IIAI Open Conference Publication Series
IIAI Letters on Informatics and Interdisciplinary Research
Vol.005, LIIR189
DOI: https://doi.org/10.52731/liir.v005.189

Development of a Cybersecurity Training System Based on SaaS

Sanggyu Shin *

Abstract

This study proposes the development of a cloud-based cyber attack and defense exercise system that enables practical cybersecurity exercises and experiences by using the Cyber Range environment built in a virtual space on the cloud via a local web browser. The platform and exercise programs will be developed on an open-source software platform built on the cloud as an ecosystem. The exercise contents will be developed using Docker, which has excellent portability, and attack and defense exercise scenarios will be produced as containers. By building the system in the cloud, this study proposes a platform that can be configured with various exercise scenarios independent of local PC performance. This study (1) develops a SaaS platform that can be connected to the cloud-based training environment via a web browser. (2) Enable users to interactively respond to attacks and defenses in a virtual space on the cloud in an internet environment. Finally, (3) The curriculums are structured based on microservices to be configured flexibly.

Keywords: Cyber Range, cybersecurity, training system, cloud, SaaS

1 Introduction

Cybersecurity has long been cited as a significant issue along with the high-speed change to an IT society [1] and DX (Digital Transformation) [2]. The government and industry are requesting higher educational institutions to enhance practical security education, and some universities are launching security education with support from companies. However, many academic institutions and small and medium-sized enterprises still need more support. They are struggling because they cannot handle the available human resources and budget.

Cyber Range is a virtual space specially constructed on a computer to train cyber attack and defense exercises, allowing students to learn more practical information security techniques [3][4][5][6].

1. To construct the Cyber Range system in a locally isolated network environment (For example, VMware, usually used at Cyber Range, is configured on physical PCs to create a dedicated environment), its management and operation require expertise.
2. There is a significant burden for flexible configuration of the contents of the exercises and introduction or modification of new scenarios.
3. When curriculum changes or safety issues are mentioned due to adding new content, a professional response to dedicated systems is required (vendor-dependent).

* Tokai University, Japan

This study aims to propose a new cybersecurity education system based on the SaaS (Software as a Service) concept to build a cloud-based Cyber Range. The conventional Cyber Range is built on a local system, which has the disadvantages of high hardware-dependent costs and difficulty in changing educational programs. This study proposes developing a SaaS architecture to overcome these disadvantages based on open-source software, which is not hardware-dependent but built on the cloud. This proposal for a cybersecurity education system can be flexibly responded to meet the individual requirements of universities and companies. Next, this study proposes incorporating an ecosystem (collaborative development of contents by open source) where educators can share educational methods (joint use) and a mutual assistance system to construct various exercise scenarios through cooperative development.

The proposed system is based on the concept of the *CyExec* system [7][8], which is previous research and has enhanced several new functions to make it practicable. The users can use the content, a sandbox built on server spaces using virtual container technology via a client-side Web browser.

2 Background

On February 8, 2022, NRI Secure Technologies released the results of its "NRI Secure Insight 2022," targeting companies in Japan, the United States, and Australia. According to the results of the survey, the total of "somewhat insufficient" and "insufficient" in terms of the availability of human resources engaged in information security management and internal system security measures was 89.8% for Japanese companies, with nearly 90% of the companies feeling a shortage of security personnel. This result is the result of a survey of the same options. This percentage is an overwhelming difference from the 9.7% in the U.S. and 10.8% in Australia, where the same options were tabulated [9].

On November 1, 2023, the ISC2 (International Information System Security Certification Consortium), a non-profit organization that develops and certifies security personnel, announced the 2023 edition of its annual global security human resources survey, "ISC2 Cybersecurity Workforce Study." According to the report, the number of security personnel in Japan increased 23.8% from the previous year to 481,000 people. However, the demand for human resources also increased by 33.0% to 591,000 people, and there is still a need for more than 110,000 people [10].

Additionally, ransomware damage to companies and organizations continued to increase in Japan. In terms of attack tactics, in addition to "double threats" to expose stolen data, DDoS attacks on victim organizations and other threat methods, such as contacting customers and stakeholders of the victim organization about the damage, have been confirmed. In the February 2022 ransomware attack on an automobile parts company, the operation of the automobile plant that supplied the parts was suspended for a day. In a ransomware attack on a medical center in Osaka City in October of the same year, the electronic medical record system of the medical center was compromised via a server after being compromised by a food service provider connected via VPN [11][12][13].

As mentioned above, a security learning system that is easy for anyone to learn and operate is necessary to counter Japan's security personnel shortage and cyber attacks that are evolving year by year. As part of efforts to develop security personnel, some higher education institutions provide education through practical exercises [7][8][14][15].

3 Conventional Approaches to Security Education

Practical exercise-style training systems used in previous security personnel training efforts include exercises using *Experiential learning tools* and Cyber Range exercises.

Experiential learning tools allow users to systematically learn basic knowledge about vulnerabilities in a practical training format, such as an overview of vulnerabilities and countermeasures. A typical example is IPA's "AppGoat," which allows users to learn about vulnerabilities by installing the AppGoat learning tool for web applications on a local PC [16]. Experiential Learning Tool is,

1. Need to build environment locally.
2. Fixed exercise scenarios.
3. Organizational attack/defense methods are outside the scope of learning.
4. Mutual exercises between attack and defense are not developed except for a few.

Cyber Range is a virtual space built on a server for learning cyber attack methods and their countermeasures. The main feature of the Cyber Range is that it enables practical learning by simulating situations that are close to actual ones [4][5][6][14].

Traditional Cyber Range is,

1. System construction in a local, isolated network environment (configuring VMware and other software on physical PCs to create a dedicated environment) and its management and operation require expertise (① in Figure 1).
2. Flexible configuration of the contents of exercises and introduction or modification of new scenarios require a significant workload and cost (② in Figure 1).
3. If the curriculum is changed due to the addition of new content or safety issues are raised, specialized support for dedicated systems will be required (③ in Figure 1).

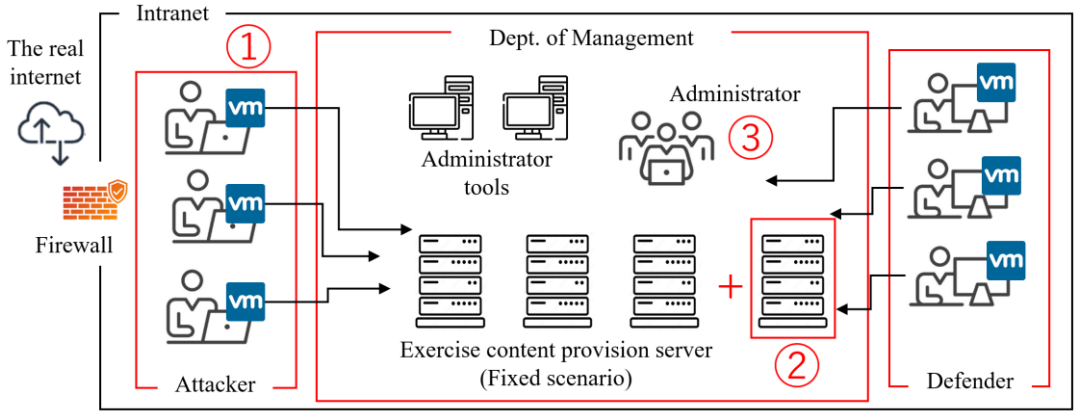

Figure 1: Existing Cyber Range architecture

These problems are solved by (1) not operating a physical system locally but placing the training resources on the cloud, and (2) cybersecurity training that can provide a practical training environment in conjunction with theoretical education. In addition, it is evident that cybersecurity

education should be learned anytime and anywhere, and that basic security knowledge and skills should be acquired. The problem, however, is that the contents of cybersecurity are so vast and complicated, and the factors involved are so intertwined that theoretical learning has apparent limitations. Therefore, the cultivation of practical skills to connect theory to practice must rely on case study learning, and herein lies a factor that hinders effective and efficient learning.

burden of development and management.

4 Proposed System

This study adopts the CyExec architecture, which enables the use of exercise contents in Docker containers as an ecosystem [7][8]. The exercise contents are developed as a system that can be provided and used via a web browser. In addition, a content management system and learning record management system are developed simultaneously, and a management screen for operators is created to provide an environment where exercises can be conducted efficiently. This system enables information security education in the same training environment as CyExec, eliminating the need to build a local system and reducing the cost of system construction and the burden of development and management.

4.1 Architecture

The preceding study, CyExec, still needs help with system construction issues on the local side and the creation and combination of flexible, practical exercise content. In this study, the question is whether this problem can be solved. Based on the improvements raised from the actual operation of CyExec and the problems with the existing Cyber Range, this proposal considered a plan for a cloud-based Cyber Range development infrastructure research using the SaaS concept that provides the development, provision, and management of exercise environments in a single package via the internet web. The plan included the following research topics. The following research issues were considered in the plan.

1. Develop network control technology, architectural models, and APIs that enable secure connection to virtual environments in the cloud via local web browsers. It will be possible to provide an exercise environment that does not depend on the hardware performance of the local side (① in Figure 2).
2. Break down the content into parts and implement them on the Docker container. These parts can be combined by dividing the contents into parts, and scenarios can be configured according to the needs (② in Figure 2).
3. Since security education involves many case studies, creating flexible exercise contents and scenarios is essential. *Container Manager Server* develops a management system construction technology that combines flexible curriculum and exercises scenarios to assemble each service according to the situation (③ in Figure 2).
4. *Contents Generator Server* provides an editing system for creating content. Since the number of ways to combine each part is enormous, an ecosystem should be created to enable joint development. The joint developers can do this at spatially distant points (④ in Figure 2).

5. Ensure virtual network technology is grouped into isolated networks for each class/scenario. Need to validate an architecture that integrates attacks and defenses of the same class to be performed on the same virtual network (⑤ in Figure 2).

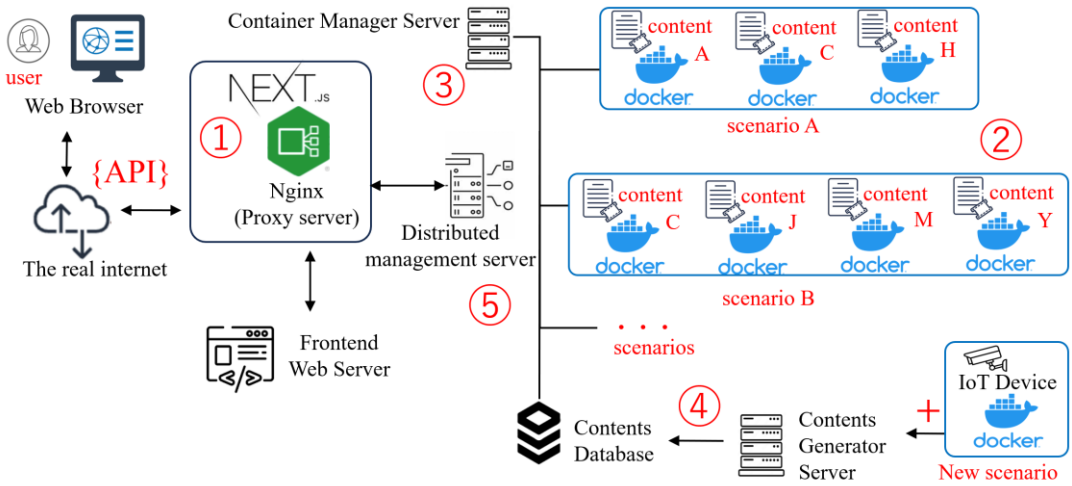

Figure 2: Architecture of the proposed system

4.2 Processing of Data Flow

The processing of data flow in the proposed system is controlled by launching each management system. Figure 3 shows the conceptual diagram of each management system.

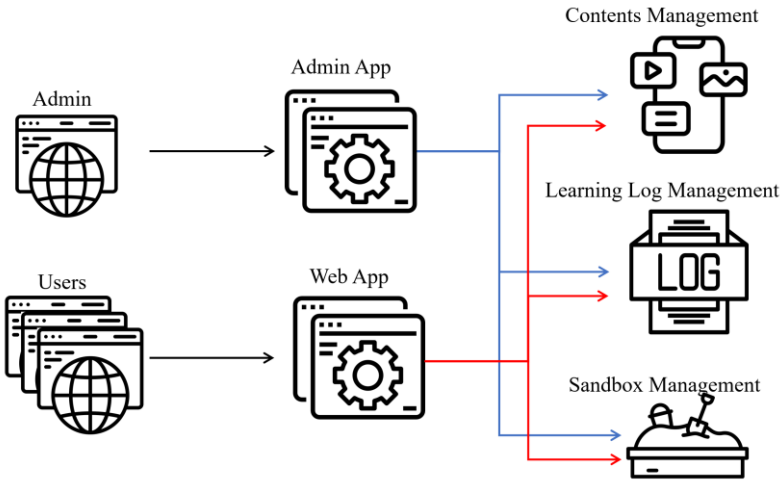

Figure 3: Concept of each management system

For scalability of future development, the applications controlling each data management system were developed separately for administrators (*Admin App*) and users (*Web App*). Administrators and users connect to each App via their web browser. Through each App, users can use

the system's functions, and each management system processes the data generated by each App. The critical process is maintaining and managing the relationship between the information generated by the administrator and the users who use it.

The management systems are roughly divided into *Content Management System*, *Learning Log Management System*, *Sandbox Management System*, and the APIs of these systems are called from the respective applications. The API of each system is called from each application. The words used in the proposed system will be explained, followed by a description of each method.

Sandbox: An isolated space is provided to users where they can safely conduct exercises. It is run in the container, which is loaded with the scenario for the cybersecurity exercise. Also, each sandbox works on a separate container.
Scenario: Refers to the exercise content that the user experiences related to cybersecurity problems. The proposal system provides some scenarios, but other new scenarios were presumed to be built on an ecosystem with other developers or teachers. To keep the ecosystem, the proposal system provides the content editor system to make new scenarios or quizzes for the study.
User Agent: A Linux container for users to access when they experience the exercises.
Course: A collection of exercise content for users to experience. Courses will be created for each exercise content area and difficulty level. The course administrator can edit the existing course or build a new course by combining the scenarios.
Section: A single piece of content in a course. Multiple sections make up a course. A section is divided into three types: sandboxes, quizzes, and explanatory articles.

4.2.1 Content Management System

The content management system manages the content required for the exercises. The main functions include creating, registering, and deleting exercise courses, sections, and explanatory articles, as well as an editor function for submitting descriptive articles. The operator provides exercises to users via this system.

4.2.2 Learning Log Management System

The learning log management system manages users' learning record data. The learning log management system manages user learning data by linking user IDs with the course IDs of exercises users have studied. An API for the management screen will be implemented.

4.2.3 Sandbox Management System

The sandbox management system will manage sandbox activation, deactivation, and resource usage. The sandbox management system will also issue URLs for users to access sandboxes that they have activated. The sandbox management system's design and activation flow are shown below.

The data created with user, course, and exercises is saved in JSON type. When a user returns, the system refers to this JSON file and loads the environment for each user.

Figure 4 shows a simplified diagram of the design of the part of the sandbox management system shown in Figure 3 and the flow of sandbox activation. The API is developed using Express, a JavaScript library, and the sandbox is activated and deactivated in response to requests from the Web App. When startup, a Docker command is executed on the server where the sandbox will be started (① in Figure 4). Information on the startup of the sandbox is recorded in the in-memory

database Redis in the reverse proxy server for user access (② in Figure 4), and the KEY is returned to the Web App. Then, the user can access the sandbox by including the KEY in the URL (③ in Figure 4). And can access the reverse proxy (④ in Figure 4).

The sandbox management system is the core of the proposed system. Since access is always concentrated during user exercises, it is necessary to take appropriate measures such as load control and logging.

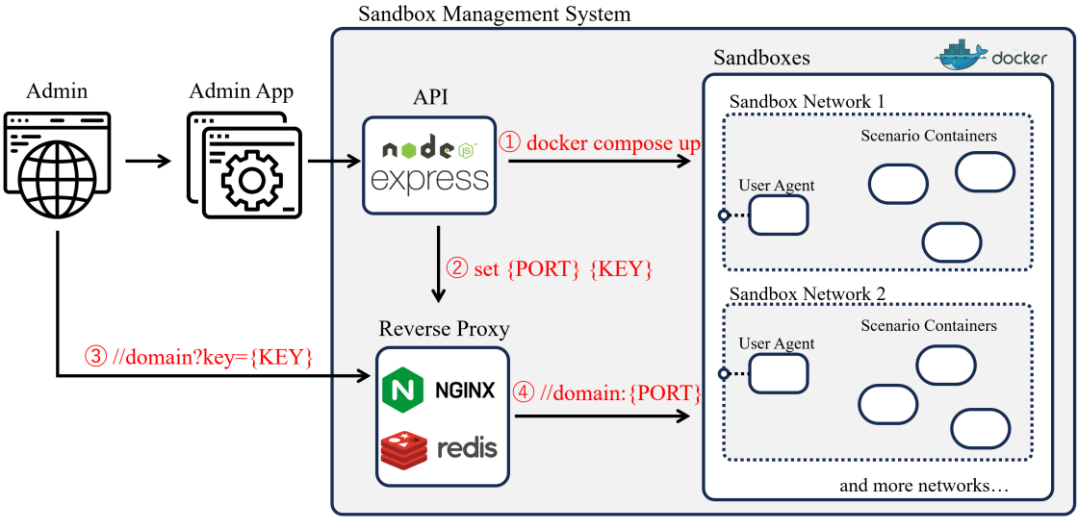

Figure 4: Sandbox management system design and sandbox activation flow

4.3 System Implementation

For the proposed system, a demonstration system was first developed equipped with scenarios for basic security exercises, and the operation of the attack and defense exercise programs was verified. Since attack methods constantly evolve, having all scenarios at the start of learning attack and defense is impossible. It is essential to improve the scenarios as the attack methods evolve continuously. Continuous improvement starts with a minimum number of exercise scenarios and increases the number of scenarios as needed.

Since the feedback from the previous CyExec research often resulted in confusing environment construction, the current proposal focused on basic research, especially for building a hacking lab on a cloud environment (so-called "virtual environment Cyber Range construction on a server/client"). First of all, the proposed system has created a method for efficiently building Kali Linux on multiple containers in the construction of a distributed system, a database server for collecting the results of each exercise, a method for personalizing educational content to enhance learning effectiveness, and an environment for personalization through text analysis. Also, the proposed system tried to build the environment for personalization by text (generated JSON data from users during the exercise) analysis.

Figure 5 shows the implementation environment of the proposed system. A simple exercise is done by providing a server-side terminal environment over the user's local browser ((a) in Figure 5). Full-scale exercises are performed by implementing a server-side Kali Linux environment over the user's browser ((b) in Figure 5).

Providing a Kali Linux environment (with exercise scenarios) implemented in a container on

the cloud (server) side to the user's browser requires considerable resources on both sides (server and local). Therefore, the proposed system solved this problem by implementing only a relatively light terminal environment for simple training scenarios in the user's browser. The administrator can choose which environment to use for the exercises through the editor.

(a) Terminal-enabled environment

(b) Environment with Kali Linux implemented on the cloud

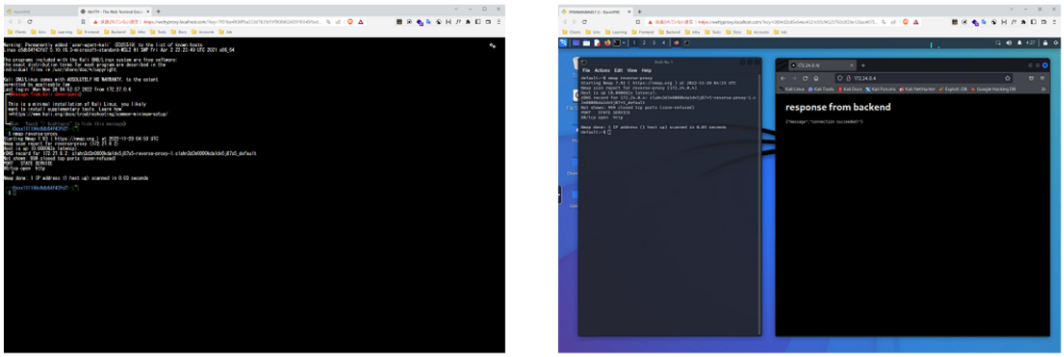

Figure 5: Example of exercise operation on the proposed system

Additional content can be easily added using the editor environment in the system (Figure 6). Administrators or registered teachers can use the editor provided by the system to edit existing scenarios or create new ones. Of course, the contents of the course within the scenario((a) & (b) in Figure 6), the creation of learning questions((c) in Figure 6), the design and editing of learning content((d) in Figure 6), etc., are all possible, and the created content can be shared.

(a) Create new course

(b) Add section

(c) Add question

(d) Content creation

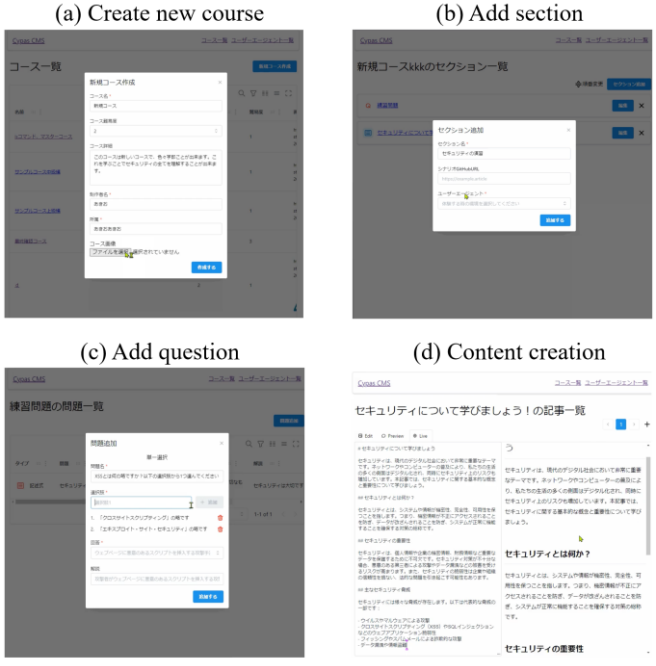

Figure 6: Create new scenarios by combining each learning content

Figure 7 shows an example of a phishing site, one of the exercise contents. The user is tricked and inputs his/her user ID and PW on a spoofed site ((a) in Figure 7). The spoofed site forwards

the user to the actual website, but the user has not been logged into the actual website ((b) in Figure 7). This exercise confirms that the user's ID and PW entered by the user are stored in a JSON-type file on the spoofing site and that personal information has been leaked ((c) in Figure 7).

Figure 7: Example of phishing site

Figure 8 shows an example of an XSS (cross-site scripting) exercise, another typical security exercise implemented in the proposed system. In this way, the proposed system aims to provide exercises appropriate to the user's level, ranging from practical attacks to finding a simple PW from the exercise website by programming.

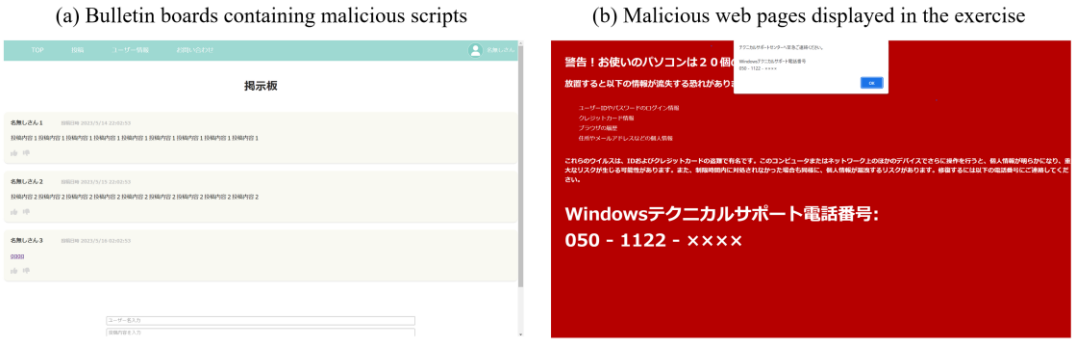

Figure 8: Example of XSS

4.4 Final Goal and Features

The final goal of this research is to research new technology proposals necessary for constructing an information security training system using virtualization technology and container technology on the cloud. To achieve this goal, the following technologies are targeted for development.

1. Establish a technology to securely connect to an environment built in a virtual space on the cloud via a web browser and this mechanism and verify its effectiveness.
2. To confirm the microservices technology to implement, distribute, and manage contents on Docker containers and to verify the effectiveness of the proposed microservices architecture.
3. Propose a technology to distribute and control security attacks and defense exercises

likely to cause system vulnerabilities on an isolated network.
4. Develop a technology to build a virtual security attack/defense exercise network on the cloud for each individual or class unit.
5. Propose a flexible curriculum assembly and a collaborative use/development environment based on an ecosystem.

This proposal is to build a Cyber Range based on cloud technology that can be exercised as long as the local side web browser runs. This enables information security education in the same training environment as the conventional Cyber Range but without local infrastructure or system construction. In addition, personalization and specialization of learning contents per individual and per class can be expected. By the final goal, the proposed system has the following features.

1. Aim to develop service provision based on the SaaS concept on the server side through OSS. There is no need to configure the local environment, reducing the cost burden and facilitating easy implementation.
2. Proposal of curriculum by multiple educational institutions, joint development of scenarios, and construction of an open environment system that can be used jointly.
3. Imaging each scenario and making it a microservice on a container. By partitioning, each scenario can be assembled from parts to enable exercises tailored to the learner's level. Eliminate development burdens.
4. Since the Cyber Range environment is a virtual space built on the server side, other systems and networks will not be affected.

5 Conclusion

This study proposed the development of a cloud-based cyber attack and defense exercise system that enables practical cybersecurity exercises and experiences by using the Cyber Range environment built in a virtual space on the cloud via a local web browser.

This study (1) developed a SaaS platform that can be connected to the cloud-based training environment via a web browser. (2) Enabled users to interactively respond to attacks and defenses in a virtual space on the cloud in an internet environment. And (3) The curriculums are structured based on microservices to be configured flexibly. This proposed system is expected to provide "an environment that facilitates the implementation of information security education" by considering not only the users' but also the operators' perspectives while solving the problems of conventional information security education.

In the future, to use server resources efficiently, the proposed system is considering optimizing the use of server resources by optimizing Sandboxes in Figure 4 as a server resource management system.

References

[1] White Paper on Science, Technology, and Innovation 2021 (Provisional Translation) Toward Realizing Society 5.0, white paper, MEXT (Ministry of Education, Culture, Sports, Science and Technology, Japan), 2021.

[2] E. Stolterman and A. C. Fors, "Information Technology and the Good Life," Information Systems Research, vol. 143, 2004, pp. 687-692.

[3] M. M. Yamin, B. Katt, and V. Gkioulos, "Cyber ranges and security testbeds: Scenarios, functions, tools and architecture," Computers & Security, vol. 88 (2020), art. no. 101636; doi.org/10.1016/j.cose.2019.101636.

[4] D. Fenton, T. Traylor, G. Hokanson, and J. Straub, "Integrating cyber range technologies and certification programs to improve cybersecurity training programs," The Challenges of the Digital Transformation in Education, Springer, 2019, pp. 632-643.

[5] K. E. Stewart, J. Humphries, and T. Andel, "Developing a virtualization platform for courses in networking, systems administration and cyber security education," Proc. the 2009 Spring Simulation Multiconference, 2009, pp. 1-7.

[6] R. Beuran, T. Inoue, Y. Tan, and Y. Shinoda, "Realistic cybersecurity training via scenario progression management," Proc. IEEE Eur. Symp. Secur. Privacy Workshops, 2019, pp. 67-76.

[7] N. Maki et al., "An Effective Cybersecurity Exercises Platform CyExec and its Training Contents," International Journal of Information and Education Technology, vol. 10, no. 3, 2020, pp. 215-221.

[8] R. Nakata and A. Otsuka, "CyExec*: A High-Performance Container-Based Cyber Range With Scenario Randomization," IEEE Access 9, 2021, pp. 109095-109114.

[9] NRI Secure Technologies Ltd., "NRI Secure Insight 2022," Feb. 2023; https://www.nri-secure.co.jp/download/insight2022-report.

[10] ISC2, "ISC2 Cybersecurity Workforce Study," 2023; https://www.isc2.org/research.

[11] Information Security White Paper 2023, white paper, IPA (Information-technology Promotion Agency, Japan), 2023.

[12] IPA (Information-technology Promotion Agency, Japan), "Top 10 Threats to Information Security in 2023," March 2023; https://www.ipa.go.jp/security/10threats/ps6vr70000009r2f-att/kaisetsu_2023.pdf.

[13] National Police Agency (Japan), "Threats to Cyberspace in 2022," March 2023; https://www.npa.go.jp/publications/statistics/cybersecurity/data/R04_cyber_jousei.pdf.

[14] R. Beuran et al., "CyTrONE: An integrated cybersecurity training framework," Proc. the 3rd Int'l Conf. Information Systems Security and Privacy (ICISSP 2017), 2017, pp. 157-166.

[15] G. Erdogan et al., "Developing cyber-risk centric courses and training material for cyber ranges: A systematic approach," Proc. of the 7th Int'l Conf. Information Systems Security and Privacy (ICISSP 2021), 2021, pp. 702-713.

[16] IPA (Information-technology Promotion Agency, Japan), "Vulnerability Experience Learning Tool AppGoat," Aug. 2023; https://www.ipa.go.jp/security/vuln/appgoat/index.html.