properties are a set of instance attribute values.

•Relation=<attack$_i$, R$_{cc}$, alarm$_j$> | <attack$_i$, R$_{ci}$, event$_j$>. Relationships represent the relationship between instances, such as subClassOf, instanceOf, beRaletedTo, and so on.

•Rule={rule | rule = <attack$_i$, newR$_{ij}$, event$_j$> | <attack$_i$, newR$_{ij}$, alarm$_j$ >, based on K}. Rules are used to constrain the association of attacks with events, attacks, and alerts.

2. Knowledge source and attribute description

Attack: crawl through the web crawler to describe information about the attack on the website and process the unstructured data into structured data and store it in the database. The attribute description of attack is shown in Table III.

TABLE III. THE ATTRIBUTE DESCRIPTION OF ATTACK

Attribute name	Description
name	attack name
type	attack type
description	attack description
severity	hazard level

Attack events: The specific implementation steps of the attack are also crawled through the network to obtain information. For example, the snort rule base. The attribute description of attack event is shown in Table IV.

TABLE IV. THE ATTRIBUTE DESCRIPTION OF ATTACK EVENT

Attribute name	Description
time	the time of each step
location	the location of each step
source	transmitted satellite
assert	assertion: pre-information and post-information
rule	attack steps: sequence, cause and effect, choice, etc.

Alarms: Alarms generated by the IDS detection system, such as the snort rule base. The attribute description of alarm is shown in Table V.

TABLE V. THE ATTRIBUTE DESCRIPTION OF ALARM

Attribute name	Description
name	alarm name
type	alarm type
description	alarm description
severity	alarm level

3. The abridged general view of ontology construction and ontology relations are shown in Figure 2.

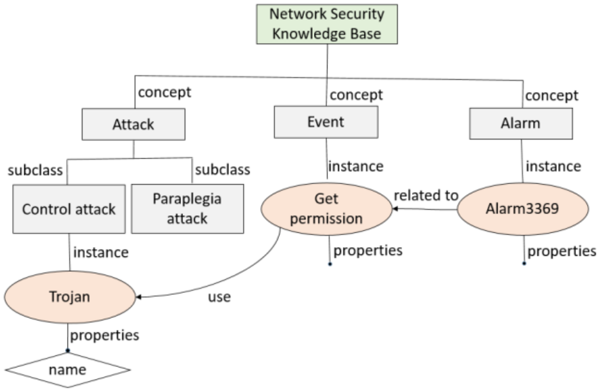

Figure 2. The network security ontology construction

The most important of the above-mentioned knowledge graph is the attack event. The event ontology is a specific refinement of the instances in the above five-tuple model. Here is a detailed description of how to construct the ontology of the attack event.

1. Event ontology construction

The event ontology contains five tuples: time, space, source (related to satellites), assertions, and rules. The definitions are as follows:

The E represents the event ontology, E=<time, space, source, assertion, and rule>, where:

- Time={time$_i$ | i=1,...,n}. The timestamp from the log information is the set of time for each step of the specific attack.
- Location={location$_i$ | i=1,...,m}. The record from the location of the log information on the location of the attack step is a collection of attack locations.
- Resource={resource$_i$ | i=1,...,m}. The collected attack information is sent to the ground analysis system through a specific satellite.
- Assert=<preconditiont$_i$, E$_{cc}$, postcondition$_j$>. Describes the causes and consequences of each step of the attack.
- Rule=<eventi, Rcc, eventj> | <eventi, Rci, eventj> | <eventi, Rii, eventj>. Indicates the sequence of occurrences between the various steps of the attack, such as order, selection and so on.

2. Event ontology source and attribute description

Source: Log information. The attribute description of event is shown in Table VI.

TABLE VI. THE ATTRIBUTE DESCRIPTION OF EVENT

Attribute name	Description
time	the time of occurred event
domain	the area of occurred event
log	the record of occurred event
resource	the source of occurred event
description	the description of occurred event

3. The abridged general view of event ontology construction is shown in Figure 3:

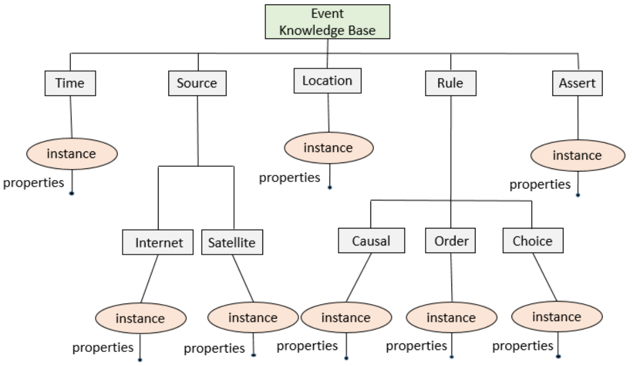

Figure 3. The event ontology construction

B. Association Analysis Algorithm

After one or several attacks occur, the data collection system collects relevant log information, including system logs, firewall logs, and IDS logs, and extracts event information from these logs and stores them in the database. The association analysis algorithm proposed in this paper firstly obtains the log information set L from the database, then parses the log and removes the redundancy, obtains the event set E, sorts the event sets chronologically, obtains the event list S, and the event list S is divided into n event lists by the length of the time window

$T_n = \{T_1, T_{2,} \cdots T_i\}$ $(1 \le \mathrm{i} \le \mathrm{n})$,each time window contains a set of event sequences $T_i = \{E_{i1}, E_{i2}, \cdots, E_{ij}\}$ $(j \ge 1)$, traverse the event window, match each event E_{ij} in the event window with the rules in the event repository, count the matching success times n, and calculate the success rate of the match. In the knowledge graph, each type of attack has a corresponding sequence of events. The matching success rate here refers to the coincidence degree between the sequence of events collected and the sequence of events attacked in the knowledge graph. If the match success rate is greater than the alarm threshold β, it is considered that the event sequence is successfully matched, and then the alarm information associated with the event is traversed, and the associated alarm information A_i is generated and put into the event-alarm set R together with the event sequence of the event window. Otherwise, the event sequence is only put into the event-alarm set, and finally the event-alarm set R is returned. The correlation analysis algorithm description is as follows:

Correlation Analysis Algorithm

```
Input:
        Log information set $L$ , Time window length $\alpha$ ;
        Alarm threshold $\beta$
Output:
        Event- alarm set $R$
$R \leftarrow \varnothing$
The event information is parsed from the log information set L and de-reduplicated to obtain the event set $E$
Each event of event set $E$ is sorted in chronological order to get the list of events $S$
Sort the sorted list of event $S$ by time window length $\alpha$ into $n$ event window list $T_n$
for $T_i$ in $T_n$ :
    $n \leftarrow 0$
    for $E_{ij}$ in $T_i$ :   // $E_{ij}$ is the No.j event in the event window $T_i$
        if event $E_{ij}$ matches the rules in the event repository:
            $n = n + 1$
    if $\frac{n}{\text{the length of } T_i} > \beta$
        The alarm information $A_i$ is generated, and the result $< T_i, A_i >$ is placed in the event-alarm set $R$
    else :
        Put the result $< T_i, \varnothing >$ into event-alarm set $R$
return $R$
```

IV. Simulation Experiment and Results

Due to lack of comprehensive understanding of the space-ground integration network and the limitations of current experimental conditions, this paper uses simulation experiments to verify the feasibility of the above algorithm.

A. SGIN-Based Attack Scenario Construction

This paper uses the Ukrainian power outage event in 2015 as an example to simulate the key steps in this APT attack.

Attack process settings: Through spear phishing emails or other means, BlackEnergy is first placed on the "springboard host", later, through BlackEnergy, the establishment was based on horizontal penetration using the "springboard host" as a main point. Then, get permission from the target machine through the Trojan. After obtaining the rights of the target computer, it steals data information and destroys the host hard disk. Simultaneously launch DDOS.

The schematic diagram of the simulation attack process is shown in Figure 4.

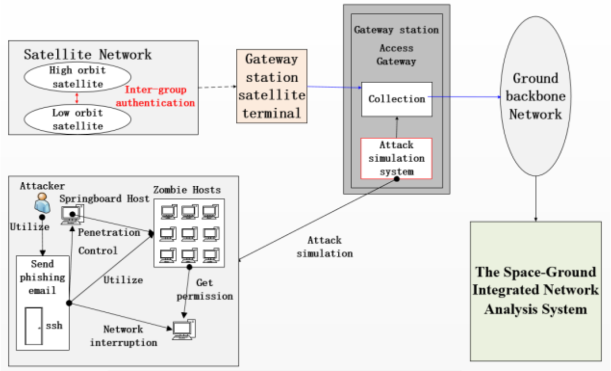

Figure 4. The schematic diagram of the attack process

B. Experiment Results and Conclusions

First, build the knowledge base according to the aforementioned ontology construction, and then integrate the association analysis algorithm into the space-ground integration information network analysis system, collect the data of the simulation attack through the collection tool, and finally, the analysis system shows the attack scenes in the form of visualization.

The process of association analysis is as follows: After the simulation attack, the data collection system collects system logs, firewall logs, and IDS logs of the bouncer, zombie and target machines, extracts 10 events information from these logs. The event information is subjected to a redundant pre-processing operation based on timestamps, and then the remaining 7 events are sorted in chronological order, and the event list is divided into 4 event sequences according to the time window length of 10 seconds. The first sequence of events contains 3 events: use ssh, open phishing email, and scan port. The matching success rate of the 3 events with the event sequences in the knowledge graph is 90%, which is greater than the set threshold of 60%. After the event is successfully matched, the alarm associated with the event is traversed, and then the 3-event sequence is associated with an alarm. The second sequence of events contains 2 events: get permission and steal information. The matching success rate between the two events and the event sequence in the knowledge graph is 75%, which is greater than the set threshold of 60%. After the event is successfully matched, the alarm associated with the event is traversed, and the two event sequences are associated with the alarm. The third sequence of events contains one event: host paralysis, which matches the event sequence in the knowledge graph with a success rate of 25%, less than the set threshold of 60%, and does not traverse the alarm associated with this event. The fourth sequence of events contains one event: network interruption, the matching success rate of this event with the sequence of events in the knowledge graph is 68%, which is greater than the set threshold of 60%. The alarm associated with this event is traversed, and the event is associated with the alarm. Finally, correlating all events and alarms in chronological order is the attack scenario of this attack.

The result of the association analysis is shown in Figure 5.

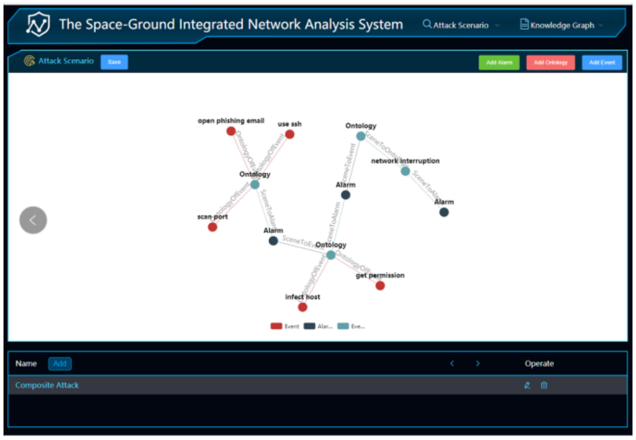

Figure 5. The result of the association analysis

The attack scenario derived from the association analysis is basically consistent with our pre-set attack procedure. But the time window and alarm threshold in the algorithm still need to be determined through repeat tentative to determine the best value.

V. CONCLUSION

The association analysis method proposed in this paper uses the knowledge graph to construct the cyber-attack scenario, which is applicable not only to the traditional Internet but also to the space-ground integration network. In order to better apply to the day-to-day space-ground integration network, it is necessary to further understand the uniqueness of the space-ground integration cyber-attack, exhaustive analyze these attacks and add these attacks to the knowledge graph. The next step need to restrict the construction of the event ontology and to make a lot of experiments to determine the selection of time intervals and alarm thresholds in the association analysis algorithm.

ACKNOWLEDGMENT

The work described in this paper is supported by National Key R&D Program of China (No. 2016YFB0800303), NSFC (No. 61732022).

REFERENCES

[1] SHEN R J. Sky-earth integrated space network idea of China[J]. Engineering Sciences, 2006(10):19-30

[2] Liu Xianzhu, Liu Xianzhu, Ding Ying, et al. Some Thoughts and Suggestions on the Integration of Heaven and Earth Information Network [J]. The 28th Annual Conference of National Communication and Information Technology, 2013.

[3] ZHANG N T, ZHAO K L, LIU G L. Thinking of newly geographical information in China[J]. Electronic Journal of Academy of Sciences, 2015, 10(3): 223-230.

[4] Yang T, Feng H, Yang C, et al. Resource allocation in cooperative cognitive radio networks towards secure communications for maritime big data systems[J]. Peer-to-Peer Networking and Applications, 2016:1-12.Zhong Zhaoman, Liu Zongtian, Li Cunhua. Event ontology model and ordering of event classes[J]. Journal of Peking University (Science and Technology), 2013, 49(2): 234-240.

[5] Zhong Zhaoman, Liu Zongtian, Li Cunhua. Event ontology model and ordering of event classes[J]. Journal of Peking University (Science and Technology), 2013, 49(2): 234-240.

[6] Zhu Zhonghua. Research on quantitative reasoning method of event ontology[D]. Wuhan University of Science and Technology, 2016.

[7] Liu Wei, Xu Wenjie, Tang Yingying, et al. Formal Representation and Reasoning of Event Actions Based on Extended Description Logic and Logic Programs[J]. Computer Science, 2014, 41(1):116-125.

[8] Shanahan M. The event calculus explained[M]// Artificial intelligence today. Springer-Verlag, 1999:409-430.

[9] Artikis A, Sergot M, Paliouras G. Reactive Reasoning with the Event Calculus[J]. Computer Science, 2015, -1(3):325-352.

[10] Denecker M, Ternovska E. Inductive situation calculus[J]. Artificial Intelligence, 2007, 171(5):332-360.

[11] Schiffer S, Ferrein A, Lakemeyer G. Reasoning with Qualitative Positional Information for Domestic Domains in the Situation Calculus[J]. Journal of Intelligent & Robotic Systems, 2012, 66(1-2):273-300.

[12] Gonzalez G, Baral C, Gelfond M. Alan : An action language for modelling nonmarkovian domains[J]. Studia Logica, 2005, 79(1) : 115-134.

[13] Zhong Z, Li C, Guan Y. Event ontology reasoning based on event class influence factors[J]. International Journal of Machine Learning & Cybernetics, 2012, 3(2):133-139.

[14] Jia Y, Qi Y, Shang H, et al. A Practical Approach to Constructing a Knowledge Graph for Cybersecurity[J]. Engineering, 2018, 4(1):53-60.