2019 IEEE European Symposium on Security and Privacy Workshops

Becoming Invisible Hands of National Live-Fire Attack-Defense Cyber Exercise

Joonsoo Kim, Youngjae Maeng, and Moonsu Jang
National Security Research Institute, South Korea
{joonsoo, brendig, moonsujang}@nsr.re.kr

***Abstract*—This paper provides an extensive discussion of the design process for a new format of national live-fire exercise, called Cyber Conflict Exercise (CCE). CCE targets a real-time battlefield drill between red teams (RTs) who try to penetrate a multi-level organization network and blue teams (BTs) who try to defend against them. This is also a competition-based exercise between red teams and between blue teams. Exercise management teams (WTs, white team) want to be invisible hands to provide a dynamic and interesting battlefield experience to both RTs and BTs with balanced set-up. This paper discusses the challenges that need to be addressed in developing and operating CCE where WTs have little control over RTs or BTs. The devised technical and operational solutions to tackle these challenges are discussed.**

***Index Terms*—cybersecurity exercise, large-scale exercise design, cyber range technologies**

I. Introduction

Cybersecurity is becoming more and more crucial to national security. To strengthen national cybersecurity capabilities, it is necessary to train young security talents, so-called national cyber defenders or warriors, who will fight against future cyber threats. Since experience has a correlation with time, nurturing young national cyber-talents requires an advanced cyber exercise program that simulates real-life experiences [1] [2]. It should provide an opportunity to practice using their knowledge and skills in realistic simulated cyber crisis situations and to build up a variety of experience quickly. It is also an opportunity to assess the capabilities of national cyber-talents and to verify their comprehensive cyber crisis response capabilities. A network of these national talents should be also set up through national cyber exercises where they can develop together and share information.

Cyberspace is a new space of conflict in which all the national capabilities of public-private-military-academia must be gathered to face threats. In this work, we attempted to develop a national cyber exercise that could bring together national talents from any relevant sectors, test their competencies and develop the deficient capabilities. In particular, we aimed to develop a new cyber exercise model that would suit both the young white-hat community and public-sector cyber professionals working for national cyber safety.

This paper provides an extensive discussion on the design process for a newly developed Cyber Conflict Exercise (CCE). CCE targets a real-time battlefield drill between red teams (RTs) who try to penetrate a multi-level organization network and blue teams (BTs) who try to defend against them. This is also a competition-based exercise between red teams and between blue teams.

Each red team should do step-by-step intrusion by pivoting through compromised machines. Red team missions (or injects) are designed mostly based on the actual cases. Vulnerabilities, mis-configurations, or malwares/backdoors were pre-built into this game network.

On the other hand, blue team should take pre-emptive action through threat analysis or execute prompt and effective incident responses. Information sharing is promoted to enable teams to cooperate to respond better.

As far as we are aware, CCE is a rare case of a large-scale exercise in which both red teams and blue teams have autonomy under no control of the exercise organization team at the same time. This paper presents an interesting case study and discusses potential problems inherent in this unique exercise format as well as technical or operational solutions for them. CCE is also a polymorphic cyber exercise that can be reused in other cyber exercise formats.

II. Background

Capture the Flag (CTF) competition is the most widely known cybersecurity exercise format. It has two main styles: Jeopardy! and Attack/Defense [3]. These CTF competitions are fundamentally offensive-technology-focused competitions.

The Jeopardy-style CTF is a contest in which one tries to solve the given problems as quickly and as much as possible. The types of problems are Web, Pwnable, Crypto, Digital Forensic, Trivia, and so forth.

In an attack/defense style competition, such as DEF CON [4], each team is given a daemon, or a (virtual) machine, or a small network to defend on an isolated network. Most of the problems in Attack/Defense CTFs are predominantly in the Pwnable type. One must attack the network daemons or running on the server. The executable binaries can be obtained from their own team server.

CTF is a competition-based exercise that requires advanced hacking techniques focused on acquiring service privileges by exploiting vulnerabilities. The defensive elements in a CTF exercise are extremely limited. It is also difficult to find a CTF exercise that provides a realistic penetration testing target that is similar to the actual work environment. Therefore, CTF may not be a suitable exercise model to enhance the national comprehensive cybersecurity capacity that is directly applicable to defend national critical information infrastructures.

DOI 10.1109/EuroSPW.2019.00015

Cyber defense exercises (CDXs), on the other hand, focus more on realistically simulating field system networks and providing BT participants to face and respond to diverse cyber incident scenarios with their best efforts. *Locked Shields* (LS) [5] hosted by the NATO Cyber Cooperative Defence Centre of Excellence (CCD COE) and *Cyber Europe* [6] hosted by the European Network and Information Security Agency (ENISA) are examples. In the LS exercise, to strengthen the comprehensive response capabilities of BTs, various game elements, such as media, law, and strategy, are combined on top of the technical hands-on exercise.

Crossed Swords (XS) [7] [8] is one of few publicly known offensive exercises. XS, organized jointly by NATO CCD COE and CERT.LV since 2016, is an annual technical red teaming cyber exercise to train penetration testers, digital forensics experts and situational awareness experts.

CDX is a good model for the main CCE audience of the exercise; it has cybersecurity or ICT workforce from public-sector, military, or critical infrastructures as the BT participants. In addition, CCE attempts to create an offensive skill-based exercise that provides an realistic penetration testing experience by inviting young national white-hat community to participate as the RT participants.

By sharing the exercise game-net that RTs should penetrate into and BTs should protect against incoming cyber attacks, CCE is intended to provide the most dynamic exercise format. CCE planners also expect to use this exercise as the platform to collect the most realistic data on cyber conflicts and produce interesting analytic research products that should be beneficial for the cybersecurity community.

III. Cyber Conflict Exercise Overview

The main components of CCE include the exercise participants, exercise contents (game-net, attack missions, defense missions, exploit scripts, availability checker, etc.), exercise scenario, exercise platform, and so on. Fig. 1 shows the conceptual diagram of the CCE components.

The RTs and BTs are assigned roles according to the exercise scenarios. The exercise management team (WT) provide proper information and education to get the players ready, monitor the progress of the exercise, fix issues on-the-fly during the exercise, score the teams' performance, provide feedback and lead the after-action review sessions.

The virtualized game-net is configured based on the base scenario to provide systems and services of the BT network in a multi-level organizational network that simulate the actual government or critical infrastructure network.

WTs implant vulnerabilities and mis-configurations that RTs can exploit to penetrate into the BTs' networks. A mission statement that describes the target machine and the minimum information needed to solve each mission is provided to RTs. WTs use every available tool not to let the game go in favor of either RTs or BTs. They want to be invisible hands whose main objectives are to provide a dynamic and interesting battlefield experience to both RTs and BTs with balanced set-up. This is a very challenging task where there is no definite correct answer. The CCE platform should have various functions that are necessary to support the WTs' tasks.

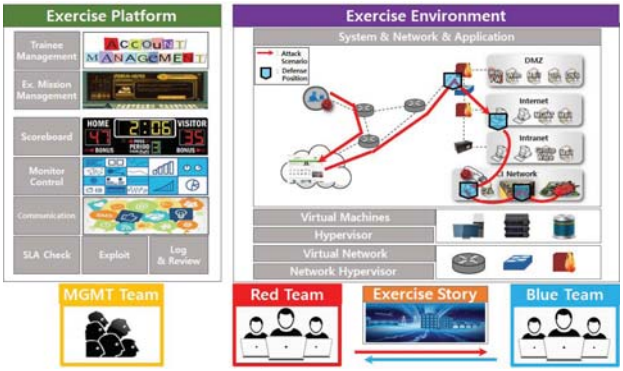

Figure 1. A Conceptual Diagram of CCE Elements

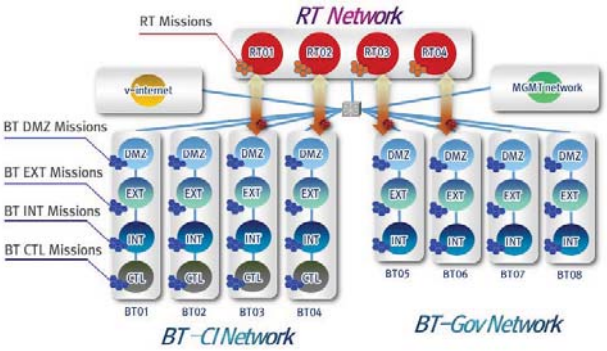

Figure 2. CCE Game-net Overview

In this section, each element will be examined in more detail.

A. CCE Game Network (Game-net)

CCE is a live-fire attack-defense exercise between multiple RTs and multiple BTs. The CCE game-net is composed of an RT network, a BT network, a management network and a virtual internet (v-internet) that simulates the real-world internet as shown Fig. 2. A virtualized-based CCE platform has been developed for automatic deployment and configuration of the CCE game-net.

The BT network is composed of 4 different networks.

1) **DMZ**: a DMZ zone that host public web services, DNS, mail servers, etc.
2) **EXT**: a work environment that can access the v-internet
3) **INT**: a logically isolated intra-net for government offices or critical-infrastructures
4) **CTL**: an isolated ICS/SCADA control network

Systems and services are configured to enable exercise scenarios that are developed based on actual cyber incidents or vulnerabilities/mis-configurations that are observed from the real field environment. Organizational security policies, such as network isolation, wireless services, Bring Your Own

Device (BYOD), outsourced network management, and so on, are considered in the scenarios as well.

Making the platform scalable to support any number of participating teams was one of the main objectives. One of the restrictions in this regard was that an ICS/SCADA exercise platform that is hard to be virtualized is not easily or cheaply reproduced for each BT. The initial plan was to connect the CCE game-net with the existing critical infrastructure cyber range. However, it could provide realistic critical infrastructure environment only for 4 BTs which will limit the number of BTs participating from the critical infrastructure sectors. (Note: There was a separate study done later to develop a scalable critical infrastructure simulation platform for large-scale exercises. However, it is not within the scope of this paper.)

Considering the practical constraints, it was determined that 4 RTs and 8 BTs would be the basic set-up (hereafter, [4:8]-battle-setup). BTs will be divided into 2 different groups. One BT group (BT-CI, 4 teams) will be recruited from the national critical infrastructure workforce whose network includes the simulated ICS/SCADA control network. The other BT group (BT-Gov, 4 teams), in the mean time, will come from the rest of the public sector, who will focus on defending the ICT systems in the generalized government and network during the exercise. However, cooperation and information sharing between two groups are allowed and encouraged,. Note that this is proposed for the first CCE set-up as a test-run and it should be flexible in supporting different numbers of teams.

B. Red Team (RT)

1) Main Target: RTs mainly comprise participants of national cybersecurity elite nurturing programs who have participated in national or international CTF competitions. They are secondary students, undergraduates, or promising young professionals in their early 20s.

2) Required Skill Level: Due to the nature of CCE, the RTs must have a certain level of skill so that cyber attacks can be carried out at an appropriate rate. If the RTs get stuck, the tasks assigned to the BTs will be limited, which will limit the effectiveness of the exercise as a whole. The WTs can also provide RTs with hints and tools to help them when the exercise progress becomes much slower than planned.

RTs should understand how the network is organized and how each configuration can be accessed or manipulated. Some networks are not directly accessible from the RT network. The INT or CTL network can only be accessed through machines in the DMZ or EXT. RTs must understand how to exploit known vulnerabilities to penetrate, perform privilege escalation, remain persistent in the compromised machines, and how to pivot through compromised machines. Since it is virtually impossible to attack using only zero-day vulnerabilities for a short time during the exercise, the game-net is composed of systems and software that are widely used in the actual work environment and have well-known more than one-day vulnerabilities or misconfiguration.

In CCE, RTs often perform tasks using only a minimum amount of knowledge of the target in the format of a mission statement. They need to follow the general cyber kill chain [9]. They need to do reconnaissance by executing a network scan, to recognize services, to create a deliverable payload, to exploit the vulnerability to penetrate into the BT's machines, to execute code or install malwares on the victim's systems and remotely manipulate the victim's systems based on the given RT injects. Some of the missing bridges between steps can be supplemented by the WTs in the form of hints or through the preset environment. While some vulnerabilities must be found through binary analysis that requires reverse-engineering techniques, in this exercise set-up, RTs should take different approaches than those take for general CTFs.

3) Team Size: Considering the [4:8]-battle-setup and the areas of knowledge and skills (e.g. network, web, OS system, performing attack work and maintaining the attack situation), five members per team is initially considered an appropriate number. Of course, one person can work on one or more areas. Moreover, the role assignment between team members can vary and team size can be larger. Practically, the difficulty in recruiting personnel to participate in exercise by being free from the busy work situations is also considered.

C. Blue Team (BT)

1) Role: BTs play a role to defend the multi-level organizational network of critical infrastructure or national public institutes. BT should take pre-emptive action through threat analysis or do prompt incident responses following 5-step procedure: detection, initial reaction, analysis, recovery and security enhancement. Information sharing is promoted to enable teams to cooperate to respond better. Countermeasures that each BT executes should not affect the basic services. At the same time, situation reports should be submitted to the top-level officials (that WTs play) to brief them regarding the overall situation. CCE will be continuously enhanced to provide challenging and comprehensive tasks that BTs should complete to prevent and respond to the cyber crisis.

2) Main Target: The main target for CCE BT participants are the ICT system administrator and cybersecurity officials who work for the public-sector, government, or national critical infrastructure organization. The BT scenarios or injects are designed and implemented to consider their daily working environment.

3) Required Skill Level: BTs should be able to manage all the given tasks as described in the BT's role. They should be technically equipped and be able to work as a team. In addition, all the necessary measures in accordance with the actual situation including so-called soft-skills, such as situation reports and public or media relationships, are also required.

4) Team Size: To consider the practicality in recruiting participants and fairness between BTs and against RTs, each team consists of 4 to 6 persons.

D. Exercise Management Team (WT)

1) Role: During the exercise, the WT does everything that is necessary to keep the exercise going as planned: e.g. pro-

viding pre-exercise orientation or training, presenting exercise injects both for BTs and RTs, monitoring exercise progress, trouble-shooting on-the-fly, scoring, analyzing logs, monitoring cheating or rule violation, reviewing reports, providing hotwash or after-action-review, managing award-ceremonies, and so forth.

2) Importance of Automatic Scoring: There is a real difficulty in securing the personnel and resources needed to prepare and operate a large-scale exercise. It is, therefore, necessary to enable an automated scoring system as much as possible. Using flags to detect the RT's accomplishment of each mission as a typical CTF model is one obvious option. Although there may be disagreement about reality aspects, it becomes possible to run the exercise with great efficiency in comparison to manual scoring, which requires huge manpower. It can also present clearer goals for RTs and BTs.

However, if manual scoring is inevitable, such as report scoring, the content and duration of the exercise injects should be planned ahead so that short or long feedback can be provided along with the score to each team.

E. Exercise Story

A high-level scenario, which is the background through which the exercise is conducted, is defined as an exercise story. This is based on the nature of the organization that the BTs belong to (e.g. 'X' Electric Power Corporation, the information protection officer of the 'Y' ministry, or the national data network service provider), the organization's information protection policies, damage caused by cyber attacks (e.g. service down of 'Z' agency's public web service due to DDoS attack, power outage due to the compromised control network), the intent or goal of cyber attackers (e.g. a hacker group aiming for monetary gain, insiders who are disgruntled with institutions, State-involved attack with political objectives), and so on. A plausible and realistic exercise story is the main instrument for enhancing the engagement of the exercise participants. Injects provided during the exercise serve as a main tool to secure the realism or verisimilitude of the exercise story. It is also assumed that a RT performs various threat assessment or penetration testing considering this background.

F. Exercise Content

In a broad sense, exercise content is everything that constitutes the CCE. Here, we briefly review the exercise content comprises the technical components of the exercise.

1) Attack Scenarios: To simulate realistic future cyber threats, system vulnerabilities in design, implementation or configuration are developed and implanted in the game-net. Necessary attack tools to exploit those vulnerabilities are also prepared for the RT's environment.

Attack scenarios are collected and selected on the basis of the most recent issue or the most frequently occurring incident type. Fig. 3 shows an example of an attack scenario that combines multiple cyber incidents that have occurred. First, a PC in the EXT becomes infected with a malware after accessing a malicious website from a vulnerable web

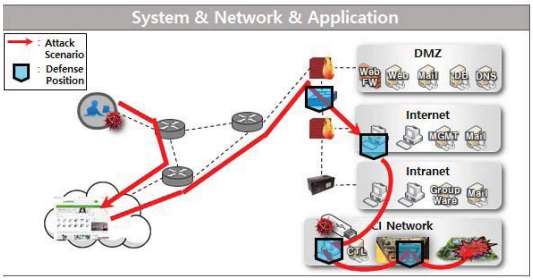

Figure 3. A CCE attack scenario example

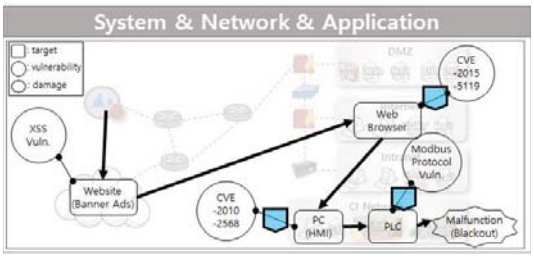

Figure 4. A conceptual diagram of a CCE attack scenario example

browser (Adobe Flash zero-day vulnerability of 'Hacking Team'). Then, an insider attack through the USB infects the PC in the CTL (LNK vulnerability used for Stuxnet). Finally, the advanced persistent threat (APT) attack breaks down the power grid system by exploiting the authentication vulnerability of the Modbus protocol that is still widely used in the control system network. Some of the steps require the WT as a virtual RT or virtual BT (insider attacker) to perform actions to link the scenario elements. Each scenario element is given as an RT mission to compromise each target machine. This is only one example, but there are many different possible scenarios by mixing up different attack routes to cross network levels.

Fig. 4 shows a conceptual diagram to illustrate the necessary components for the implementation of these scenarios and for deriving defense scenarios. Each rectangle represents a vulnerable system, and the description in each attached circle represents the vulnerability used for the attack. Arrows indicate the attack route used for each system infiltration, and it eventually will cause damage represented by the polygon shape.

2) Defense Scenarios: It is necessary to prepare the BT's anticipated defense scenario that can be taken against cyber attack scenarios. Possible defensive actions and tools for the five step incident procedure against cyber attacks are discussed, developed and validated before the exercise. Even though it is not possible to predict all the actual actions of the BT, it is necessary for the verification of the exercise content and serves as a scoring criterion for the BT.

3) Service-Level Agreement (SLA) Checker: An SLA checking system is introduced to realize a high-degree of freedom exercise for BTs. An SLA is a commitment between

BTs and WTs that certain functionality must work under any circumstance for any service. The automated script continually checks each service in seconds until the end of the exercise (EndEx). If an SLA check fails, it will automatically deduct the BT's score in certain frequency.

BTs, therefore, should carefully consider randomly blocking network communication, patching the systems, modifying the system configurations, and so forth. It can be compared to an actual situation in the critical infrastructure field. If the administrator does not pay attention in modifying the configurations or patching the system, it can cause extensive damage.

An SLA is used as an essential element in attack-defense-style CTFs. However, SLA checks can lead to a major dispute by causing problems during the exercise if they are not thoroughly verified in advance. If the exercise is run without SLA checks, the WT load in manual functionality or usability checking will significantly increase or additional rules should be added to reduce the degree of freedom of the BTs. Neither of these options is desirable.

4) Exploit Script: Automated mission-specific exploit scripts should also be implemented. This allows WTs to check in real time whether the RT missions are still alive. If the RT progress is slower than expected, WTs run the exploits across all the BTs on behalf of RTs.

For some RT missions, if the flag acquisition of one RT is confirmed, it can be set-up to automatically run an exploit that includes a damage simulation function. Accordingly, the BT mission can be given by presenting the detectable damage situation.

5) Scoring Plan: According to the attack and defense scenarios, the RT and BT performance is automatically or manually or semi-manually scored. Evaluation criteria for this is required. Since the RT has a clear criterion of registering an obtained flag, it can be prepared with the location and permission settings of the flags and the score for each mission.

In the case of a BT, a scoring criterion in the form of a checklist is required for the submitted report or inject responses. This can be implemented so that submitted reports can be automatically analyzed and scored based on the prepared checklist. Completely automatic report scoring is not desirable and most likely it is not possible. Eventually, the scoring WT team will have to finalize the BT score. How to develop a systematic methodology in BT scoring is one of the most challenging topics because BT scoring for defense-focused large-scale exercise requires a lot of highly trained or experienced manpower.

6) CCE Mission Manual: A manual for each RT and BT mission should be prepared. It should include the objectives, general attack and defense scenarios, the reason for selecting mission scenarios (e.g. recent cyber accident), how to set up the mission environment, source code, vulnerability information, exploit code, SLA check, possible defense methods, virtual environment specification, and so on. The checklist for scoring will be generated based on this information.

G. Exercise Platform

The exercise platform is a set of solutions that implements all the technical functions needed for overall exercise management. Below is the set of representative functions:

- Trainee registration and management
- Exercise mission management and automatic game-net configuration
- Data collection (exercise information, events, syslogs, network packets, and so on) from network systems or agents
- Exercise monitoring and control center and scoreboard
- Multiple communication platforms (chatting, announcement, ticketing, e-mail, etc.) between WTs and RTs/BTs

Detailed technical implementation of each exercise platform component is not the subject of this paper. However, how to present the missions to RTs and BTs is an important function that should be implemented on the platform. It can be designed in various ways to maximize the merits of exercise operation. It is also important to be prepared to make changes depending on the developing situation of the exercise.

Generally, RT missions will be provided sequentially from the outer-most networks. At the start of an exercise (StartEx), only the missions at the DMZ will be open at the exercise platform. Other RT missions in different networks (EXT, INT, CTL) are disclosed when any mission in a more external network is solved. We assume that the RT has acquired the ability to pivot into a more internal network via the machines that have just been compromised.

On the other hand, all the BT missions should be open from StartEx till EndEx because BT's role of protecting the entire network should not be constrained. However, to push BTs further to achieve the exercise objectives, some additional injects that will require BTs to take actions accordingly will be presented based on the planned schedule or contingently depending on the exercise situation.

H. Exercise Operation Manual

The exercise operation manual should include action plans for all predictable issues that may arise during CCE. It should also provide general procedures for exceptional circumstances.

IV. Technical and Operational Considerations

In this chapter, we discuss several technical and operational challenges that need to be addressed to effectively operate CCE and solutions devised for each.

A. System Virtualization

PCs, servers, and other systems used during the exercise should be configured to be similar to the actual work environment and to have as many vulnerabilities as possible or to become infected with malicious code or backdoor in advance. The systems thus constructed should be automatically deployed immediately before the exercise and should be restored as they were after the exercise. System virtualization is essential for this. However, problems with ICS systems or embedded hardware devices that are difficult to virtualize must be addressed individually.

B. Network Virtualization

Network virtualization needs to be considered in order to flexibly utilize physical network resources such as routers, firewalls, and switches. Virtual LANs (VLANs) are very useful for creating small networks. However, it is difficult to build on a large-scale network, and there is a disadvantage that it lacks manageability and scalability. It is necessary to consider the introduction of technologies such as network functions virtualization (NFV) or software defined networking (SDN) to provide a flexibility in configuring a complex game-net composed of hundreds or thousands of VMs according to the number of participating teams to efficiently allocate resources and to improve network service.

C. Multi-Level Network Configuration

A multi-level BT network has different firewall policies at each level. The default access policies are shown in Table I. Based on these policies, a separate detailed policy can be established for each exercise mission.

Table I
DEFAULT FIREWALL ACCESS RULES

Src \ Dst	v-Int	DMZ	EXT	INT	CTL
v-Int		Allow	Reject	Reject	Reject
DMZ	Allow		Allow	Reject	Reject
EXT	Allow	Allow		Reject	Reject
INT	Reject	Cond. Allow	Cond. Allow		Reject
CTL	Reject	Reject	Reject	Reject	

D. Pivoting

If the RT succeeds in penetrating the BT machine, the RT will bypass the firewall and attack the other server via that machine. However, if the RT spends a lot of time in pivoting set-up during the exercise, the dynamics of the exercise will be reduced. In order to mitigate the temporal limit to some extent, it is necessary to set-up a pivoting environment in advance so that the RT can use the compromised machine as a base machine.

For CCE, SOCKS servers are installed on all mission servers. The authentication feature is activated to disable the proxy server before the mission is completed and the authentication password is obtained for the proxy server.

E. Virtualization of Embedded Systems

It is interesting to reproduce cyber attacks using vulnerabilities in embedded devices. In this case, however, physical systems are required for each team, and it greatly increases the configuration complexity to manage their physical integration into the virtual game-net.

Using the QEMU [10] that supports ARM or MIPS architectures that are widely used in embedded systems, one mission scenario regarding wireless routers has been ported on a VM. The main executable binary and all the configuration files and temporary files related to functionality are extracted from the router and reconfigured according to the virtual system environment. Through this series of processes, the web server of the router operating normally on the virtual physical system is set-up, and the known vulnerabilities are successfully exploited.

F. Delayed Flag Registration

For CTF-style exercise, it is a widely accepted strategy that RTs do not register the obtained flags until just before the EndEx. However, in CCE, this causes the BT exercise not to proceed smoothly.

To solve these problems, a flag system that is periodically updated is introduced Each flag is updated with a predetermined period of time. When an RT acquires and registers a flag, it validates the flag in terms of the registration time. Flags cannot be registered after the expiration time. The validation time for flag registration must be at least two times longer than the generation period of the flag. This prevents the flag being updated between the time of flag acquisition and the time of flag registration.

To generate the hash value for the flags, the variables listed in Table II are used.

Table II
PARAMETERS USED FOR GENERATING FLAGS

Parameter	Description
Flag Salt	To prevent collision and input guessing
BT Number	The target BT where flags reside
Mission #	Mission ID
Timestamp	Timestamp that determines the valid time of the flag

Then, flag validation is done based on the domain variables as presented in Table III.

Table III
PARAMETERS USED FOR VALIDATING FLAGS

Parameter	Description
StartEx Time	This is the reference point for flag generation cycle synchronization.
EndEx Time	After EndEx, the flag generation request will be rejected.
Flag Gen. Period	Interval of flag generation
BT List	List of Target BTs
Mission List	List of RT Missions
Flag Salt	Hash salt used for flag generation
Flag Effective Time	The maximum time allowed to validated flags based on the flag generation time

G. Balanced Attack across BTs

The RTs can freely set the targets in a way that may cause unfairness of incoming RT attacks across BTs. The skill levels

of participating teams, the number of BTs, or other factors can make the situation worse. If such a situation arises, there may be a disruption in the process of fair and smooth competition between exercise participants.

Though it is difficult to completely solve the problem, there are incentives given to RTs to induce balanced attacks to alleviate the situation. The earlier RTs succeed in flag registration for each RT mission on a particular BT, the more points the RTs will get. Also, if attacks against all the BTs are successful against the same BT mission (the same target machine), a bonus point will be given (a.k.a 'Bingo-Rule').

H. Periodic Flag Update

In a real situation, when a server is occupied by an attacker, an attacker frequently uses the server later as a stopover point or re-occupies it continuously to obtain more information. To reproduce this situation in the exercise format, it is necessary to induce attackers to regularly re-attack the same RT mission targets.

Therefore, periodic flag updating for each RT mission target is introduced. If an RT re-compromise the server after a certain period of time, they will acquire a new flag and get additional points. This requires automated flag updates and validation system updates.

Considering that the flag update tasks are similar to the management of large server farms, the tools used for the service deployment process are considered to enable this. There are many available options, such as Ansible, Puppet, Chef, or Fabric [11].

I. Check for Patched Vulnerability

After configuring the environment, the WTs need a series of verification processes to check each mission server to determine whether service is properly running or it is actually attackable. As the above-mentioned problem, it is necessary to build up a service that automatically checks the mission status and notifies the WTs if the verification fails.

J. Deleting Work History

There are traces of setting up the environment, exploit verification, and checking the countermeasures while configuring and verifying each exercise mission. This should not be exposed to exercise participants. There were cases in which CTF participants easily solved problems using the remaining methods of logging, or recovering deleted files and extracting exploits. To prevent this problem, it is necessary to completely clean up the administrator's work history.

It is necessary to delete the file system entry as well as the file data. Some important items to be deleted are summarized below.

- Shell history
- Daemon log related to the problem
- System log
- Other test scripts

Storing a VM snapshot with traces removed can also eliminate unnecessary repetition.

Table IV
AN EXAMPLE FOR CCE SCHEDULE

	Day 0	**Day 1**	**Day 2**
0900-1300(4H)	Orientation	Ex. Phase 1	Ex. Phase 3
1300-1400(1H)	Lunch	Lunch	Lunch
1400-1600(4H)	Preparation	Ex. Phase 2	Ex. Phase 4

K. Other Challenges

Several issues that CCE format might cause and that could not be easily resolved were identified during the exercise preparation phase. One example is potential collusion between RTs and BTs who only focus on winning the prize. Clear and decisive rules were established to punish such dishonest conduct and were explained to the exercise participants beforehand. However, more efforts is needed to develop technical solutions along with operational methods and to root out such intractable problems while accumulating experience.

V. EXERCISE RESULT

The first test-run of CCE has been executed according to the schedule shown in Table IV.

RTs and BTs who participated in the CCE had difficulties in adapting to the new exercise format early in the exercise phase. As the exercise progressed, it became clear that the degree of commitment and the dynamics of the exercise greatly increased as participants gradually a better understanding of the format. After the exercise, individual surveys were collected and in-depth interviews were done with sampled participants. Exercise participants expressed deep satisfaction with the overall exercise operation and the technical contents of CCE. However, they also provided detailed technical and operational improvement points for future CCEs, such as some inexperienced operational mistakes, some more required documents, exercise preparation time allocation, team size change, and so on. All matters will be considered for further correction and improvement.

VI. CONCLUSION

This paper discussed the challenges that need to be addressed in developing and operating a new national cybersecurity exercise, called CCE, and the designed technical and operational solutions for those challenges. CCE has a parallel competitive format in which WT has no control over RTs and BTs during the exercise execution. Therefore, it takes a great deal of effort to make the WT as invisible as possible during the exercise but to still maintain the overall exercise balance. This paper has summarized and described the design considerations necessary for developing the new exercise format in the preparatory steps from initial design to operation.

Just as exercise is an endless repetition of learning, review, and improvement, so is exercise design and operation. Some additional issues that have been found through multiple CCE operations will be discussed further in our future work. Among

the identified issues, some priority list to solve are (1) uncontrollable CCE game pace, (2) the need to provide a more stable service availability checker, and (3) the need for more balanced attacks across BTs.

REFERENCES

[1] D. Fox, C. McCollum, E. Arnoth, and D. Mak, "Cyber wargaming: Framework for enhancing cyber wargaming with realistic business context," *HSSEDI, The MITRE Corporation*, 2018.

[2] R. S. Dewar, "Cybersecurity and cyberdefense exercises," tech. rep., Center for Security Studies (CSS), ETH Zürich, 2018.

[3] wikipedia.org, "Capture the flag." https://en.wikipedia.org/wiki/Capture_the_flag. Accessed: 2019-02-13.

[4] C. Cowan, S. Arnold, S. Beattie, C. Wright, and J. Viega, "Defcon capture the flag: Defending vulnerable code from intense attack," in *Proceedings DARPA Information Survivability Conference and Exposition*, vol. 1, pp. 120–129, IEEE, 2003.

[5] NATO Cooperative Cyber Defence Centre of Excellence (NATO CCD COE), "Locked shields 2013 after action report," in *After Action Report*, 2013.

[6] European Network and Information Security Agency (ENISA), "Cyber europe 2018: After action report," in *After Action Report*, 2018.

[7] N. CCDCOE, "Crossed swords." https://www.ccdcoe.org/exercises/crossed-swords/. Accessed: 2019-04-18.

[8] S. Waterman, "The reason a recent international cyber-exercise was so unique." https://www.cyberscoop.com/nato-cyber-wargames-crossed-swords/, 2018.

[9] E. M. Hutchins, M. J. Cloppert, and R. M. Amin, "Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains," *Leading Issues in Information Warfare & Security Research*, vol. 1, no. 1, p. 80, 2011.

[10] F. Bellard, "Qemu, a fast and portable dynamic translator.," in *USENIX Annual Technical Conference, FREENIX Track*, vol. 41, p. 46, 2005.

[11] P. Venezia, "Review: Puppet vs. chef vs. ansible vs. salt," *InfoWorld*, vol. 21, 2013.