An Intelligent Learning Method and System for Cybersecurity Threat Detection

Yuan Tao¹ ², Wei Hu¹ ² and Moyan Li¹ ²
¹The Third Research Institute of Ministry of Public Security, Shanghai, China
²National Engineering Laboratory for Key Technology of Classified Information Security Protection, Beijing, China
Email: programfox@126.com

Abstract. To protect information system and data effectively is the core of cybersecurity. At present, most technologies are based on known static protection strategies. Through the cybersecurity threat is dynamic, so it is necessary to detect the cybersecurity in advance, so as to update the protection strategy in time, which means the dynamic and active defence. The intelligent learning method and system for cybersecurity threat detection is proposed. And the immune selection of cybersecurity threat detection is proposed to select the appropriate threshold, and the knowledge map of cybersecurity threat detection are constructed from suspicious time, security management centre, security computing environment, security area boundary and security communication network. The threshold is built by immune selection, and the cybersecurity intelligent detection can be achieved by knowledge map. The architecture of intelligent detection system is proposed. So that the cybersecurity threat detection can be dynamic and intelligent, and the performance of cybersecurity threat detection and analysis can be greatly improved.

1. Introduction
Cybersecurity protection technologies are included by firewall, intrusion detection, virus protection, data encryption and authentication technology. The most of these technologies are passive protection, and these technologies are based on the known static protection strategy [1, 2]. Through the cybersecurity threat is dynamic, so it is necessary to detect the cybersecurity in advance, so as to update the protection strategy in time, which means the dynamic and active defence [3-5]. And the intelligent learning method and system for cybersecurity threat detection is necessary [6, 7].

Because of the complexity of attack intrusion mechanism and the limitation of traditional detection methods, it is difficult to improve the accuracy of security analysis threat detection. Therefore, it is urgent to research the technologies of fast and accurate security threat analysis so as to provide support for the cybersecurity and supervision of the important information system. The intelligent learning method and system for cybersecurity threat detection is proposed. And the immune selection of cybersecurity threat detection is proposed to select the appropriate threshold, and the knowledge map of cybersecurity threat detection are constructed from the suspicious time, security management centre, security computing environment, security area boundary and security communication network.

The research is focused on the detection technologies of cyber flow, domain name, message and malicious code oriented multi-level abnormal behaviours, which is based on intelligent learning method. The threshold is built by immune selection, and the cybersecurity intelligent detection can be achieved by knowledge map. The multi type cybersecurity threat data statistical analysis modelling technologies and multiple security event correlation analysis technologies are constructed according to the model. And the architecture of intelligent detection system is proposed. So that the cybersecurity...
threat detection can be dynamic and intelligent, and the performance of cybersecurity threat detection and analysis can be greatly improved.

2. Cybersecurity Threat Detection Model for Heterogeneous Multivariate Security Data

Immune selection algorithm is proposed to realize intelligent rule selection and intelligent updating of threshold. So that the complex problems of cybersecurity threat detection are solved by using functions, principles and models of some immune systems. The basic rule set are adopted from the rule acquisition method which combines manual maintenance and automatic acquisition. In order to facilitate the intelligent selection of the cybersecurity threat detection model for heterogeneous multivariate security data, the immune characteristics of cybersecurity threat detection are defined as formula (1).

\[ \text{CST} = [B, TS, R, A, E, Rules] \]  

In formula (1), CST is cybersecurity threat. B is the belief, which indicates the credibility of the rule when using the immune selection algorithm for rule evolution. The value range is [0,1], 0 is the lowest and 1 is the highest. TS is Threshold. R is the relations, which indicates whether the rule runs directly when using the immune algorithm for rule evolution or is related to multiple rules and then runs. The value is Boolean, when R is true, which means that multiple rules need to be associated for re running. When R value is false, it means that the rule is running directly. A is cyberspace actions. E is cybersecurity events.

The immune selection algorithm is proposed to calculate cybersecurity threat level adaptability and data decision attribute adaptability of the corresponding rules. The rules are selected and the threshold are updated intelligently according to the immune characteristics of the rules. The \( f(x) \) function is the threshold fitness calculation function, which is the standard function for rule selection. The rules in the rule set A are should be able to select according to the cyberspace actions, and the rules in the rule set E are should be able to basically meet the cybersecurity events. So that the \( f(x) \) function, which is the threshold fitness calculation function, is defined as formula (2).

\[ f(x) = f < \text{argmax}(B), R, A, E, Rules > \]  

In formula (2), \( \text{argmax}(B) \) is the highest adaptability threshold, which is the rule with high immune belief. The \( f(x) \) function is to find a set of quasi optimal classification rules. The error between it and the optimal classification rules should meet the predefined value. If the fitness of threshold decision attribute of individual rule meets the pre-set requirements, the rule is marked, and then immune selection is carried out according to the mark to obtain the rule. If the obtained rules meet the threshold classification conditions of the system, a new classification rule set will be developed, otherwise, regeneration will be performed until the threshold classification conditions are met. The immune selection threshold of cybersecurity threat detection is shown as Figure 1.

![Figure 1. The immune selection threshold of cybersecurity threat detection](image-url)
3. Rapid Threat Perception Model for Intelligent Dynamic Behaviours Analysis

3.1. Intelligent Learning Rules for Knowledge Reasoning

In order to use the cybersecurity level to analysis the threat detection knowledge graph, the following illation rules are proposed:

1. Firewall access rules: (device type = / security device / firewall) and (event type = access)
2. DNS access rules: (event type = access) and (common DNS port for destination port)
3. Access rules from external host to local host: (event type = access) and (destination address local address) and (source address external address)
4. Access rules from local host to external host: (event type = access) and (source address local address) and (destination address external address)

In order to illation and analyse the network behaviour in abnormal time period, the working time and non-working time are specified as time illation rules, time illation rules are shown in Table 1.

Table 1. Time illation rules

| Non-working time                           | work time                           |
|--------------------------------------------|-------------------------------------|
| 19:00 every Monday - 7:00 every Tuesday    | 8:00 every Monday - 18:00 every Monday|
| 19:00 every Tuesday - 7:00 every Wednesday | 8:00 every Tuesday - 18:00 every Tuesday|
| 19:00 every Wednesday - 7:00 every Thursday| 8:00 every Wednesday - 18:00 every Wednesday|
| 19:00 every Thursday - 7:00 every Monday   | 8:00 every Thursday - 18:00 every Thursday|

3.2. The Knowledge Graph of Cybersecurity Threat Detection

Compliance with “GB/T 22239-2019 Information Security Technology- Baseline for Classified Protection of Cybersecurity” [8], “GB/T 28448-2019 Information Security Technology- Evaluation Requirement for Classified Protection of Cybersecurity” [9] and “GB/T 25070-2019 Information Security Technology- Technical Requirement of Security Design for Classified Protection of Cybersecurity” [10], and according to the intelligent learning rules for knowledge reasoning, the knowledge graph of cybersecurity threat detection is obtained.

According to time illation rules in table 1, threat detection knowledge graph of suspicious time is shown as Figure 2.

Security computing environment is storage and processing of the information of the grading system and relevant components for implementing the security policy, such as database, operating system, cloud platform, big data platform, etc. Threat detection knowledge graph of security computing environment is constructed from the illation rules of security computing environment. And the knowledge graph of security computing environment high-risk activity knowledge graph is shown as Figure 3.
Figure 3. The knowledge graph of security computing environment high-risk activities

Security management center is a platform or area for unified management of security policies and security computing environment, boundary of security area and security mechanism on security communication network, such as security management platform, cybersecurity situation awareness platform, etc. Threat detection knowledge graph of security management center is constructed from the illation rules of security management center. And the knowledge graph of security management centre association analysis is shown as Figure 4.

Figure 4. The knowledge graph of security management centre association analysis

Security area boundary is the boundary of the security computing environment of the rating system, and the relevant components of the connection between the security computing environment and the security communication network, such as IPS, firewall, etc. Threat detection knowledge graph of security area boundary is constructed form the illation rules of security area boundary. And the knowledge graph of security area boundary suspicious activity is shown as Figure 5.

Figure 5. The knowledge graph of security area boundary suspicious activity

Security communication network is the information transmission between security computing environments of rating system and related components of implementing security policies, such as switch, router, etc. Threat detection knowledge graph of security communication network is
constructed form the illation rules of security communication network. And the knowledge graph of security communication network non-conformance security policy is shown as Figure 6.

**Figure 6.** The knowledge graph of security communication network

4. The Architecture of Cybersecurity Threat Detection System Based on Big Data and Intelligent Computing

Cybersecurity threat intelligent detection system are distributed by big data storage technologies to store log and cybersecurity policy data. By providing large-scale distributed computing ability to quickly analyse log and cybersecurity policy data, the system can provide agile, reliable, safe and flexible service capabilities to intelligent detect the cybersecurity threat. The big data storage technologies used by cybersecurity threat intelligent detection system are OpenStack and CEPH. The cybersecurity threat intelligent detection system architecture design topology is shown as Figure 7.

**Figure 7.** Cybersecurity threat intelligent detection system architecture design topology
OpenStack is used to build a unified public storage warehouse for cybersecurity data, which can realize the deep encapsulation of the data resource platform, reduce and eliminate the generation of duplicate data and copies. So that the data storage can be isolated from other devices and networks.

CEPH is used to store and analyze in combination with the logs and cybersecurity policy data. The logs are included from device logs, system logs, application logs and database logs, the cybersecurity policy data are included from security management centre policy, security computing environment policy, security area boundary policy, security communication network policy.

In Figure 7, the cluster of the Cybersecurity threat intelligent detection system is composed of application server node, control node, computing node and storage node.

Application server node is responsible for receiving external access from users and uploading logs and security policies. In order to ensure the safety, it is deployed in DMZ area, and hot standby is carried out to ensure the normal operation of the business.

Control node is responsible for the control of other nodes, including virtual machine establishment, migration, network allocation, storage allocation, etc. Deployed in the isolated intranet area, in order to control the performance, the computing node and storage node can be called quickly. On this basis, flexible expansion of computing and storage nodes can be made.

Computing node is responsible for the operation of the virtual machine for log and security policies audit analysis. It is deployed in the isolated intranet area, which can meet the calculation requirements of audit services and can be flexibly expanded and reduced according to the change of the number of customers.

Storage node is responsible available storage space, and is responsible for storage management of virtual machine and other than it, such as storage and backup of log files and security policy files. The storage nodes are highly available and can be dynamically expanded.

5. Conclusion
The core of cybersecurity protection is effectively to protect information system and data. Through the cybersecurity threat is dynamic, so it is necessary to detect the cybersecurity in advance, which means the dynamic and active defence. An intelligent learning method and system for cybersecurity threat detection is proposed so as to defence information system and data dynamically and actively.

The immune selection of cybersecurity threat detection is proposed to select the appropriate threshold, and the knowledge map of cybersecurity threat detection are constructed for suspicious time, security management centre, security computing environment, security area boundary and security communication network. The threshold is built by immune selection, and the cybersecurity intelligent detection can be achieved by knowledge map.

The architecture of cybersecurity threat detection system based on big data and intelligent computing is proposed. So that the cybersecurity threat detection can be dynamic and intelligent, and the performance of cybersecurity threat detection and analysis can be greatly improved.

6. Acknowledgments
This research was financially supported by the National Key R & D Program of China (2018YFB0803503).

7. References
[1] Schinagl S, Schoon K, Paans R. A framework for Designing a Security Operations Centre (SOC) [C]. The 48th Hawaii International Conference on System Sciences (HICSS). IEEE, 2015:2253-2262.
[2] TAO Yuan, ZHANG Yuxiang, MA Siyuan, et al. Combining the Big Data Analysis and the Threat Intelligence Technologies for the Classified Protection Model [J]. Cluster Computing, 2017, 20 (2): 1035-1046.
[3] JIA Yan, QI Yulu, SHANG Huaijun, et al. A Practical Approach to Constructing a Knowledge Graph for Cybersecurity [J]. Engineering, 2018, 4:53-60.
[4] WU Yunchao, MAO Shaojie, ZHOU Fang. Domain Knowledge Graph Construction Technology for Simulation Deduction [J], Command Information System and Technology, 2019, 10 (3): 32-36, 82.

[5] Olga Russakovsky, Jia Deng, Hao Su, et al. ImageNet Large Scale Visual Recognition Challenge [J]. International Journal of Computer Vision, 2015, 115 (3): 1-42.

[6] Hasim Sak, Oriol Vinyals, Georg Heigold, et al. Sequence discriminative distributed training of long short-term memory recurrent neural networks [C]. INTERSPEECH 2014: 1209-1213.

[7] Niu Yulei, Lu Zhilei, Wen Jirong, et al. Multi-Modal Multi-Scale Deep Learning for Large-Scale Image Annotation [J]. IEEE Transactions on Image Processing, 2019, 28 (4): 1720-1731.

[8] MA li, CHEN Gongyong, ZHANG Zhenfeng, et al. GB/T 22239-2019 Information Security Technology- Baseline for Classified Protection of Cybersecurity. Beijing: Standards Press of China, 2019.

[9] CHEN Gongyong, LI Ming, LI Shuiling, et al. GB/T 28448-2019 Information Security Technology- Evaluation Requirement for Classified Protection of Cybersecurity. Beijing: Standards Press of China, 2019.

[10] JIANG Yong, LI Chao, LI Qixiang, et al. GB/T 25070-2019 Information Security Technology-Technical Requirement of Security Design for Classified Protection of Cybersecurity. Beijing: Standards Press of China, 2019.