Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Supplementary materials for

Dandan WU, Jie CHEN, Ruiyun XIE, Ke CHEN, 2024. OntoCSD: an ontology-based security model for an integrated solution of cyberspace defense. *Front Inform Technol Electron Eng*, 25(9):1209-1225.
https://doi.org/10.1631/FITEE.2300662

1 A security defense mechanism for cyberspace

In Fig. 1, the left area represents the execution steps. The right area represents the execution elements in each specific process.

The left area is explained as follows:

Monitoring layer: The network monitoring system analyzes the network threat information and reports the vulnerabilities, illegal operations, attack behaviors, and other risks in the system.

Decision layer: According to the threat information, including the risk level, the threat type, the coverage, the weakest point, etc., the optimal decision-making solutions (as the security defense service set) based on optimization algorithms or reasoning algorithms under the current network risk or requirements are generated dynamically.

Response layer: The network security service node sends out the security defense service set for the area where there is a network threat to the key security platform, which plays an important role in secure resource organization.

Defense layer: The security platform (including backbone network security platform, boundary security platform, and terminal security platform), which obtains the security defense service set, installs and updates the security service set automatically and then reports new security information to the management center.

The content of the right area is explained as follows:

The network monitoring system mainly carries out the network threat information collection, analysis, and evaluation, which mainly provides risk intelligence.

The security platform is the executor of the decision-making scheme, which realizes the security behavior in different node locations or different time domains by loading the security service set dynamically. It is divided into a backbone network security platform, a boundary security platform, and a terminal security platform according to different functions.

The security resource service node receives the security service set issued by the management center, and then downloads it to the security platform in each user subnet.

Security atomic service is an inherent attribute of the security characteristics of systems, including confidentiality protection, integrity protection, undeniable protection, identity authentication service, access control service, etc.

Security component pools are a series of security defense component resources that can be implemented independently or jointly to realize the inherent properties of the security characteristics, such as a resource component pool and a service component pool.

The security service set aims to provide a security atomic service and relies on an optimization algorithm or a reasoning algorithm to aggregate different safe atomic services into a corresponding security service. It is

a set of components that provide security capabilities for systems and eliminate specific network threats.

2 The detailed design process for classes and relationships for the OntoCSD security model

The information system ontology class (ISComponents) is a collection of system entities such as management centers, network monitoring systems, firewalls, switches, routers, security resource service nodes, servers, data centers, computer terminals, and various security platforms. The security platform includes a terminal security platform, a boundary security platform, and a backbone network security platform.

The vulnerability ontology class is a collection of vulnerabilities. One is a backdoor vulnerability located at both the technical level and the management level of information components such as operating system vulnerabilities, database vulnerabilities, and application vulnerabilities. It can be exploited by some specific attack methods to launch network attacks, to damage system security, and even to obtain control rights as the worst case. The vulnerability database is obtained by the common vulnerabilities and exposures (CVE) leak library, and the severity score is obtained by common vulnerability scoring system (CVSS). The existing CVE knowledge base generally contains CVE numbers, versions, and a text description. For example, the vulnerability CVE-2022-21350 from the CVE leak library is located on the Oracle WebLogic Server, and other vulnerabilities related to different versions of Oracle WebLogic Server, such as CVE-2023-21931 and CVE-2023-21839, are found by querying. The second is the lack of physical security measures of the security platform. For example, the terminal security platform does not provide the required encryption parameter resources or encryption service resources. It will enable the confidentiality and integrity of the information data. Lastly, the physical security policy setting in the internal security platform is insufficient or lacking. For example, if the backbone network security platform has redundant interfaces and no security policy, this will cause viruses, network attacks, illegal access, and other risks. Additionally, the security state can be tracked by describing the relationship between the vulnerability ontology class and the information system ontology class. For example, the vulnerability of a network can be judged by the following statements:

$\text{hasVulnerability} \subseteq \text{ISOnto} \times \text{Vulnerability}$,

$\text{ExploitedWith} \subseteq \text{Vulnerability} \times \text{NetworkAttack}$.

The network attack ontology class is a collection of network attacks including (but not limited to) denial of service (DoS), code execution, buffer overflow, memory corruption, SQL injection, cross-site scripting (XSS), directory traversal, HTTP response splitting, IP address spoofing, man-in-the-middle attacks, and sniff attacks. When combined with corresponding network vulnerabilities, the system will face some risks.

The network attack ontology class can be used to configure attackers in modeling and reasoning. So, by the following statements, “Uses” means that the attacker can use some network attacks to achieve their goals:

$\text{Uses} \subseteq \text{Attacker} \times \text{NetworkAttack}$.

The defensive measures ontology class contains a collection of security service sets, which are used to describe the security defense implementation generated against security threats. The security services with multiple dynamic orchestrations are implemented by an algorithm based on a case legend. The security component pool subclass includes the security resource component pool subclass, the security service component pool subclass, and the security policy component pool subclass.

The security state ontology class is a collection of security attributes. Five subclasses—the confidentiality service subclass, integrity service subclass, identity authentication service subclass, undeniable service subclass, and access control service subclass—are included. It can be traced by describing the network attack.

Indicates, hasVulnerability, Mitigates, EquippedWith, ExploitedWith, CompromisedBy, SubClassOf, AttributedTo, Uses, CommunicateTo, and LackOf are included for the description of relationship about ontology classes. For example, “hasVulnerability” means that there are vulnerabilities in the system. The principal entity belongs to the network information system class, and the object entity belongs to the network vulnerability class. “ExploitedWith” means utilization. The principal entity belongs to the network attack class, and the object entity

belongs to the network vulnerability class. “CommunicateTo” means communication. The principal entity belongs to the network information system class.

3 Decision-making scheme generation based on CBR

3.1 Decision-making paradigm

The decision-making paradigm based on a similar case analysis is represented in the form of triplets, that is, Case:<Problem, Solution, Result>. Problem, Solution, and Result are used to describe the “problem” involved in the Case, the “solution” for the problem, and the “solution implementation effect,” respectively (Guo et al., 2014; He et al., 2020).

The decision-making paradigm based on a similar case analysis is expressed as

$$\text{Case}:\langle P,S,R\rangle. \tag{S1}$$

According to the historical cases C_1, C_2, …, C_n (n is the number of cases in the database) and the target case C_0, the feature attributes q_i=(q_{i1}, q_{i2}, …, q_{im}), q_{0i}=(q_{01}, q_{02}, …, q_{0m}), and feature weight vectors w=(w_1, w_2, …, w_m) are involved in the target cases (m is the number of feature attributes); the alternative decision-making scheme of the target cases is generated and optimized.

The system-related elements and historical case schemes are included in the feature attributes set. Six kinds of information forms that may be involved in practical decision-making problems with complex characteristics are considered, namely, clear symbols, clear numbers, interval numbers, fuzzy language variables, random variables, and texts. The multi-dimensional index set is used to evaluate the effect of the historical case scheme, including high efficiency, robustness, availability, security, and reliability.

3.2 Similarity calculation

Similarity calculation is divided into global similarity and local similarity, which can realize the similarity retrieval of a new case.

The global similarity is calculated as follows:

$$\text{SIM}\left(C_0,C\right)=\sum_{i}^{m} w_i\times\text{sim}\left(x_i,y_i\right), \tag{S2}$$

where x_i and y_i are the corresponding feature attributes in the new case and the historical case, respectively; sim is the local similarity of each feature attribute; w_i is the weight of features satisfying w_1+w_2+…+w_m=1.

The higher the value of SIM(C_0, C), the higher similarity the new case has.

The local similarity calculation: there are multiple types of data in the case, so the local similarity calculations need to be performed by type before retrieval.

Numerical similarity calculation:

$$\text{sim}=1-\frac{\left|x_i-y_i\right|}{x_i+y_i}. \tag{S3}$$

The main calculation determines the numerical characteristics, such as the number of security devices.

Interval similarity calculation:

$$\mathrm{sim} = \frac{x_i \bigcap y_i}{x_i \bigcup y_i}. \tag{S4}$$

The fuzzy similarity is calculated as follows:

$$\mathrm{sim} = 1 - \left| x_i - y_i \right|. \tag{S5}$$

The fuzzy type data, such as the environmental risk level, are calculated. Different degrees of data are assigned, such as strong, medium, and weak, corresponding to 0.9, 0.7, and 0.3 respectively, and then the similarity is calculated.

3.3 Case revision

Case revision is an important link to avoid empiricism. Generally, there are very few completely similar cases that can be found in the case base. If the current conditions are not carefully analyzed to solve the problem according to experience, it is empiricism. In the process of decision-making scheme generation, the scheme based on similarity is usually a suggested solution, which can only be used for reference. The final scheme also needs to be given by the decision-maker according to the actual situation, personal experience, and the corresponding technical means. That is, the solution of the historical cases can be modified and upgraded safely.

References

Guo M, Qian HZ, Huang ZS, et al., 2014. Intelligent road-network selection using cases based reasoning. *Acta Geod Cartograph Sin*, 43(7):761-770 (in Chinese). https://doi.org/10.13458/j.cnki.11-2089.2014.0120

He HW, Qian HZ, Duan PX, et al., 2020. Automatic line simplification algorithm selecting and parameter setting based on case-based reasoning. *Geomat Inform Sci Wuhan Univ*, 3:344-352 (in Chinese). https://doi.org/10.13203/j.whugis20180250