UNIVERSITY OF TARTU
Institute of Computer Science
Cyber Security Curriculum

Allyson Hauptman

Designing Digital Forensics Challenges for Multinational Cyber Defense Exercises

Master's Thesis (30 ECTS)

Supervisor(s):
Patrycjusz Zdzichowski
Rain Ottis
Raimundas Matulevičius

Tartu 2016

Designing Digital Forensics Challenges for Multinational Cyber Defense Exercises

Abstract:

This thesis seeks to design and evaluate a digital forensics challenge for inclusion in a multinational cyber defense exercise. The intent is to narrow down the key skills a state-based organization requires of its digital forensics experts and design and integrate technical tasks that adequately test these skills into a larger cyber defense exercise. It uses the NATO Locked Shields cyber defense exercise as a test case, for which the thesis author joined the digital forensics design team at the NATO Cyber Defense Centre of Excellence in designing and implementing a three day digital forensics challenge. This thesis establishes a series of technical and procedural skills state-based organizations require of their experts, determines ways to test these skills, and develops a scenario-based digital forensics challenge. Using first hand observations, participant feedback, and challenge scores to evaluate the effectiveness of the challenge, it finds that the scenario adequately tested a majority of the skills at the appropriate difficulty level and needs improvement in timing and reporting standards. Finally, it explores ways to improve upon the selected methods and tasks for future exercises.

Keywords:

Digital forensics, NATO, cyber exercise, malware, system forensics, network forensics, forensic reports

CERCS: P170, Computer science, numerical analysis, systems, control

Digitaalse ekspertiisi ülesannete disain rahvusvaheliste küberõppuste kontekstis

Lühikokkuvõte:

Töös kujundatakse ja hinnatakse digitaalse ekspertiisi teemalist ülesannet, mida kasutada rahvusvahelisel küberkaitse õppusel. Eesmärk on keskenduda põhioskustele, mida üks riiklik organisatsioon oma digitaalse ekspertiisi ekspertidelt vajab ja disainida ning integreerida tehnilisi ülesandeid, mis adekvaatselt testivad neid oskusi suuremahulise küberkaitseõppuse raames. See töö kasutab Locked Shields küberkaitseõppust näitena, mille jaoks väitekirja autor liitus digitaalse ekspertiisi arendusmeeskonnaga NATO Cooperative Cyber Defense Centre of Excellence juures, kui nad kavandasid ja rakendasid kolm päeva kestvat digitaalse ekspertiisi ülesannet. See lõputöö identifitseerib rea tehnilisi ja protseduurilisi oskuseid, mida riiklikud organisatsioonid vajavad oma ekspertidelt, määrab viisid, kuidas testida neid oskusi ja arendab välja stsenaariumipõhise digitaalse ekspertiisi ülesande. Kasutades õppusel vahetult saadud tähelepanekuid, osalejate tagasisidet ja ülesande tulemusi, leitakse lõputöös, et loodud ülesanne testis osalejate oskusi õigel raskustasemel ja vajab parendamist ajastuses ning aruandluse standardites. Lõpetuseks uuritakse erinevaid viise, kuidas parendada valitud meetodeid ja ülesandeid tulevaste õppuste tarbeks.

Võtmesõnad:

Digitaalne kohtuekspertiis, NATO, küberülesanne, pahavara, kohtuekspertiisi süsteem, võrgu kriminalistika, kohtumeditsiini aruanded

CERCS:P170,arvutiteadus, arvutusmeetodid, süsteemid, juhtimine (automaatjuhtimisteooria)

Table of Contents

1. Introduction

Cyber security exercises are a cornerstone of an organization's ability to gauge and increase its level of preparedness and technical expertise for countering cyber espionage, theft, and attacks. These exercises include a broad range of areas to test, including: network penetration, file security, public affairs, legal considerations, and data protection. An area of testing that is relatively new to these exercises is digital forensics, the process of investigating cyber incidents. For States this area is extremely important, because the ability to attribute cyber incidents and recover data can mean the difference between war and peace. While States conduct their own internal exercises, cyberspace is an international domain, and this makes multinational exercises extremely important. For these exercises, it is vital that the challenges test the main goals of all of the involved parties, both efficiently and effectively. These exercises are comprehensive and take a long time to plan; each testing area is nested within the larger exercise scenario. Hence, the digital forensics challenge needs make sense within the context of the overall exercise story, fit within a few days of game-play, and accurately test the main skills the States require of its experts.

The main question that this thesis asks is: how can organizations design and implement an effective and efficient digital forensics challenge for a multi-state cyber exercise? This is a complex question, because it requires a series of considerations. First an organization needs to consider the elements of a cyber exercise and what are its constraints. Second, the organization needs to decide what are the most important digital forensics skills to test. Third, the organization has to design a realistic challenge that tests those skills within the context of a larger cyber exercise. Finally, it needs to develop a way of evaluating its design. This last part requires a test case. This thesis uses the NATO Cyber Defense Centre of Excellence's annual Locked Shields cyber exercise. Locked Shields is a good test case, because the digital forensics team could compare its results with the results and experiences from the 2015 digital forensics challenge. The exercise is multinational, comprehensive, and limited in duration. The digital forensics team used the performance and feedback from the exercise designers and participants to judge how well the challenge efficiently and effectively tested the digital forensics skills and if those skills were, indeed, the most import to test.

This thesis author worked with a group of three technical experts at the NATO Cyber Defense Centre of Excellence in designing, implementing, and evaluating the digital forensics challenge for the 2016 Locked Shields cyber exercise over the course of seven months. This thesis resulted in a three day digital forensics challenge that included hard disk image and memory dump acquisition, network and memory analysis, file carving, and forensics reporting tasks. It found that the Locked Shields 2016 design tested nearly all of the most important forensics skills to test with room for development on preparation of the teams, use of anti-forensics, and reporting design. Comparison of results from Locked Shields 2015 to Locked Shields 2016 showed a vast improvement in the design and implementation of the digital forensics challenge in terms of difficulty and timing.

2. Background

In this first section, the thesis author reviews various types of existing cyber defense exercises and the digital forensics portions they include. She then researched and determined the main goals and limitation of State cyber defense exercises.

2.1 State of the Art

Types of Exercises

While the type of exercise on which this thesis is focused is an inter-team competition, many other varieties of cyber exercises exist for which digital forensics challenges are also required. These exercises go from focusing on individual persons to hundreds of participants. The most basic type of exercise is skills improvement, in which a single person or small group of people must perform the exercise in order to increase a specific skill set. For example, a forensic analyst may be required to recover all photographs from a corrupted New Technology File System (NTFS) in order to improve his file carving abilities.

Slightly larger groups can participate in Capture the Flag or Workshop exercises, in which a series of challenges are constructed and a team must retrieve some sort of value or token to prove that they successfully completed each task. In such an exercise, a digital forensics team may need to locate the login credentials of a user who downloaded a corrupted file and also recover the file itself. They also may need to submit the credentials and file hash as the "flags" to the event organizers or just discuss the solutions if in a Workshop forum [1].

At the strategic level, Table Top exercises test the plans and procedures of an organization or a group of organizations (known as Distributed Table Top exercises). In these types of exercises, management level individuals discuss how they would respond in given situations according to established routines and procedures [1]. For instance, a government agency may conduct an exercise to test the authorities it has in responding to a ransomeware incident. No technical solution will actually be implemented, and policies and procedures are more important to the discussion than the technical responses. In a similar vein, Command Post and Building Block exercises occur at the policy and procure level as well but are primarily focused on inter-organization and inter-body coordination. These exercises often occur in phases. Each level of the exercise is started and completed and passed up to the next resolve to resolve their layer of the issue [1]. These exercises are common amongst government agencies in which response to a cyber incident involves several bodies with different organizational policies and authorities.

Additionally, on a more defensive side, there are general training and awareness exercises that organizations give to their users and clients. These can range from employee phishing awareness training to security policy testing. These types of exercises may be relevant to digital forensics in terms of teaching people what information they provide (to websites, servers, ect) is easily recoverable by a malicious actor and how to mitigate against it.

Existing Multi-Actor Cyber Defense Competitions

In recent years digital forensics challenges have been added to several well-known cyber defense competitions. One such competition is the Cyber Olympics, a series of cyber competitions amongst high school and college level students in the United States. The

Cybersecurity Challenge (part of the Olympics open only to players over the age of eighteen) is a hybrid Capture the Flag, competition exercise. The first phase of the challenge is a digital forensics challenge. This phase lasts four hours and requires teams to locate evidence of intrusion and analyze one or more of the following: malware files, memory dumps, hard drives, logs, and network traffic. Teams are scored according to the number of artifacts they find [2]. Note, in this exercise all teams are given the items to analyze, so there are no acquisition tasks.

One exercise well-designed for the inclusion of digital forensics injects is the United States National Collegiate Cyber Defense Competition (NCCDC), which presents competitors with an already constructed network and services that they must defend. The exercise claims to focus on the "more operational task of assuming administrative and protective duties" [3]. Teams are required to respond to outside threats, determine which services are vital, and maintain or remove services as they see fit. This is a good set up for testing procedural and policy level considerations in concert with the technical capabilities of the competitors. Teams are scored according to automated assessments of their services as determined by the services' value by the competition coordinators [3].

Similarly, Cyber Panoply is an exercise that provides each team with a network; however, this competition is a zero-sum game. Teams compete over common resources and services. The competition requires teams to perform both defensive and offensive actions, protecting the resources and services they have under their control and penetrating the network of rival teams in order to gain new ones. As in NCCDC, teams are scored according to automated service scanners [4]. The forensics piece of the challenge (finding intrusions and the sources of those intrusions) is scored indirectly in this way.

One of the most advanced cyber defense competitions in terms of digital forensics scenarios is the National Security Agency (NSA) and Cyber Security Service (CSS) Cyber Defense Exercise, referred to as NSA/CSS CDX. This exercise pits the students at the United States service academies against one another as Blue Teams (defending teams). The Red Team (the attacking team) is composed of approximately forty experts from the NSA and military reserve components. The competition requires each team to configure and defend a network with specified services. In 2010 the competition introduced its first digital forensics challenge. A gray cell of system users simulates issues in the network created by average users, one of whom is operating from a suspicious computer system on which team had to conduct forensics analysis [5]. Teams receive points for maintaining services and detecting and responding to threats [6].

The exercise that this thesis will use as a test case is the NATO Cooperative Cyber Defense Center of Excellence (CCDCOE) Locked Shields exercise. This annual event has reached a size of twenty teams of NATO member and ally military and government organizations. The CCDCOE first ran the exercise in 2010 and it now includes technical, forensic, policy, legal, and media level challenges [7]. The CCDCOE introduced the first digital forensics challenge in the 2014 competition and widely expanded on in the 2015 competition. For the 2015 digital forensics challenge teams were given a PCAP file of network traffic and access to a live virtual machine for analysis [8]. The challenge required teams to analyze a compromised machine and compose a digital forensics report, including acquiring an image of the virtual machine [9]. The coordinators scored the teams based upon the forensics reports they submitted, to include how and what they found in their analysis. They also provided a simple template for the players to use in completion of their reports [8].

2.2 Considerations

Actors Involved in a Cyber Incident

Prior to developing any specific scenario for a cyber exercise it is important to understand the actors who may be involved. There are a wide variety of actors who would be responsible for responding to a cyber incident, including: technical, legal, managerial, and political. These actors play their parts at different points in the incident response cycle, which consists of five phases: [10]

1. Prevent and Protect
2. Detect
3. Analyze
4. Respond
5. Resolve

Starting with the first phase, Chief Information Officers (CIO) and Cyber Security Officers (CSO) are primarily responsible for conducting risk assessments and approving a cyber security policy for an organization. These key documents identify the possible threats, vulnerabilities, and impacts of potential incidents and how to mitigate the risks. To craft these documents, the officers analyze the organization's services, IT assets, client base, network and physical infrastructure, and priorities of upper level management. They need to "think like a hacker" [11]. Specific personnel are assigned to fix or manage these risks depending on the organization's resources and risk tolerance. It then becomes the responsibility of these people to detect a threat if an incident occurs. This would usually be the responsibility of system and network administrators.

These administrators have to quickly take action to isolate the threat and communicate with the CIO and/or CSO about the steps they should take. This involves using memory and PCAP files and system, network, and firewall logs. Essentially, they must detect the precise location and actions of the threat. Once they pass the information up the proper Chain of Command (CoC), the analysis phase begins. In this phase a variety of actors determine what steps should be taken in order to respond to the incident. Business and branch managers consider what human resources are necessary and available to resolve the incident according to the organization's Internal Role plan [12]. This includes primarily staffing and budget information, particularly if there exists a need to outsource. At the same time, the board members and executives (or political leaders if it is a government agency) need to consider the impact to the bottom line and mission of the organization, as compared to the costs of solving the issue. To do this they require an impact analysis and real-time updates on any media coverage of the incident.

There are a variety of actors involved in incident response. The system and network administrators and any special technical personal are responsible for implementing whatever patches or technical fixes have been deemed appropriate. Board members, politicians, and public relations officials must handle communications with media concerning the incident and how it is being resolved. Communication with media should first occur no later than thirty minutes post incident detection. The CIO and/or CSO must communicate with system and network users concerning what actions they can and cannot perform during response to

the incident. It is sometimes also necessary to communicate with specific users whose actions facilitated the incident.

The resolution phase of the incident involves more than just paperwork. It is here that legal and law enforcement entities may become involved. The entities need access to forensics reports, evidence they want to image and/or analyze, software and hardware documentation, system and network logs, and user information. CSO and/or CIO need to communicate with users to prevent a future occurrence, often involving the creation and deployment of user training. Board members, political entities, and public relations officials must handle any additional media fall out, largely oriented around conveying to the media what plans are in place to prevent a future incident. Finally, investors and constituencies need to be informed of the final impact and what is being done to prevent a future incident [12].

Exercise Goals

Digital Forensics experts are expected to possess a broad range of technical capabilities. The discipline of digital forensics includes the acquisition, processing, analysis, and reporting of digital artifacts and evidence. Two well known certification authorities for digital forensics experts are the SANS Institute, which administers the Global Information Assurance Certificate [13], and the International Society of Forensic Computer Examiners, which administers the Certified Computer Examiner (CCE) Certificate [14]. The CCE is a prestigious non-vendor specific forensic certification, used in over twenty-eight countries to validate individuals' forensic competencies. These two certificate programs are designed to test the "core skills required to collect and analyze data [13]." They will be used to establish the professional competencies of digital forensics experts.

To begin with, forensics experts need to perform various acquisition methods. These include physical and remote acquisition. The devices from which forensic images can be acquired include external and internal hard drives, removable media such as Universal Serial Buses (USB), mobile devices, and network storage devices (such as share folders and cloud storage). In recent years, memory acquisition has also become more important, due to the "growing importance of temporary files" [15]. This type of acquisition is also extremely important in situations where powering off the device would result in data loss [15]. Apart from forensic images, digital forensic experts should also be capable of acquiring network traffic [14]. Every time a forensics expert performs a procedure on digital artifacts, they risk corrupting the evidence, and for this reason digital forensic experts must be able to verify the integrity of their acquisitions. Methods to do this include metadata reading and hashing algorithms. Experts must also be capable of demonstrating the reliability of the tools/methods used to acquire the images/traffic through proper procedures, such as safe boot [14].

Once finished acquiring images and traffic, forensic examiners must also obtain additional artifacts valuable to the investigation. These are the files and metadata attributes themselves. Forensic experts need to read file systems for multiple operating systems and perform data carving to obtain files from memory, both allocated and unallocated [14]. In order to properly carve data, forensics experts need to demonstrate the abilities to read and analyze the Master Boot Record (MBR) [13] and other file system indexes and registries using hexadecimal values. This means they must also be capable of the proper installation and use of digital forensic software. In some instances, forensic experts may not be certain which files contain the information they need to obtain. Experts can use strings analysis to search for key terms within a series of directories for specific information, such as "password" [13].

In addition to data recovery, digital forensic experts need to analyze data for the evidence it represents. Through system and network log analysis, forensic experts can demonstrate when and how certain software entered and affected computer systems and networks. This concerns the fundamental issue of attribution. An example would be a forensics expert using network traffic to trace the domain of origin of a Trojan or other form of malware. Another important aspect of attribution includes Prefetch Analysis, which concerns executable file metadata and logging. Digital forensics experts should also be capable of tracking user activity within networks and systems for the purposes of evaluating user account abnormalities or events [13]. This can be useful, for example, to show what user account an attacker compromised and used to access confidential information. It is also useful for determining the source of that compromise, such as users who use the same passwords for multiple credentials. Browser forensics is also crucial to attribution and the patching of vulnerabilities and is another valuable digital forensic skill [13]. For instance, an expert can determine what site a user last visited before a system crash through accessing the history.dat file that Firefox automatically writes [16]. The biggest issue with attribution, however, is that forensic data is only as good as the trust in which others place in its integrity.

Digital Forensics experts not only need to be able to obtain and analyze data but also present it as pertinent, complete, valid, and legal. The European Union Agency for Network and Information Security (ENISA) is the center of expertise in information security for the European Union, both its member states and citizens [17]. A Cyber Emergency Response Team (CERT) from ENISA composed a comprehensive digital forensics handbook that establishes some basic guidelines for the proper handling of evidence during the collection and analysis phases of a digital forensics examination. If the artifacts the expert will be acquiring and/or analyzing will be used as evidence at any point, then it is vital that the expert follows proper procedures. This means that the analyst obtains the data in compliance with applicable law, is qualified to perform the actions he/she performs, and is capable of proving the data's authenticity and veracity [17]. This proof comes from an important set of documents the examiner must be able to produce: the report.

In short, the examination itself should be traceable and repeatable by a third party [14], and thus the purpose of proper documentation of the investigation. One method for achieving this is to use built-in forensic software logging tools and exporting data items to comma-separated (CSV) and text files [18]. This documentation needs to include four key components: case summary, acquisition steps, analysis processes, and conclusions. The examiner must be able to, in simple terms, describe the context and importance of the case to which the examination is relevant. The key here is simplistic. A forensics expert must be able to convey to a non-expert what and how all steps in the investigation occurred. If necessary the report can include a glossary in order to support this goal [18]. Next, the expert must include all actions taken on data objects, including the methods used to preserve the integrity of data and verify their acquisition. In describing the findings of the investigation, the expert must to describe the tools used such that a third party could repeat the experiments exactly. Finally, they need to convincingly summarize the conclusions of the investigation.

In the construction of this report, a digital forensics expert needs to include a few very important aspects, often referred to as audit data. Evidence is only as good as it is presented. An examiner must preserve the audit data and logs during the investigation for use as appendices in the report [19]. During an examination more than one person will likely access the evidence items, and for this reason an expert must properly document the chain of

custody and purpose of access. The forensics expert has to construct a proper time-line of data transport, storage, and analysis to include persons, places, precautions, and actions [19]. This means the expert must understand what qualifies a person to access different types of data (i.e. certified individual having access to network firewalls). All tools used during the investigation need to be verifiable not only as a tool but within the investigation itself. This means the expert must be able to perform and document the calibration of tools for the purposes of the investigation [19]. Finally, in the event that improper procedures or hasty measures are taken, the expert should be capable of explaining within the report why. Such explanations are key to supporting the forensic evidence as pertinent and valid.

Having established what technical and procedural capabilities forensics experts should possess, a challenge design team must also consider what are broader than the aims of state governments. In state-based exercises, the goals of the government drive the skills the exercise should test. State based exercises are unique in that they must prioritize actions against a broader set of goals and take a "full spectrum approach" [20]. They must test a variety of capabilities and functions of personnel and constructs. In 2001 the United States designed the inter-military academy cyber defense exercise previously discussed, NSA/CSS CDX, in which participants had to "design, implement, manage, and defend a network of computers" [21]. Since then this annual exercise has grown to test the broader range of aspects, because state-based exercises need to examine "legal, ethical, forensic, and technical components while emphasizing a team approach" [21]. The team-based aspect is significant here. Digital forensics experts must operate within a team in the exercise, and sometimes may even be responsible for activities beyond forensics due to personnel constraints. Thus, testing a unit's communication channels and efficiency is almost more important than testing technical skills [21].

State-based exercises include a wide variety of participants, not just technical experts, and thus the input of those participants significantly affects the activities of digital forensics experts. Exercises need to consider the priorities and directives of government policy bodies [21], which may change during a dynamic exercise. For this reason, it is imperative that these exercises "test participants readiness when faced with a realistic cyber event in a stressed environment against a dynamic and skilled adversary" [22]. This means digital forensics experts must be forced to react to ongoing challenges, not just the static analysis of an image. It also means prioritizing. Policy bodies will direct what forensics experts should deem most important. In particular, for state-based organizations the protection of confidential information is often most important [21]. They must also maintain certain key services, including Domain Name Service (DNS), Windows Active Directory, web, chat channels, email, and Voice Over IP (VoIP) [23]. A unique aspect of state-based exercises is that they need to test how to respond to untrained users, because the majority of government personnel are non-technical and the organization's biggest liability [23]. In Locked Shields, there are designated White Team (scenario team) players for these roles. Additionally, States have a vested interest in appearing to be legal players. This means that digital forensics experts may need to prove the innocence or guilt of a party, which makes proper documentation and reporting even more significant [21]. Such legal constraints begin to address some of the many constraints on exercise goals.

Cyber exercises also have many constraints. The most obvious one is the time-issue. While exercises that last months do exist, they are rarely used paradigms for state-based exercises because of all of the key players involved. These players often include key policy makers and powerful military personnel that can only devote a few days to the exercise. This means that

all of the digital forensics injects need to be solvable in a limited amount of time, regardless of whether the challenge includes dynamic or static analysis. This severely limits testing in terms of acquisition. Large cyber defense exercises include multiple teams, meaning there is a need to establish tasks that present equality of challenge [18]. This will include bandwidth and tools. For most cases, this means the challenges should test experts' abilities to obtain and use open source tools.

State-based exercises apply even more constraints, particularly that the infrastructure with which participants are presented cannot be significantly altered. It needs to realistically represent what platforms and resources the state currently possesses. For most States this limits the operating systems to Windows, as Microsoft is the biggest government contractor for computer systems [21]. Additionally, because of the sensitive nature of State cyber infrastructure, the game environment must be isolated, making virtualization a must. This isolation is due to the negative consequences governments face if their activities adversely affect the private sphere, particularly in democratic States. This means that the digital forensics injects should also test experts' abilities to deploy, use, and analyze virtual technologies. State cyber organizations possess certain recruitment constraints. Hiring practices are rigid, and training is relatively fixed, as well. This means out-sourcing is usually not an option, so exercises should reflect this. In essence, the exercise needs to test the cyber teams' abilities to do more with less people [21]. A summary of the main skills to test is shown in Table 1.

Type	**Skill**
Technical	NTFS/MBR analysis
Technical	Image acquisition from hard disk
Technical	Installation/use of open source tools
Technical	Carving some deleted file
Technical	Memory dump analysis
Technical	PCAP/Netflow analysis
Technical	Windows systems logs reading
Technical	Use Windows Systems Admin tools
Technical	Windows prefetch analysis
Technical	Locate/identify malware
Technical	File hashing
Technical	Tool calibration
Procedural	Use case logging tools
Procedural	Timelining
Procedural	Description of activities
Procedural	Presentation of data in visible, simplistic terms
Procedural	Use and document data preservation methods
Other	prioritization
Other	In-time communication to team

Table 1: Goals for a Digital Forensics Challenge in a State Cyber Exercise

3. Design

In this second section, the thesis author worked with a team of three other computer technicians at the NATO Cyber Defense Centre of Excellence to determine what specific tasks would test the skills from Fig. 1 and how to evaluate teams for completing them. The author supported the team in research, technical design and implementation of the competition tasks and environment, including script coding and user history data creation, and the drafting of forensic report templates and examples.

3.1 Design Methodology

Reporting

Following the technical forensics investigation, teams must submit a forensics report that fulfills the procedural goals listed in Table 1. It needs to be both concise and detailed. The US National Forensics Computer Institute (NCFI) researched and published a guide for forensics students that is intended to describe how students should draft forensics reports for academic exercises as part of its Network Intrusion Responder Program. The Institute's methodology highlights one key fact: the report needs to be a one-go read [22]. This means that the report needs to include only the most relevant details, presented in such a way that a non-expert can read and understand the investigation in a few minutes. In order to do this, the Institute demands that students include a clear timeline of events, all individuals associated with the investigation (chain of custody, device owners and operators, resources), all the items analyzed (physical and logical) and all the programs used to conduct the analysis. It suggests that students organize the analysis in whichever way makes the most coherent story, such as time, relationship, or device [22]. In other words, the method of dividing up the report is not set in stone: it may vary depending on the scenario for which the report is written.

The Association of Chief Police Officers of England, Ireland, and Wales (ACPO) conducted similar research and published a best practice guide for forensics reporting. This guide is used by the majority of forensics training programs in these countries. It discusses the initial report specifically, advising that is should be brief and, if possible, include screen shots [23]. One of the main points the guide emphasizes is that a report should clearly separate opinion from fact.

Melia Kelley, a senior computer forensics consultant for First Advantage Litigation Consulting, conducted a report on the most effective way to organize forensics reports. Her main finding is that there needs be a template, because “templates are easy to create and will end up saving you many hours of work” [24]. In a time-pressed situation like a cyber exercise, these hours saved are priceless. For organizations-- or exercises-- that involve actors of many different backgrounds, templates enable standardized formatting and language that those responsible for reading and assessing the reports can easily comprehend [24]. She suggests that the templates include a summary, objective, evidence analyzed, steps performed, and findings sections. If they are of a sizable length, they should also include a title page and table of contents [24].

In the Locked Shields 2015 exercise, teams were not given a typical forensics report template or requirement. Instead, the digital forensics team asked them to submit a preliminary and final report. The preliminary report consisted of a series of questions concerning the investigation, such as Internet Protocol (IP) addresses and file locations, and given a very

simple template consisting of a two-column table in which to input their answers [25]. The final report asked teams to answer the who, what, when, and how of the investigation using a two-column template of time and description columns (essentially a timeline chart) [9]. The result of using these templates was that multiple competing teams provided terse answers to the challenge with little description [21]. Thus, the 2016 digital forensics team decided that a more robust report was needed for future exercises, such as those discussed in the methodologies above.

Forensics challenges for a state-based government exercises involve actors from various backgrounds, are pressed for time, and are nested within a larger scenario. As such, it makes sense to adopt the template principle for the teams to fill out, with specific guidance on what the teams need to provide. Instead of using two charts, the information obtained in both can be consolidated into a more professional forensics template. This template should be heading-based so that it is uniformly organized. Such a template pushes analysts to provide a coherent story. In this venture, the timeline is also one of the most important aspects of the report, as noted by all the methodologies listed above. Because a government forensics scenario will contain only a few artifacts that all teams can access, the report template does not require a title page and table of contents, particularly because this report should be fashioned more after the initial report than a polished report going to trial.

While the report should be concise, there are a few details teams absolutely need to include in the report in the section for findings. The United States Department of Justice (USJ) published a report to guide law enforcement in forensics practices and reporting. The report asserts that the details that should be included in regard to reproducibility of the findings are specific searches performed, such as string searches, details related to ownership, and snapshots. In terms of verifying those findings, it is important that teams submit hash values of all the items on which analyses were performed and specific versions for utilized hardware and software [26].

Environment

The United States military conducted a research project into the development of the NSA/CSS CDX previously discussed, which was purposed towards training and testing the students of the country's military academies when developing the environment for the exercise. COL David Ragsdale (Ret.) was largely responsible for this study. One of the most important aspects of the environment for a successful exercise, he found, was the need for it to be isolated during gameplay [21]. The reason for this is that if the exercise network touches real world networks, players need to be extra cautious about their actions. Additionally, if anything goes wrong during the exercise and negative effects are exacted on third parties, it will bring bad press to the exercise and harm future exercise attempts [21]. Still, players need access to the internet in order to set up their machines for competition. This is why a Day 0 is necessary. Teams need a day of access to the internet in order to set up their networks and obtain any resources they will need to successfully accomplish the exercise [21]. This is extremely important in terms of the kind of resources the teams should be utilizing: open source. Requiring teams to use open source resources levels the playing field, due to the disparity on nation's defense budgets and access to specific software [21].

In terms of the network environment itself, it is realistic and very beneficial to include systems of various operating system types. In this way, the exercise can test the players' abilities on multiple platforms [32]. Multiple operating systems also enables exercise set-up,

as unpatched versions are easy ways to introduce vulnerabilities into the teams' networks for the purposes of the exercise [21]. These patches can serve as the low hanging fruit of the exercise. In essence, the exercise needs to have multiple levels of challenge in order to keep all teams motivated to continue [21]. Unpatched operating systems are easy to fix vulnerabilities for teams with less technical proficiency.

For a government cyber defense exercise, and particularly for digital forensics exercises, the scenario that supports the exercise is extremely important. Vital skills to test include prioritization, communication, and procedures, all of which relate to the *story* of the exercise. The National Institute of Standards and Technology composed a report for introducing forensics into incident response. It explained the most important questions that an effective scenario needs to answer: [26]

1. How does the scenario dictate the sources of data?
2. How does the scenario dictate the most likely available resources and tools?
3. How does the scenario create, maintain, and require communication channels?
4. How does the scenario restrict and manage incident and response times?
5. How does the scenario shape the physical and logical environment?

The 2015 Locked Shields exercise included many elements of these methodologies. In this exercise, the teams had a preparatory day in order to download any open source resources they would need to their virtual network environment; however, the teams could not use this day for any parts of the forensics challenge, including digital forensics acquisition [27]. At the end of Day 0 the teams were disconnected from the game network(gamenet) such that they could make no further adjustments until the start of Day 1. The network itself included Windows 7 and 8 and multiple Linux operating systems dispersed among various subnets [28]. The teams were all given network diagrams at the start of the exercise. This is in line with the methodology developed for NSA/CSS CDX [21]. Fig. 1 shows the forensics related subnets for this thesis' test case, Locked Shields 2016.

There is also the question of how much knowledge the Red Team will have of the Blue Team and the gamenet environment. Exercises in which the attackers start from ground zero and have to conduct the full reconnaissance phase are classified as “black box exercises” [29]. For a large scale state-based exercise, this is not very feasible. Red Teams have a short span of time to compromise and attack multiple teams, which means the reconnaissance phase would be too time consuming [30]. If the exercise is limited in number of Blue Teams, then it would be possible to bring in this more realistic aspect.

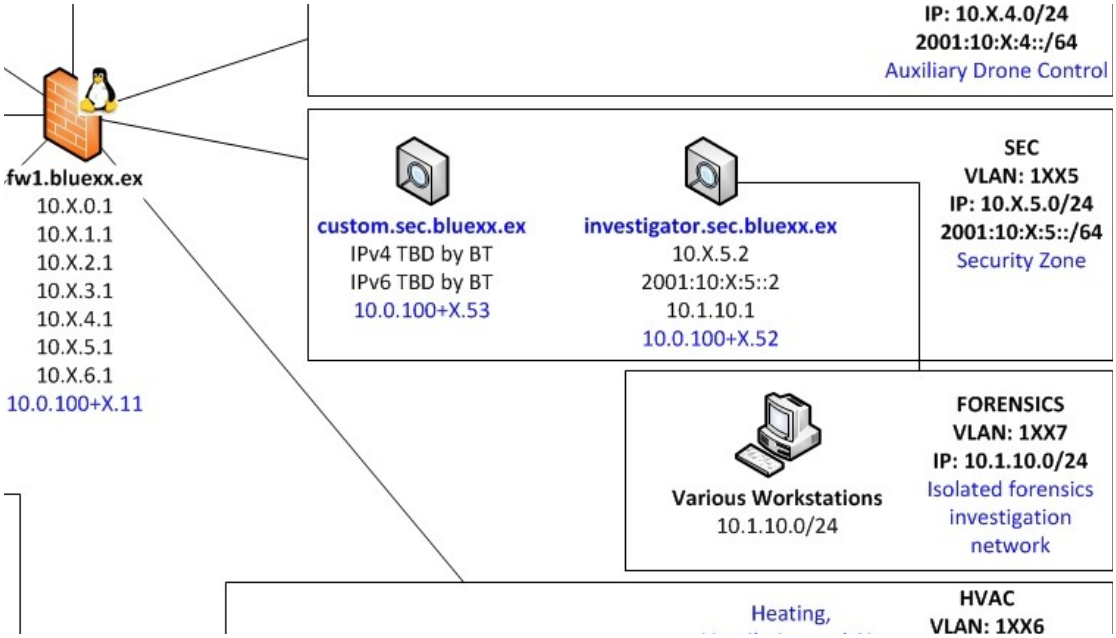


Figure 1: Locked Shields 2016 Forensics Subnets [28]

Technical Components

A digital forensics challenge for a state-based cyber exercise needs to have several components in order to achieve all the goals outlined in Table 1. The United States Department of Justice (DoJ) report suggests a number of possible challenge scenarios for digital forensics exercises. These include Denial of Service (DoS) attacks, rogue wireless access point attacks, mistaken identity attacks, uploading unwanted images, phishing scams, and encryption attacks [26]. For these to be implemented, the digital forensics team needs to create a number of components. A home server would be necessary for a DoS attack, as well as some method of creating botnets. A rogue wireless access point requires an internet access point. To upload an unwanted image they need the image itself, as well as the website or server to which it is being uploaded. Phishing scams require the creation of an email account-- and email support, in general. Encryption attacks require the files to be encrypted and an encryption algorithm [26].

Then there are the items on which the participants need to perform analysis. These can be derived from the tools that the exercise wants to require the participants to use. According to the goals articulated in Table 1, this should include pcap files for network analysis, access to a computer system that will allow for memory dumps and file system analysis, and web sites. In Locked Shields 2015, teams were given virtual access to the target virtual machine (VM), a pcap file of network traffic, a memory dump file, and the malware file that caused the incident [8]. This means that a more robust method would include some sort of web page, as well as more than one machine to analyze in order to require teams to deal with more than one type of file system.

Scoring

The NSA/CSS CDX methodology suggests using a well-defined, uniform method of scoring, one that scales well to teams of different sizes and skill levels, for state-based exercises [31]. This method should include both automated and manual scoring for different parts of the exercise. Automated scoring should be implemented for maintenance of services, which periodically checks the teams' vital services, such as Simple Mail Transfer Protocol (SMTP) and web servers. Penalties for down services should be cumulative over time, such that teams lose more points the longer the services are down [31]. Teams should also be penalized for breaking predetermined rules of conduct, such as those that mimic legal limitations [31].

The NCCDC is intended to provide curricula with a "competitive environment to assess their student's depth of understanding and operational competency in managing the challenges inherent in protecting a corporate network infrastructure and business information systems" [31]. In developing its scoring method, the competition divided scoring into three main categories: critical services, injects, and written reports [32]. As suggested in the NCCDC methodology, an effective exercise should use automated scoring for maintenance of critical services. More important to this thesis is the other two categories of scoring. Injects, the methodology dictates, should have time limits, where teams are scored at certain points for having achieved a certain part of the inject objective [32]. This methodology is helpful for three main reasons. First, it enables the scoring body to handle the scoring burden. Second, it allows teams to assess their current standing in real time. Third, it motivates teams who may only be technically capable of completing parts of the inject.

The third category of scoring is for written reporting. Teams should submit written reports for their responses to all injects, specifically revolving around the forensics portions. These reports should be scored according their inclusion of the following components: amount of evidence uncovered, discussion of impacts, detail of procedure to locate/obtain the evidence, and supporting proof (e.g. logs, screen shots, IP addresses) [32]. Teams should receive additional points for using more practical and/or safer methods [21].

Locked Shields 2015 incorporated the timeline suggestion by using two phases for the digital forensics scoring, a preliminary and a final report. Teams were scored according the their answers to a series of questions in each report [9]. This was in addition to the exercise's larger automated functionality scoring. Teams also received time bonuses for the rate at which they could complete the challenge [9]. An issue uncovered with this scoring method was that the phases were not based on ability to perform parts of the exercise, and many teams struggled with the initial acquisition phase [8]. Thus, it would be beneficial this year to combine a phased scoring method with the tasks the exercise requires the teams to perform. Additionally, the two-report method stressed the small forensics team in terms of time to grade the reports, and the 2016 exercise is larger.

Anti-Forensics

An additional layer to consider in developing the exercise methodology is the role of anti-forensics, the process of frustrating the forensics investigation and the tools used to conduct it. Anti-forensics techniques have three main focuses. The first is data hiding, in which methods are used to cover data related to the incident. This concerns things such as covering script with images, hiding artifacts in file system slack space, and using confusing metadata such as long file names [33]. The second main focus area is artifact wiping. In essence, this is

the step beyond artifact deletion, in which the artifact data is destroyed beyond the point of simple recovery methods [33]. Finally, there is trail obfuscation, used to cover the tracks and frustrate attribution, such as the use of anti-forensics techniques to confuse email forensics [33].

In a state-based cyber exercise, is it worthwhile to employ such techniques? Certainly, it would enhance the challenge, but would it be too much of a challenge given the exercise constraints and goals? Gary Kessler, Director of the Champlain College Center for Digital Investigation proposes that it is essential, because the importance of anti-forensics techniques increases as the allotted analysis time decreases, as the role of anti-forensics techniques is not to cover something forever-- it is to slow it down enough for the incident to achieve its goal [33]. Thus, a few day exercise seems to be the perfect testing ground for such techniques. Additionally, it adds an additional layer of challenge for the teams with more advanced digital forensics experts.

Selected Methodology

The methodology the Locked Shields 2016 digital forensics challenge designers selected was an isolated virtual network that is connected to the outside internet for Day 0 for the purposes of performing the acquisition phase of the exercise. Included in the challenge network will be multiple operating systems: Windows versions 7, 8, and 10 and Linux Ubuntu. The designers decided to require competitors to acquire an image of the infected machine and submit proof of the image's integrity. The designers chose to provide the teams with a pcap file, as well, in order to test their abilities to perform network, in addition to system, forensics. They also chose to incorporate a website and image for analysis.

The inject will be handled in phases. First is proper acquisition. Teams will be able to start this process on Day 0 and given technical support by the digital forensics team over the communication channel for the exercise. Next is the analysis phase, for which teams can receive points for proper network, system, memory, and web forensics components. At the end of Day 1 teams will be able to request technical answers in exchange for caps on the number of points they could earn for that category. Finally is the forensics report. The designers chose to provide teams with a template (see Appendix 1) and a sample (see Appendix 2) and require them to complete one for all evidence items analyzed. In this way, even if teams do not have digital forensics expertise in certain areas, they will be able to at least complete the report and earn some points, thus motivating them to attempt the challenge. Finally, some anti-forensics will be integrated into the inject in order to provide an additional layer of challenge for more sophisticated teams.

3.2 Test Case

Overall Scenario

As previously mentioned, digital forensics injects need to exist within the overall exercise scenario for a government cyber exercise. The Locked Shields 2016 scenario involves three States: Crimsonia, Berylia, and Revalia, three rival states within a local region. Essentially the Red Team is the villain nation of Crimsonia, and the Blue Teams play the role of Berylia. Revalia, while not initially at war with either, is a rival state within the region [34]. Berylia's primary industry is drones, an industry that Crimsonia is interested in advancing. Crimsonia

will begin attacking the civilian-operated drone facilities of Berylia, and the Berylian government deploys rapid reaction teams (the Blue Teams) to the facilities [34].

The scenario asks the digital forensics designers to create a situation in which Crimsonia conducts a cyber attack on Revalia but tries to make it look like Berylia conducted the attack in an effort to get Revalia to join the conflict on Crimsonia's side [34]. Berylia needs to prove to Revalia that Crimsonia is the responsible party. Additionally, the exercise creators ask the designers to show that Crimsonia has been stealing documents related to drone use and development from the facilities [34]. Locked Shields, like most large scale state-based cyber exercises, is intended to mimic the worst case scenario where multiple attacks and exploits are executed. This is why it is important that the digital forensics challenge contain multiple layers [34].

Berylia's technical environment consists of various operating system types, including Linux, Ubuntu, Windows 8, and Windows 10. The exercise coordinators specified that the compromised machine for the forensics challenge needs to be Windows 10. The coordinators granted permission for the acquisition phase to take place on the exercise's preparation day, Day 0. The rest of the challenge will take place over Day 1 and Day 2 (the entire exercise is one preparation and two game-play days) [34].

At this point in the challenge development, the thesis author and the other designers need to answer the scenario questions from the methodology section:

How does the scenario dictate the sources of data?
There needs to be data with clues to Crimsonia present on the infected machine. Because Berylia is the team conducting the investigation, and the damaged party is Revalia, somewhere between the creation and execution of the attack, artifacts have to infect something under Crimsonia's control that the blue teams can analyze.

How does the scenario dictate the most likely available resources and tools?
All tools need to be available open source, as not all teams have licensed forensics programs. Because the infected machine will be Windows 10, any tools used to deliver or execute the attack need to exploit Windows 10 vulnerabilities. Additionally, because the network includes various systems, it is advantageous for the delivery mechanism to be capable of delivering to more than one operating system.

How does the scenario create, maintain, and require communication channels?
The scenario requires Blue Teams to present proof of attribution of an attack to the victim party, Revalia. This means that the teams need to log and keep proof of their acquisition and analysis activities. It requires them to give a persuasive report to a legal team. Because the scenario will occur within an ongoing conflict with Crimsonia, prioritization may require stop/start of analysis. There is also high risk of a loss of communication channels during the analysis process.

How does the scenario restrict and manage incident and response times?
Teams will be limited to the timeline of the exercise, with one day provided for acquisition and two for analysis and presentation of findings. The incident response time may speed up depending on escalation of attacks. Because there is a media team involved in the exercise [34], digital forensics teams will be pushed to present findings quicker in order to prevent the

entrance of Revalia into the conflict. This allows for the use of techniques that are speedier over safer, such as live analysis.

How does the scenario shape the physical and logical environment?
There are at least three separate government networks involved in the scenario-- Crimsonia, Revalia, and Berylia. This means three public address spaces. Operating systems will be heterogeneous. The teams themselves will operate from all over the world over virtual networks/ virtual machines. This means the acquisition will also be virtual. Teams need to choose a form of connecting to the virtual machines, such as Secure Shell (SSH) or Remote Desktop Protocol (RDP).

Launching the Attack

In developing the injects, step one is mapping what effects of the incident would create analysis challenges that mapped to the goals outlined in Section 1, the methodology discussed above, and the overall Locked Shields 2016 scenario. The thesis author and her team of three digital forensics designers have to create a challenge that fulfills as many of them as possible within the confines of the overall Locked Shields 2016 scenario. This means that any scenario needs to incorporate an image, a web site, a Windows 10 machine, and a malware file. The scenario itself needs to include three actors in the execution of the attack: Crimsonia, Berylia, and Revalia. The easiest way to do this is to literally turn Berylia into the executor of the attack by turning one or more of its machines into botnets, because "bots run almost exclusively on Windows" [35]. Ideally, Crimsonia will deliver malware that gives it command and control over a Berylian system and execute the attack on Revalia from the machine. Windows 10, being a relatively new system, is fairly secure to well known malware; however, it is important to recall that botnet programs originally were not intended to be malicious. Remote control programs are advantageous to administrators, and thus they exist for all versions of operating systems. Malicious attackers take advantage of these programs and use them to gain control over systems to which they are unauthorized to access [35]. Such a program could be used to gain control over a Crimsonian machine.

According to this thesis' selected methodology, the attack on Revalia needs to include a website and an image. One of the most common forms of cyber attack conducted against government entities is website defacement [36]. These attacks are often semantic attacks, directed at disseminating false information and inciting fear or anger [36]. This matched the scenario well. Thus, Crimsonia will deface the Revalian government website with an image that linked to Berylia. For this to occur, the Revalian website needs some type of a vulnerability to exploit. The five most common web server vulnerabilities are remote code injection, Structured Query Language (SQL) injection, format string vulnerabilities, cross-site scripting, and username enumeration [37]. The purpose of the vulnerability is for Crimsonia to deface the website with an image. SQL injection is enough to do this; however, if the designers want the image to do anything beyond that (i.e. execute some type of code), then the server also needs to be vulnerable to cross-site scripting.

For Locked Shields, the thesis author and fellow designers tested a multitude of Wordpress exploits and discovered one that works on Windows 10. This exploit relies on a compromised plugin called RevSlider version 4.1.1. The attacker uses the infected machine to browse to the victim web-server with the additional command /wp-admin/admin-ajax.php? action=revslider_show_image&img=../wp-config.php. This downloads the file with the

database credentials, such that the attacker obtains login information for the server and uses it to deface the website with the image [38].

Delivering the Attack

In order for the attack to occur in this manner, first Crimsonia needs to turn a Berylian computer into a Botnet. There are many ways to do this. The delivery mechanism has to be one that adds evidence in a way that helped test the skills listed in Table 1. Because network analysis is a key skill, the delivery mechanism should introduce network traffic. Additionally, the environment contains multiple hosts and operating systems, so the delivery mechanism has to be easy to multiply and deliverable to multiple operating systems, thus ruling out operating system specific deliveries. Options for this include e-mail, ftp, video chat, voice over IP, advertising, and social media sites. Because in this type of exercise the designers need to trick the victim machine user into downloading some type of malware, email is a wise choice, essentially because it is easy to disguise an executable attachment as a pdf in Microsoft Outlook.

For the specific email, it needs to be something a user would want to read and would realistically download an attachment from while at work. Phishing emails are specifically written to convince a target that they come from a trustworthy source [39]. The most common subjects of these emails include online payments, security violations, and IT department messages [39]. It is important that the email make sense in the context of the exercise scenario. Given that the workers are drone research scientists in the Locked Shields scenario, the email needs to be clever enough to trick someone whose organization is heavily concerned with security. In specifically targeted attacks like this one, attackers favor Spear Phishing as the attack vector [40]. Spear Phishing campaigns use information gathered about individuals to compose communications that appear personal and legitimate in nature [40]. That this email be intricately supported by the scenario is important, as previous exercise reviews cited competing teams' decreased motivation when the event did not seem realistic [41]. This also means there needs to be a background story on how the attacker conducted the reconnaissance for the spear phishing attack.

For Locked Shields 2016 this thesis' author and her fellow designers created an accountant persona within the organization. The accountant has a Facebook account that registers her as an employee in the drone organization, as does the victim machine user. The user and the accountant are friends. Facebook and similar social networking sites are primary resources of reconnaissance, because they list not only personal information that allow attackers to craft convincing emails, but they also show the relationships between people that can be utilized in a spear phishing attack [40]. In order to be plausible, the designers need to create some email history between the accountant and the user from their work email accounts, supporting their work relationship. For the attack, the attacker makes an email account that looks similar to the victim's friend's legitimate email. This is normally done by changing one or two letters in the domain name [42]. The email itself needs a legitimate excuse to have an attachment. For this scenario, because the designers chose an accountant, they will send him an email from his friend the accountant asking if he could view the invoice that she received from an executive within the company. The real domain name for the victim's workplace droneworld.site, and the attacker's domain will be dronevvorld.site, replacing the letter *w* with two of the letter *v*.

This leads to the question of what malware to use in order to turn the target machine into a bot. According to the head digital forensics inject developer for the NATO Cyber Centre of Excellence, the main criteria for selecting the malware is that it is open source, reliable, supports file download and upload, enables encryption, and is easily customizable [38]. Open source malware has a few advantages. First, the code is simple to modify to meet the exercise's needs, including leaving clues. Second, it is usually accompanied by explanations of its use. Finally, with open source malware there is little risk of an unknown developer backdoor that could compromise the exercise [38]. This also supports the reliability of the malware. The malware should also be proven reliable by testing it thoroughly for bugs.

Features important to the malware include file upload and download, because the attacker needs to deliver artifacts to the victim in order for it to upload the defacing image to the web server. In order for the network analysis part of the exercise to present a decent challenge, it also needs to be possible to encrypt the communication between the attacker and the bot [38]. Finally, because exercises requires the designers to leave specific clues for the teams, the malware needs to be easily customizable. Areas that the designer will want to customize include metadata, file location, ports, and process names [38]. A Remote Access Trojan (RAT) that meets this criteria is the Qaesar RAT, which was used in the 2015 Locked Shields and the designers decided to reuse in a modified manner in the 2016 exercise. Additionally, the RAT allows for encrypted file upload and download, an essential part of the scenario [38]. Once the RAT takes over the machine, it will use this encrypted channel to download files, such as the defacing image.

File Theft

Recall that a large constraint in government cyber exercises is the need to prioritize and respond to political concerns. Locked Shields incorporated this by making the discovery of leaked documents a priority. Thus, the teams will need to discover if and what documents the Crimsonians copied from the infected machine. Because the RAT uses encrypted file download, this presents a challenge. The two day exercise is not enough time for teams to decrypt the file upload, and just timestamps are not enough to identify copied documents, since in the scenario the Windows 10 user is unaware of the system's compromise for an extended period of time[38].

Thus, the process of copying the files needs to leave clues. Because a main goal of the exercise is the use of system administration tools, this can be integrated into this section of the exercise by executing a process. The thesis author crafted a python script that walks the Windows 10 user's directories for a keyword in the file names (in this scenario, “drone”) and copies those files to a new folder. When the attacker downloads the image to the victim, he also downloads this script. After its execution, he will upload the entire folder using the RAT. After file upload, the entire folder will be deleted. This entire process leaves various clues for the teams. The script is shown below:

```
import os
import shutil

global targetDir, keyword, destFolder
targetDir = "C:\\Users\\codeRunner"
destFolder = "C:\\Users\\codeRunner\\walk"
keyword = "drone"

def walkDir(targetDir, keyword, destFolder):
    counter = 0
    #print targetDir, keyword, destFolder
    for dirname, subdir, files in os.walk(targetDir):
        for fname in files:
            try:
                fname = fname.lower()
            except Exception:
                fname = fname
            try:
                if keyword in fname:
                    fullpath = os.path.join(dirname, fname)
                    newpath = os.path.join(destFolder,fname)
                    try:
                        shutil.copy(fullpath, newpath)
                    except Exception:
                        break
            except Exception:
                break

def main():
    global targetDir, keyword, destFolder
    walkDir(targetDir, keyword, destFolder)

main()
```

Anti-Forensics

Given the selected methods and clues, the designers must decide what anti-forensic techniques to employ in order to increase the challenge. Techniques available include those focused on frustrating acquisition, stenography, source elimination, fabrication of false positives, data destruction, virtualization, memory related, and forensic tool exploitation [43]. Given the time constraints of a government exercise, the methods should not be unduly time-consuming. This rules out most acquisition frustration, as this process has already proven time consuming in such exercises [38]. Instead, the anti-forensics should to be applied to the analysis portion itself.

While several tools exist for this, the forensics challenge in Locked Shields 2015 proved too difficult for many of the teams. The anti-forensics technique chosen that year was simple file deletion. The files important to the attack were put in one folder that the attacker deleted and

needs to be recovered. Teams failed to do this [38]; thus, this year the designers decided to repeat the use of a single folder with all of the attacker's files that is deleted. For a greater challenge, they will also employ metadata hiding methods. When the RAT executes it appears in the running processes, but the designers will rename the process name to a common Windows process, CCleaner, a free program to free up hard disk space by removing cookies and temporary files [38]. If teams compare the time of opening the malware file to the process execution will be obvious that the process is related to the attack.

The Inject

Now that the thesis author and her fellow designers possess the entire scenario, components, and evaluation methods, there remains the written inject for them to present to the competitors. An inject is essentially the "task and purpose" [46] of the challenge. The written inject needs to include the story of the scenario, which has to be both simple and realistic [46]. It needs to boil down the scenario to the most important factors; further clarification and information can be provided later if it becomes pertinent. The inject must articulate what the main tasks are for the team to complete and specify the objects they will be provided in order to complete them [46]. The object descriptions should include any technical information that the teams need to perform the investigation. The inject also has to include the communication methods of the findings, including the reporting format and destination. It is important to include a specific timeline for completing the specified tasks and reports [46]. The Locked Shields 2016 inject written by the thesis author is shown below:

Attention! RRT is requested to perform a digital forensics investigation to prove that the recent defacement of the Revalian government web server was not performed by Berylia. At approximately 11:15 A.M. on 04 April 2016 revalia.gov was defaced. Revalia's web server logs show that the attack came from an IP address used by the Berylian Armed Forced Drone Control Facility. They are now threatening to declare war against Berylia if it cannot prove it was not responsible! Your team is tasked with performing an investigation on machines within the subnet of the facility that is linked to the attack. This subnet includes the following machines and accounts:

OS: Windows 10 (32bit)
User: Sheldon Jobs
Username: coderunner, password: LS16Sheldon
E-Mail: jobs.sheldon@droneworld.site, password: LS16M@il
Skype:jobs.sheldon@outlook.com, password: LS16M@il
IP: 10.1.10.17

OS: Windows 7 (32bit)
User: Raj Woz
Username: webmaniandevil, password: LS16Raj
E-Mail:woz.raj@droneworld.site, password: LS16M@il
Skype: woz.raj@outlook.com, password: LS16M@il
IP: 10.1.10.26

OS: Windows 8 (32bit)

User: Howard Gates
Username: developerduck, password: LS16Howard
E-Mail: gates.howard@droneworld.site, password: LS16M@il
Skype: gates.howard@outlook.com, password: LS16M@il
IP: 10.1.10.54

OS: Ubuntu 15.10
User: Penn Dell
Username: adminfudd, password: LS16Penny
E-Mail: dell.penny@droneworld.site, password: LS16M@il
Skype: dell.penny@outlook.com, password: LS16M@il
IP: 10.1.10.22

In order to conduct your investigation, you will have two sources of evidence: the virtual machines cited above and randomly recorded network traffic from that subnet. Additionally, it is known that the Crimsonian and Revalian government public IP address ranges are:

Crimsonia: 90.0.0.1-99.255.255.255
Revalia: 120.0.0.1-129.255.255.255

In addition to the analysis, you are tasked with writing a forensics report for each evidence item related to the attack (template and example attached) no later than 211000Z (13:00 GMT+3) to white@mail.ex with the email subject line: Scenario Inject # Forensic Blue (Team Number) Report. The report may be attached in text or pdf form. Remember, the results of your investigation are vital to preventing Revalia from joining forces with Crimsonia!

Hints

In 2015 the range of expertise amongst the teams was very widespread. For this reason the 2016 designers chose to provide teams with hint sheets upon request for pcap, hard disk, and memory analysis. These hint sheets could not be full guides. This is because, as part of the NATO program, a full workshop on the challenge and digital forensics in general will be conducted a month after the exercise. Instead, the thesis author chose to create sheets of five hints that pointed the teams towards where/what to search in their files (pcap, dump, image). These five hints would be derived from the findings of the designers' test investigation (see Appendix 3 for all three sheets). In determining the conditions for the hints, a penalty to a team's existing score would likely deter many of them from requesting hints, even when they needed it. Instead, it might be wiser to put a cap on the number of points they could earn in the category in which they received help, such that the teams can only earn 50% of the category's points.

Team Investigator Machines

All teams will be requested to prepare and submit a virtual investigator machine. This machine can run any operating system that the teams wish. It needs to contain all of the forensic

tools that they may need for the challenge (the specifics of which will not be disclosed prior to the exercise commencement). Teams will be told that they cannot connect any USB or other external devices in order to conduct the investigation, which prevents them for using some software that requires such devices to function. All of the resources they require will also need to be added to a share folder that the teams can access from the victim machine, such that no additional software will to be downloaded to the victim machine, which would taint the evidence [9].

4. Results

In this section the thesis author helped map the artefacts of the example investigation to point values for use in the test case exercise. She took part in answering team questions during the exercise and scoring the final reports. During and after the exercise she logged first hand observations and conducted interviews with both participants and exercise staff for use in the evaluation of the results.

4.1 Expected Results

Investigation Process

At this point the design team has implemented the design described in Section 3. In order to determine what the teams should find and the appropriate level of difficulty, first the digital forensics challenge thesis author and fellow designers needed to conduct their own investigation to establish a baseline of expected findings. This investigation also helped the designers in drafting the scoring guidelines and determining the difficulty of the tasks they designed in the challenge. This investigation had to include the pcap file, hard disk analysis, and memory dump analysis.

For the pcap, they began by filtering the traffic by port. From here it was easy to spot an abnormal port, 88. Following the Transmission Control Protocol (TCP) stream of these ports shows two sessions between the victim and the attacker. The traffic is encrypted; however, the second session is initiated by the victim machine, explaining how RDP is possible despite the firewall. The malware makes the connection from the victim, and the Windows firewall allows for outgoing remote connections [38]. Since the main attack the team is tasked with investigating, it is important to inspect HyperText Transfer Protocol (HTTP) traffic over port 80 to the web server. The inspection showed 50 sessions between the victim machine and the web server. This traffic is unencrypted and quite revealing. It shows the failed and successful logins to the server, including the credentials the attacker used. These streams give a very detailed timeline of the attack and even showed the specific requests made to the server, one of which reveals the vulnerability that the attacker exploited on the Revslider plugin.

Following the attack, another stream shows the upload of a zip folder that contains an image. If this image's properties are analyzed (the designers used EXIF editor), then the image properties show that the attacker embedded some type of executable code into the image, a key indicator of its malicious intent. Unfortunately, the Hypertext Preprocessor (PHP) script added to the EXIF data of the defacing image was not properly included. Rather than adding the code to the EXIF data itself, the designers needed to have saved it as a separate file that the EXIF data calls. The former method results in a NO CODE EXECUTION ALLOWED error. At this point, to fix the error would require redeploying the entire attack with a new image and PHP file, and because the clues contained in the PHP script are duplicated in other areas of the exercise (attacker country email and language), the designers decided not to redeploy. If the teams think to check the EXIF data they will still see the attempted code execution.

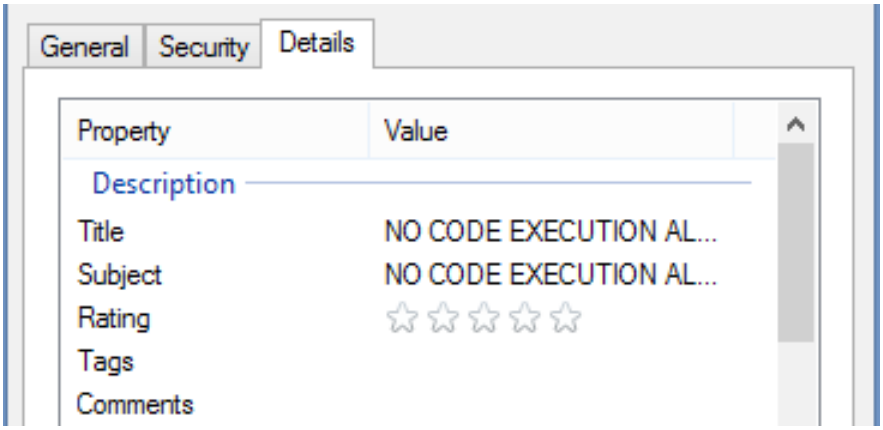


Figure 2: EXIF Data of the Defacing Image

The memory dump analysis is key to overcoming the main anti-forensics technique, metadata manipulation. Using Volatility Framework 2.5, the ccleaner.exe is shown running with a suspicious mutant, CCle@ner2016. What visibly ties this process to the attack is that the process made a connection over port 88, the port connected to the attacker machine. In Startup Manager, there is a registry persistence associated with Shelby Cole (the persona used to conduct the spear phishing attack). Note: Fig. 3 shows the results on the test machine, Windows 64-bit, and in the exercise the victim machine is Windows 32-bit.

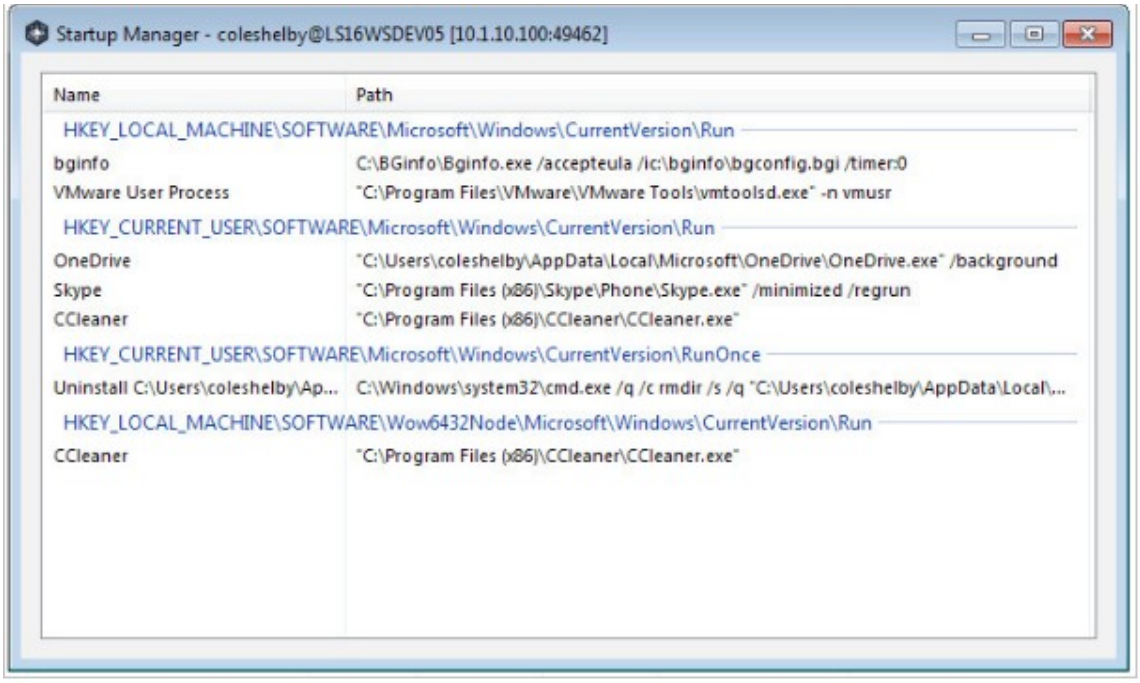


Figure 3: Systems Manager on Test Machine

A strings analysis of the process file for CCleaner gives a big clue to the teams: the phrase “Crimsoni@2016” (note: this is also the encryption password for the connection). It also

shows the transfer of a number of documents by extensions, such as Portable Document Format (PDF).

Hard disk analysis reveals the most information. For the test investigation the designers used the following open source tools: SIFT Workstation, Plaso 1.4.0, vshadowmount. SIFT exports outlook emails and attachments [46]. In this file is the spoofed email address of cole.shelby@dronevvorld.site, different than the droneworld.site email domain of the victim. The file also shows that there is an email attached within that email which contained a suspicious object...which is the ccleaner.exe. Uploading this file to virustotal.com reveals its malicious content. An analysis of the emails sent around the time of the attack shows that the victim sent one out stating that he would be away from work for a doctor's appointment, indicating that the activity during the time of the attack was not his own. Next, the designers used the Debian package Plaso to view a superuser timeline [47]. This timeline analysis incorporates a number of skills from Table 1.

A Microsoft warning reveals that when the user tried to open the document he received in his email he received a warning that the attachment may contain viruses. There is a registry entry with the temporary location for ccleaner.exe. From the prefetch file its execution created it is possible to obtain the exact time of execution. The attacker created a number of files in the CCleaner folder and also created hidden folders, buried in the Skype application folder, that was used to house his files and also used to gather files using his python script. It is possible to see, too, that it executed on the system from that hidden folder's location. Opening this script file shows what the attacker was looking for (files related to “drone”) and gives clues to the attacker (comments in Crimsonia's native language). Prefetch file analysis also reveals that the attacker used a program called Mimikatz to look for credentials, and shortly after a file called pass.txt was created in the hidden folder. It also shows the opening of files in notepad during the time of the web server attack. The RDP protocol enable time is shown in the system log file, just before the time of the attack. The Windows event log (WinEVTX) gives further details about this connection. The Plaso tool also displays web browsing history, which shows all of the browser commands executed visited during the attack, including logins, file upload, and file deletion posts. See Table 2 for the hard disk analysis timeline.

Score Sheet

This exercise is, in the end, a competition. Thus, the digital forensics challenge has been allocated a set number of points by the exercise director. Recall that a multinational event has to take special care that the scoring method is well-defined and fair [21]. Thus, the thesis author and her fellow designers created a table of all the artefacts and pieces of evidence (including email addresses, IP addresses, running processes, hidden folders, ect) and associated a set number of points for each. For the reports themselves, as an evaluation of writing is inherently subjective, they chose to use the same scoring method as Locked Shields 2015, in which the teams were divided into four tiers of ranking for how well written and organized their reports were and allocated points based on the tier. This was well received as fair during Locked Shields 2015 [38]. Because similar methods may be used in future years, the exact scoring sheet is not included in the report.

Date	Info
30.03.2016 06:35:37 UTC (tool: pffexport)	User received a message from suspicious e-mail address: cole.shelby@dronevvorld.site Delivery time: Mar 30, 2016, 05:15:52 UTC Creation time: Mar 30, 2016, 06:35:37 UTC E-mail contained Invoice from InterDrone.msg which contained suspicious ole object: 1_oledata.mso (ccleaner.exe) – MD5sum: 341e2e4bab-cbe675b6b4eddacfa13dec File uploaded to virustotal.com – 2/54 (ole) 46/54 (exe)
30.03.2016 06:36:36 UTC (tool: Plaso 1.4.0)	Microsoft Office shows warning – Some objects contain viruses… indicate that user tried to open suspicious at-tachment.
30.03.2016 06:36:37 UTC (tool: Plaso 1.4.0)	Registry entry with temporary location of ccleaner.exe (c:\Users\ls16wsdev01\AppData\Local\Temp\ccleaner-.exe) /Windows/System32/config/SYSTEM hive
30.03.2016 06:37:34 UTC (tool: Plaso 1.4.0)	ccleaner.exe appears on the system (c:\ProgramFiles\CCleaner\CCleaner.exe)
30.03.2016 06:37:51 UTC (tool: Plaso 1.4.0)	Ccleaner.exe executed (-10sec), prefech file created /Windows/Prefetch/CCLEANER.EXE-D4D76A60.pf
30.03.2016 07:05:54 UTC (tool: Plaso 1.4.0)	Attacker populates /ProgramFiles/CCleaner/folder with legit files: uninst.exe/ccleaner64.exe/Lang(folder con-tains multiple dlls)
30.03.2016 07:05:54 UTC (tool: Plaso 1.4.0)	c:\Users\ls16wsdev01\AppData\Roaming\Skype\DataRw Folders: ldp, lds, ldr (hidden)
30.03.2016 07:11:47 UTC (tool: Plaso 1.4.0)	Python script appears on the system: lds.exe (location: Skype\DataRw) File contains crimsonian language in comments. Script is design to search any document with “drone” in its name and save it to Skype\DataRw\lds\ folder.
30.03.2016 07:13:38 UTC (tool: Plaso 1.4.0)	Python script executed (-10sec), Prefetch file: PY-THON.exe
30.03.2016 07:13:57 UTC (tool: Plaso 1.4.0)	Python script populates Skype\DataRw\lds folder with drone documents
30.03.2016 07:20:46 UTC (tool: Plaso 1.4.0)	Mimikatz.exe appears on the system: Skype\DataRw\ldr\
30.03.2016 07:20:56 UTC (tool: Plaso 1.4.0)	Mimikatz.exe executed. Prefetch file: MIMIKATZ.exe
30.03.2016 07:20:57 UTC (tool: Plaso 1.4.0)	Pass.txt created. File contain user (jobs.shel-don@outlook.com) password sha1 hash. Location: \Skype\DataRw\ldr\pass.txt
30.03.2016 11:41:04 UTC (tool: Plaso 1.4.0)	RDP protocol enabled on the system. /Windows/System32/config/SYSTEM hive
04.04.2016 07:25:46 UTC (tool: pffexport)	Users sends information about doctor’s appointment at 0800 EET. Logs off around 0800. There is also browsing history – google maps – hospital in Tallinn… also pdf was created with map how to get to hospital.
04.04.2016 08:10:01 UTC	Berylia.zip appears on the system: Skype\DataRw\ldp

(tool: Plaso 1.4.0)	File contain image from defacement. MD5sum header.jpg: 36d8e0952f949fc2d35471f-daacb3481 MD5sum berylia.zip:61926c34166b651a72872bc0edef-ddf7
04.04.2016 08:15:14 UTC (tool: Plaso 1.4.0)	WinEVTX shows events related to RDP connection to the maichne. Connection from 10.1.10.17 indicates possible malware involvement (reverse proxy).
04.04.2016 08:15:32 UTC (tool: Plaso 1.4.0)	WinEVTX shows events related to RDP connection to the maichne. Successful connect – resolution 1024/768, user: jobs.sheldon@outlook.com, fits timeframe of the second session (92.223.16.65) from pcap.
04.04.2016 08:16:02 UTC (tool: Plaso 1.4.0)	User accessing website revalia.gov – same event in pcap
04.04.2016 08:16:27 UTC (tool: Plaso 1.4.0)	User accessing – revaila.gov/wp-login.php – possible log in attempt
04.04.2016 08:17:25 UTC (tool: Plaso 1.4.0)	User running url – revalia.gov/wp-admin/admin-ajax.php?action=revslider_show_image&img=../wp-config.php as a result admin-ajax.php is being downloaded.
04.04.2016 08:17:26 UTC (tool: Plaso 1.4.0)	admin-ajax.php appears on the system: c:\Users\ls16ws-dev01\Downloads\admin-ajax.php contains configuration of the wordpress web server with plain text credentials to database: wpadmin/redhat2012
04.04.2016 08:18:33 UTC (tool: Plaso 1.4.0)	User accessing – revaila.gov/phpmyadmin – possible log in attempt
04.04.2016 08:18:34 UTC (tool: Plaso 1.4.0)	NOTEPAD.exe prefetch file indicates user opened admin-ajax.php file in order to get the sql server credentials
04.04.2016 08:19:57 UTC (tool: Plaso 1.4.0)	User accessing – revaila.gov/wp-login.php – possible log in attempt, pcap indicates successful log in attempt
04.04.2016 08:20:45 UTC (tool: Plaso 1.4.0)	User accessing berylia.zip – Skype\DataRw\ldp\berylia.zip
04.04.2016 08:20:47 UTC (tool: Plaso 1.4.0)	User uploading berylia.zip - URL: Visited: codeRunner@http://revalia.gov/wp-admin/update.php?action=upload-theme
04.04.2016 08:20:51 UTC (tool: Plaso 1.4.0)	User activated new theme: URL: http://revalia.gov/wp-admin/themes.php?activated=true
04.04.2016 08:21:48 UTC (tool: Plaso 1.4.0)	User deletes content from revalia.gov: URL: Visited: codeRunner@http://revalia.gov/wp-admin/post.php?post=2&action=trash
04.04.2016 08:21:51 UTC (tool: Plaso 1.4.0)	Visited: codeRunner@http://revalia.gov/wp-admin/post.php?post=5&action=trash
04.04.2016 08:22:11 UTC (tool: Plaso 1.4.0)	User logged out from revalia.gov: URL: http://revalia.gov/wp-login.php?loggedout=true
04.04.2016 08:22:18 UTC (tool: Plaso 1.4.0)	User checks if the attack was successful by accessing revalia.gov
04.04.2016 08:23:42 UTC (tool: Plaso 1.4.0)	WinEVTX shows events related to RDP connection to the machine. Possible log off.

Table 2: Hard Disk Analysis Timeline

4.2 Evaluation Method

Nine key steps are involved in evaluating the effectiveness of an exercise: [48]

1. Appoint evaluator head
2. Organize resources
3. Formulate questions
4. Prepare evaluators
5. Observe the exercise
6. Analyze the data
7. Disseminate the report
8. Process lessons learned
9. Incorporate into the next exercise

The author of this thesis is in charge of developing evaluation materials for the exercise. In terms of resources, the best way to gather feedback is to use the same platform used for all of the exercise's communications, the online web portal, because the teams would already be providing other documents and reports over this platform [38]. The formal questions need to be linked specifically to the exercise objectives [48]. For this reason, the author turned the objectives in Fig.1 into the feedback questions. One issue with intense cyber exercises is that by the end of the exercise teams tend to be exhausted and unlikely to give thorough written feedback. In Locked Shields 2015, during the exercise question and comment session immediately following the exercise, few teams took the opportunity to comment. Thus, a rating system instead of open ended questions is advisable (See the Summary section for the questionnaire). For the test case, this sheet will be added onto the web portal for delivery to the evaluators on the final day of the exercise. This type of evaluation provides experiential feedback, which is a valuable way to test what the creators' expectations of the exercise were against the experiences of the participants [48]. Primary materials also provide a good basis for evaluation, and from the exercise this will come mostly from logs of the communication channels. There is a communication chat room specifically for the Blue Teams to discuss forensics and also ask questions to the forensics team. These logs will serve as records of what the teams found most challenging.

It is also important to involve in the evaluation the participants not directly tied to the forensics exercise. Recall that State-based exercises first must serve their political masters, which makes it all more important that the exercises strategic leaders feel that the forensics challenge fulfilled their goals. Essentially, the evaluation needs to “involve the management” [48]. For this, personal interviews during and after the exercise will be used. Evaluators must also consider where they should place themselves during the exercise itself. The Locked Shields digital forensics designers will be physically with the infrastructure (Green) and scenario (Yellow, and White) teams. In this way it will be possible to both manage and track the technical moves of the Blue Teams and also see how the rest of the exercise is playing out as a whole and the forensics challenge's role inside of it. It is not considered good practice for the evaluators to be too close to the challenge's participants, because “this can disturb them” [48]. Even so, most of the participating Blue Teams for such a large State exercise are not physically located in the same place as the rest of the teams, connecting remotely to the exercise network.

4.3 Observations

Here begins the reflection and analysis part of the thesis. At this point, the challenge has been fully designed and implemented.

Day 0

Originally, the start of the exercise was scheduled for 09:00 on the first day, at which time the teams should have download the pcap file that was uploaded to the web portal and started determining what additional acquisitions (memory and hard disk) they should make. The overall exercise included a half hour kick-off web session, at the end of which the gamenet would open. The web session started and ran late, delaying the teams' access to the network by over an hour; however, the teams were able to use this time to download and begin pcap analysis. At the end of the web session, teams needed to send a communication check into the White team, and only once all teams had successfully communicated could the White team could release the inject, for purposes of fairness. Multiple teams contributed to a delay of another hour.

Once the White Team sent the competing teams the inject, the thesis author intended for them to use the pcap in order to decide what machine(s) in the victim subnet they should investigate. This would enable the teams to perform all the acquisitions and begin analysis on this first day. During this time period a couple of teams experienced issues with their investigator machines, mostly issues concerning their familiarity with the machine environment that had picked. One hour before the gamenet would be shut down for the day, teams were asked to send an update of their acquisition status. To the all of designers' surprise, multiple teams reported having a certain number of memory and/or hard disks acquired out of four. Essentially, many teams did not use the pcap in order to narrow down the acquisition requirements as the designers intended them to do. At the end of the day, about one quarter of the teams had completed acquisition, one half of the teams in the progress of making acquisition, and one quarter of the teams struggling to understand how to make the acquisitions.

Day 1

On the first day of the full exercise, when the rest of the exercise's injects began, there again was a delay in opening the gamenet, but this time only by thirty minutes. Many teams complained about not having enough time to perform the acquisitions, an issue linked to the fact that they were attempting to acquire hard disk images and memory dumps from all four machines in the victim subnet. Still, a few teams pulled far ahead, with one completing the challenge by the end of the morning, heading into a bonus IPv6 network traffic inject in the afternoon, unrelated to the main digital forensics challenge. This bonus challenge came from a developer outside of the digital forensics challenge design team and focused on data ex-filtration by splitting packets between IPv4 and IPv6 channels.

One team even uncovered a flaw in the inject design: the designers had used an outside mail server for all of the machines involved in the scenario (attacker, victims, and third parties), and the team found this address in the header information for all of the packets. Although this did not affect the challenge much, as all of the packets had this header uniformly, it is an oversight to be corrected for next year's challenge.

In the afternoon teams were reminded that they could take advantage of the hint sheets if needed. One team asked how to receive these sheets but did not follow up with a request. Meanwhile, the media team from the exercise placed pressure on the forensics analysts by running numerous stories about the supposed Berylian attack on Revalia and the teams' lack of evidence for disavowing Berylian involvement. In the late afternoon, the designers opened each team's investigator machines to look at the teams' progress. A few teams were actively analyzing memory and hard disks, many were still performing acquisition, and some teams were logged out of the machine, indicating either commencement or foregoing of the challenge. One hour before the close of gamenet for Day 1, teams were sent out another reminder of the hint sheets. One team replied right away, asking for the pcap hint sheet. Another requested clarification for how it affected the scoring and chose not to take a hint.

At the end of Day 1, status reports were again gathered from the teams. Four teams stated they were finished analysis and in the progress of writing their reports. Multiple teams responded that they had completed acquisition and were in the analysis phase. A few teams were still struggling to make acquisitions. The thesis author advised them over the chat channel to shift their main focus to the pcap analysis.

Day 2

At the open of gamenet on Day 2, another team requested a hint sheet, this time for memory dump analysis. Another team requested this same memory hint sheet two hours later. The first reports arrived approximately two hours before the deadline. It became very apparent early on in the scoring process the wide spread of results. A few teams scored nearly 75% of the total possible points, while a few teams barely managed to scrape any points at all. The reports themselves were varying in both quality and composition. A couple teams submitted novel-length reports of appendices and a couple with single page findings. For scoring, the designers had created a point category for report quality, in which teams were divided into one of four categories based on the overall organization and composition of their reports. The winning team's report was not the longest, but it clearly over-viewed the findings, contained all the important details, and had only the pertinent supporting appendices.

After submitting the scores, the thesis author received a desperate message from one of the teams asking why they had not received any points for memory or hard disk analysis. Upon investigation, she and her fellow designers discovered that they had only received a report for the pcap analysis. The White team had missed the report in its email in-box. The report was quickly printed and graded in order to adjust the team's score. This is a clear indicator that there needs to be an acknowledgement of report receipt set up next year. After this adjustment, the scores for the team were considered final.

4.4 Scores

The final scores are summarized in Fig.6, broken down by category of scoring. The maximum score for the challenge including the bonus network analysis inject was 10,600 points. The highest scoring team earned 7,750 points, and the lowest scoring team earned 50 points. When the scores are separated among the categories of scoring, it appears that the highest scoring teams excelled in hard disk analysis. These teams also had the highest scores for well written and organized reports.

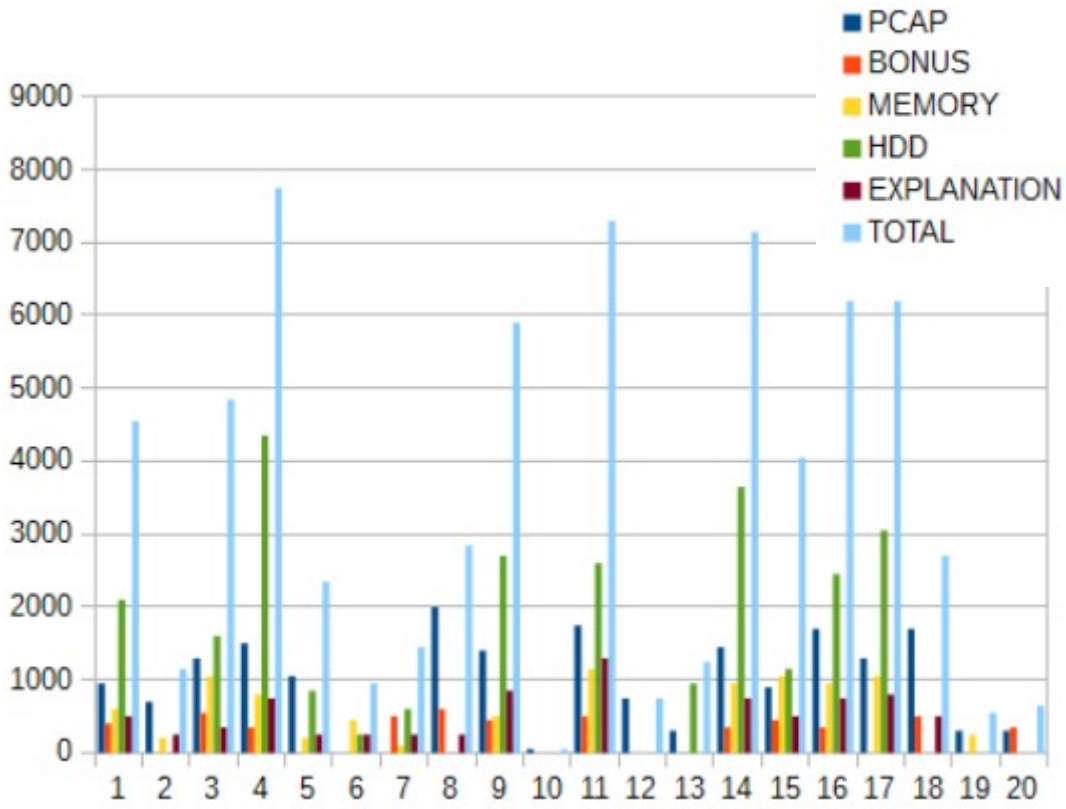


Figure 6: Digital Forensics Challenge Scores

Many teams struggled with acquisition and instead focused their efforts on pcap analysis. Thus, earning points outside of the pcap analysis gave the winners a leg up on the competition. Additionally, many of the teams focused more on the attack on the web server than on the attack that allowed Crimsonia to use Berylia to conduct the attack, missing the point of the investigation: not to show what happened but how it happened. The teams with the best reports earned a lot of points for clearly showing the whole story, from the spear phishing to the use of the RAT to the Wordpress vulnerability used to attack the website. A large portion of the teams failed to find the second attack they were asked to investigate: the suspected file theft. The winning teams discovered the python script that copied these files, as well as the clues inside the code that pointed towards its Crimsonian authors.

4.5 Analysis of Results

The big question is, how well did the inject tested the skills the thesis author designed it to test and how effective was the of the scenario. For this analysis, the thesis author combined her first hand observations with the exercise chat logs, written participant feedback, and personal interviews with multiple management level participants from the exercise.

Category 1: Technical Skills

1. Ability to make an acquisition from hard disk
Teams needed to acquire a hard disk image of the Windows 10 machine in order to find a large number of the artifacts. All the teams except one managed to at least make one hard

disk acquisition, and the one that did not had other technical difficulties that kept it from performing the acquisition.

2. Ability to make memory acquisitions
Teams also needed to acquire memory dumps in order to find artifacts related to the processes of the malware sample. Nearly half of the teams did not make a single memory dump acquisition. This is an indicator that the scenario needs to place more emphasis on the need to acquire such a dump.

3. Ability to choose and apply appropriate forensics tools
For their investigator machines, teams created their own virtual investigator machines and submitted them to the digital forensics team. Most of the teams successfully chose and installed analysis software for these machines, with the exception of the losing team that unwisely chose a new platform with which both they and the challenge designers were unfamiliar. Newcomers to the exercise were less prepared with analysis software, but this is something that they will improve only with experience [49].

4. Ability to read NTFS
The machine hard disk that the teams needed to investigate was the Windows 10 machine. Even if the teams chose to analyze the wrong machine, three of the four hosts on the victim subnet had NTFS file systems.

5. Ability to find and recover deleted files
All of the attacker's tools and documents were placed in a folder that the attacker deleted at the end of the attack. Teams could recover all of these files through the proper reading and carving of the Windows 10 file system. The teams with the highest points were those with the best analysis of the Windows 10 hard disk image, a good indication that this skill was properly tested with the inject.

6. Ability to analyze pcap files
The very first thing the challenge designers provided the teams was the pcap file. About 25% of the total exercise points came from pcap analysis. As stated before, the best teams realized that not only did they need to analyze the pcap file for artifacts but also for clues as to what hosts in the victim subnet they needed to analyze. Only one team failed to obtain any points from the pcap file.

7. Ability to read Windows system logs
One scored area that only a few teams managed to obtain points in was Windows alerts. A reading of the Windows event logs showed that when the victim opened the malware attachment, his computer generated a warning. Considering that only a couple teams located this, and the rest of the artifacts could be found through other analysis means, more artifacts should be placed in the logs.

8. Ability to analyze Windows prefetch files
The teams with the best timelines used Windows prefetch files to show the exact times of execution, not only for the malware, but also for other programs the attacker used, such as using Mimikatz to search for passwords and using notepad to view and alter the web server credentials. A large portion of the teams found at least one of these prefetch files to reference in their timelines.

9. Ability to locate malware files
Recall that an anti-forensics technique the challenge employed was masquerading the RAT as a legitimate program, CCleaner. A little less than 50% of the teams located and pinpointed as malware this ccleaner.exe. Considering that anti-forensics was also used to cover the malware, 50% is a good indicator that the skill was appropriately tested [49]. Less of the teams noted the location from which this malware was executing, and this oversight also prevented teams from seeing the attacker populate that folder. There should be increased emphasis on the physical location of the executable in future exercises.

10. Ability to use hash verification
Teams should have provided hash verification for every file acquired, found, and carved. This included the hard disk and memory dump acquisitions, malware file, zip file, and defacement image file. Almost every team provided hash values for disk image and memory dumps they acquired. Many did not think to hash the pcap file after downloading it. The winning teams not only did this but also provided hash values for every file they found pertaining to the attack. 75% of the teams provided at least one hash value; however, the importance of using hash values throughout the investigation needs to be better stressed. For starters, teams should have been rewarded on the score sheet for using hash verification on the pcap download.

11. Ability to use Windows system administration tools
The big artifact the teams should have found using system administration tools is the persistence entry for the malware file. Almost 50% of the teams found this, most using Startup Manager. Many teams did not think this skill was tested enough, and in future other artifacts should be set to test this skill [50].

Category 2: Procedural Skills

1. Ability to use logging tools
This skill was not really tested well. Although teams had the opportunity to use logging tools in their investigation, none used them, or at the very least none presented their use in their reports. This should be stressed somehow in order to encourage better investigation documentation.

2. Ability to describe forensics activities performed
The ability to write a good forensics report really set teams apart. The inject was designed such that the teams were not only asked to find artifacts but also to produce a report that would help their team convince the afflicted third party not to attack them. The designers gave the teams templates to help them draft this report. They received mixed feedback on these templates. Some of the teams appreciated the separation between evidence items (i.e. a different report for each evidence item analyzed). This approach was cited as being more legal in approach [49]. Yet, teams with more incident response experience found the division of reports distracting to the overall story of the incident [50]. Teams drafted both reports shorter than half a page and novel-length reports. The best reports very well described the full attack. One way to better test this skill may be less stress on appendices and more on the overview of findings, as many teams submitted pages out output from their analysis reports with little description of what they indicated.

3. Ability to present data logically
Most of the teams agreed that the reporting requirement did test their ability to logically present data, because it stressed the need for good timelines [50]. From the perspective of the exercise coordinators, the skill was tested most with the blind scoring methodology. Teams were not given detailed explanations of how they would be scored; thus, they could not tailor their efforts just towards earning points [34]. This forced the teams to try and logically show in their reports their technical expertise and the importance of the activities they performed [50].

4. Ability to document forensics procedures
A review of the chat logs shows that teams asked about the level of detail to be provided in the report. Chat logs show the designers stressing to the teams the importance of repeatability, which was also emphasized in the template. Repeatability is the best way to test if the procedures were detailed enough [18].

Category 3: Teamwork

1. Ability to prioritize the investigations
A major intention of the inject design was to force the teams to prioritize investigating the most suspicious machine given the small time window the teams had to turn in their report. The inject tested this well, because the winning teams realized this and focused on the Windows 10 machine[50]. Another area in which the inject successfully tested prioritization was in applying pressure to the teams through the use of media injects. Some teams felt the challenge almost forced this too much, because they felt very time crunched [50]. However, that many of the teams achieved close to 75% of the points through proper prioritization shows it was about the right level of pressure [49].

2. Inter-team communication
At one point during the exercise, a team leader expressed frustration with the digital forensics designers for communicating directly with the digital forensics experts on the blue teams in requesting status updates. They felt that this skipped proper chain of command. This raises a question of if it is more realistic to speak directly to the technical experts or go through a management level and force more inter-team communication. Although not ideal, it is more realistic to insert a management level for state-based exercises [50].

3. Designation and division of roles and responsibilities
Ideally, the teams should have divided up the labor to look at different evidence items simultaneously. It appears that the best teams had someone analyzing the pcap while the others worked on acquisition. Teams that scored poorly failed to make this division of roles. One of the best teams, however, was a one-man operation, so it was possible to do the challenge without too much division of labor. An idea for improvement would be using more than one victim machine in the subnet, but this would also increase the acquisition requirements with which the teams already struggled [49].

Level of Technical Difficulty

The wide differentiation between teams makes it hard to judge the difficulty of the challenge based on total scores alone (which range from 50 to 7750). Still, when the scores are plotted

(see Fig. 6), it appears to be well balanced. The highest scoring team earned about 75% of the possible points, which is good, because it means the most capable teams could perform most of the tasks but were still challenged [49]. Unlike last year's Locked Shields, in which none of the teams managed to find a large number of artifacts, every scored item was uncovered by at least one team this year [49]. This shows improvement in designing a better inject and better score sheet.

In terms of specific difficulties, the forensic challenge chat room logs show that teams first stumbled in connecting to the virtual machines. It took them longer than anticipated to realize they needed to disable the firewall, but leaving this firewall up was an acceptable hurdle, because it was an obstacle all the teams without other technical issues managed to surmount. Next, teams struggled with managing their time. Recall that at the end of Day 0, the designers realized that the teams were trying to acquire images and memory dumps of all four of the machines. The written challenge inject clearly indicated that they should only acquire these from relevant machines. This thesis author provided an extra hint in the chat room that the pcap could help narrow their acquisition requirements: “pcap analysis may narrow down you acquisition requirements.” This should be more clearly hinted at in the inject directions. Technically speaking, almost all the teams were able to make hard disk and memory acquisitions.

In terms of the analysis, the lowest scoring team found only the attacker IP address. Teams received in the inject the IP range for the Crimsonian government, and the team was able to search for IP addresses in this range and follow the sessions between them and the victim subnet. The next lowest scoring team was able to find the web attack in the pcap by locating the uploaded theme. Both of these teams earned points only for analyzing the pcap (the second lowest scoring team was able to acquire but not analyze the memory). The next highest scoring team also found the attacker's failed and successful log-ins to the web server. On the other end of the spectrum, the highest scoring team found nearly everything in the pcap except for the image file with EXIF data. Only one team was able to locate this image inside of the uploaded zip file and the attempted code execution. This indicates that the pcap challenge was appropriate in difficulty with the image being obtainable but very difficult. Altogether, there was nothing in the pcap that was scored that went undiscovered. A step up in the challenge would be successfully adding the PHP script to the image.

The highest scoring team did not excel in memory analysis, but many teams did. The biggest miss by the teams that did the best in memory analysis was in locating and noting the parent process and mutant related to the malware process. However, a few teams did find these items. Out of a possible 1350 points for memory analysis, almost half of the teams obtained nearly 50% of the points. This indicates the designers could slightly increase the challenge for the higher performing teams. In terms of hard disk analysis, as noted above, the teams that scored best overall scored best in hard disk analysis. One of the biggest point losses for many of the teams was not finding the hidden folder where the attacker put all his tools, including the python files. Still, it is likely that this is more of an issue of the teams forgetting about that part of the inject, not a technical inability to find it [49]. Few teams noted the original folder to which the attacker downloaded the malware, and only one team showed the additional files with which the attacker also populated it. This attempt at covering the malware with legitimate files may be a bit too complicated for an exercise of this length, as it involves looking further back into the history of the attack.

The winning team lost the most points for not discovering the RDP session information that showed the malware created a remote session back to the attacker for the duration of the web server attack, but many of the other teams did locate and explain the importance of these sessions in their reports. The most difficult part about these sessions was that the sessions were encrypted. The challenge did not score decrypting these sessions formally, but if teams were able to do some of the encryption then they were given bonus points. Two of the teams managed to partly decrypt the traffic. This should remain a bonus and not part of the formal inject, as even the top scoring teams could not finish the task.

Another consideration for difficulty was the hint sheets. Only three teams requested them. The teams reported being scared by the penalty of only being able to earn 50% of the points in that category [50]. In practice, this would have affected barely any of the teams, anyway, based on their scores. The hint sheets did help those three teams obtain some points in the category, in one case up to 25% of the points in that category where they would have earned less than 10% [50]. This indicates that the hint sheets were well crafted but the way they were advertised was not, and the chosen the penalty needs to be adjusted. Although the cap method is good in theory, it actually did not affect any of the teams that used the hint, as all scored under 50% of the possible points regardless, and the 50% number scared too many of the teams from requesting help.

Timing

If there is one certainty to conclude from this test run, it is that the use of an exercise's preparation day for acquisition is essential. Without this time most of the teams would never have made it to the analysis portion. At the start of the acquisition phase, the designers gave the teams the pcap file. This was intended to help the teams decide what was pertinent to acquisition; the teams that realized this gained a vital time advantage, and this should be maintained for future exercises. While many of the teams complained that there was not enough time for acquisition, these were the teams that tried to perform hard disk images and memory dumps from all four machines in the victim subnet. The teams that used the pcap appropriately finished acquisition early in Day 0.

The teams had until about halfway through Day 2 to turn in their report. The idea behind this was that it gave the designers four hours to collect and score the reports before the end of the exercise. Even with this allocation, the designers finished scoring the reports just thirty minutes before the end of the exercise, and they missed a report that they had to go back and regrade after the exercise ended. It would actually be preferable to push the report deadline up, having the teams submit them an hour earlier. This hour could be used for verifying the scoring. This should be acceptable, because teams can spend the night between Day 1 and Day 2 writing the report if they are smart. Even though they cannot connect to the gamenet, they have all of the files (pcap, hard drive, memory dump) locally. With this hour, at five hours and twenty, it would give the scoring team approximately fifteen minutes per report.

One of the most frustrating aspects of this year's forensics challenge was that multiple of the teams wasted time getting acquainted with their investigator machine and the virtual environment during Day 0. One team never comprehended how to use its chosen machine. For other areas of the exercise, teams are given an acquaintance period, a few weeks before the exercise. It would be helpful to force the digital forensics experts of the teams to submit their investigator machine prior to this period and have them get acquainted with it at the same time that the rest of their team is getting familiar with the gamenet environment [49].

Place in the Overall Exercise

Remember that a key part of evaluating such an exercise is to involve the management levels. The leader of the Green Team, responsible for the technical infrastructure of the exercise, noted that the integration of the digital forensics infrastructure and the rest of the gamenet was quite weak. It would add to the realism and stress of the scenario if the events in the rest of the gamenet took some affect on the subnets where the digital forensics challenge took place [51]. The digital forensics scenario this year was more involved with the storyline of the rest of the scenario, particularly with adding in the file theft piece [49]. A data collector for the exercise commented that it would be interesting to make some of the data ex-filtration important information the teams need elsewhere in the exercise [52]

5. Summary

In this section the thesis author drew conclusions on the results of the test case and used interviews with exercise participants and management to determine sites for future development of cyber defense exercises similar to Locked Shields.

5.1 Conclusion

This thesis set out to establish the main goals of a digital forensics challenge in a multinational cyber defense exercise, develop a complete scenario for which to test them, and implement the challenge into a larger cyber defense exercise. The thesis author joined a team of designers at the NATO Cyber Defense Centre of Excellence in designing, developing, and implementing the digital forensics challenge for the 2016 Locked Shields cyber defense exercise. She then used first hand observations, written participant feedback, management level interviews, and competition scores to evaluate the effectiveness of the challenge.

Of the goals the thesis established in Table 1, the challenge most strongly tested pcap analysis, prefetch analysis, hard disk and memory acquisition, timelining, prioritization, and presentation of findings. The areas that need the most improvement are in systems log forensics, memory dump analysis, case logging, and description of activities. In particular, the thesis found that what a forensics report should look like varies greatly across teams, and there needs to be a good balance between providing guidance to inexperienced teams and leaving leeway for more experienced teams to describe the investigation in the method that is most logical for the scenario.

Areas in which the challenge excelled in terms of overall design and implementation included levels of difficulty and variety of tasks. The highest scoring team was still challenged, and the lowest scoring team still managed to obtain some points. The vast array of scores reflects the vast diversity of level of expertise across the teams. Areas in which the challenge fell short are in overall pre-exercise preparation, such as the need for the teams to have time to familiarize themselves with the virtual environment, and the need for more third party oversight, such as having a test run performed by individuals outside of the design team and having a check on the exercise scoring in order to prevent oversights such as the nearly missed report from one of the teams in Locked Shields 2016.

5.2 Future Work

Counter Forensics

One area for future development is the use of counter forensics, which is a division of anti-forensics aimed directly at frustrating forensics tools [53]. One such method is a compression bomb. Compression bombs are “small data files that consume a tremendous amount of storage when uncompressed” [53]. The purpose of these bombs is to cause forensics tools to fail or crash [54]. They can be highly frustrating to forensics investigations; yet, they are also easy to detect with anti-virus and other scanners if the bomb has a known signature [54]. The most well known compression bomb is the infamous 42.zip, which contains 42 kilobytes of compressed data and is widely available for download on the internet [54]. This is an ideal tool, because it is easy to both implement and detect. One idea is to disguise this zip as a file

that the analysts want to analyze (such as the malware sample). The designers chose not to use this bomb this year, as the 2015 Locked Shields proved so difficult for many teams, but given the increased performance levels in 2016 it is viable for next year.

Customization

An early intention in the design of this inject was to use an anti-forensics tool that would hide the malware from acquisition tools. Dimentia is a tool that supports two methods of data hiding: user-mode, which hides the program from hard disk acquisition tools, and kernel-mode, which hides the program from memory acquisition tools [43]. The issue this year is that both modes require all of the parameters to be specified at the time of program execution, which reveals to the investigators everything about what the program is hiding [30]. For future use, the program could be customized to include the parameters of the program within the executable code. This requires the design team to have a programmer skilled in .NET. The RAT that was employed in this test case scenario was also written in .NET. It would be nice to customize the program to have additional attribution clues.

Attack Vector

The last two Locked Shields cyber exercises have used email as the attack vector. While this is very realistic, cyber exercises should not become predictable [50]. An infection vector category predicted to grow immensely over the coming years is the use of Universal Serial Bus (USB) devices [55]. One of the most infamous examples of this is Stuxnet, malware that destroyed uranium enrichment centrifuges at the Natanz nuclear facility in Iran, which was introduced to the facility's computer system through a USB [55]. Many computer anti-virus programs detect and will not open anything from a USB without user approval [55]. A possible tool to get around this is the USB Rubber Ducky. This is a hacker tool that tricks the computer into thinking it is a keyboard [56]. Best of all, it works across multiple platforms, including Windows, Android, Mac, and Linux[56]. Another bonus is that the programming of the device is easily customizable. Such a method should be considered as a future attack vector.

Network Forensics

This year there was a bonus challenge involving IPv6. Teams had to investigate a data ex-filtration incident and figure our the source and destination IP addresses, type of ex-filtration method, and content of the ex-filtration. Only one team fully completed the challenge, but a handful at least solved part of it. The method used both IPv4 and IPv6 to split the packets among stacks. A suggestion from one of the exercise contributors would be to have an encryption key or important file ex-filtrated in this way for next year's exercise [52].

Test Run Inclusion

The Locked Shields exercise has an annual test run about one month before its scheduled execution. Since its inception, the digital forensics challenge has not been a part of the test run. Although the digital forensics team managed to do its own baseline investigation to develop the score sheet, the investigation is biased by the fact that the investigators know what for what they are searching [49]. Cyber exercises should aim for a third party to perform the investigation prior to the exercise's execution. It would help not only gauge the difficulty but

also get a more accurate idea of how long the challenge will take. For instance, the score sheet for Locked Shields 2016 awarded very little points just for acquisition, assuming this would be an easy task. It would have been more appropriate to give more points to this, including points for prioritizing the acquisitions of the Windows 10 machines.

References

[1] "ENISA Cyber-Exercises Analysis Report V1.0," ENISA, Oct. 2011.

[2] "About the Games," Cyber Olympics and EC-Council Foundation (2016), from https://www.cyberlympics.org/about-the-games/

[3] "History," National Collegiate Cyber Defense Competition (2016), from http://www.nationalccdc.org/index.php/competition/about-ccdc/history

[4] "Panoply," University of Texas (2012), from http://cyberpanoply.com/

[5] "Fact Sheet: NSA/CSS Cyber Defense Exercise- After Action Report," National Security Agency (2010), from https://www.nsa.gov/public_info/_files/press_releases/cdx_fact_sheet.pdf

[6] "US Naval Academy Triumphs in NSA Cyber Defense Exercise," US Navy (2015), from http://www.navy.mil/submit/display.asp?story_id=86634

[7] "Locked Shields 2015," NATO CCDCOE (2015), from https://ccdcoe.org/locked shields-2015.html

[8] M. Sadlon, P. Zdzichowski. 'Locked Shields 2015 Forensic Challenge.' 2015, 06 March. Available from NATO CCDCOE.

[9] M. Sadlon, T. Svensson, P. Zdzichowski. 'Locked Shields 2015 Final Report.' 2015, May. Available from NATO CCDCOE.

[10] K. Andreasson. *Cybersecurity: Public Sector Threats and Responses* (CRC Press: 2011).

[11] S. Jordan, "A Roadmap for CIOs & CSOs After the Year of the Mega Breach," Information Week (2014), from http://www.darkreading.com/attacks-breaches/a-roadmap-for-cios-and-csos-after-the-year-of-the-mega-breach/a/d-id/1269679

[12] T. Bailey, J. Brandley, J. Kaplan, "How Good Is Your Cybersecurity?" McKinsey&Company (2013), from http://www.mckinsey.com/insights/business_technology/how_good _is_your_cyberincident_response_plan

[13] "GIAC Forensic Examiner Certification: GCFE," SANS Institute (2015), from http://digital-forensics.sans.org/certification/gcfe.

[14] "CCE Certification Competencies," International Society of Computer Examiners (2012), from http://www.isfce.com/policies/CCE%20Certification %20Competencies.pdf.

[15] C. Gross, “Digital Forensics RAM Analysis,” University of New Mexico (2014), from nest.unm.edu/index.php/download_file/view/52/156/

[16] R. Belani and K.J. Jones, “Web Browser Forensics, Part 1,” Symantic Corporation (2010), from http://www.symantec.com/connect/articles/web-browser-forensics-part-1.

[17] *Digital Forensics Handbook, For Teachers*, ENISA, 2013.

[18] B. Garnett, “Intro to report Writing for Digital Forensics,” SANS Institute (2010), from https://digital-forensics.sans.org/blog/2010/08/25/intro-report-writing-digital\ forensics/.

[19] F.B. Cohen, “Fundamentals of Digital Forensic Evidence,” California Sciences Institute (2010), from http://all.net/ForensicsPapers/HandbookOfCIS.pdf.

[20] “Academy Teams Do Battle in Cyber Defense Exercise,” *Defense Systems* (2014), from https://defensesystems.com/articles/2014/02/19/serviceacademycyberexercise.aspxadmgarea=DS

[21] L.J. Hoffman and D. Ragsdale, “Exploring a National Cyber Security Exercise for Colleges and Universities,” Cyber Security and Policy Research Institute, George Washington University, Rep. CSPRI, Aug 2004.

[22] “Network Security Responder Program,” Ver. 2, National Computer Forensics Institute (2009), from https://info.publicintelligence.net/NITROstudentV2.pdf

[23] “ACPO Good Practice Guide,” Association of Chief Police Officers (2012). from, http://www.digital-detective.net/digital-forensics-documents/ACPO_Good_Practice_Guide_for_Digital_Evidence_v5.pdf

[24] M. Kelley, “Report Writing Guidelines,” Digital Forensics Investigator News (2012), from http://www.forensicmag.com/articles/2012/05/report-writing-guidelines

[25] M. Sadlon, T. Svensson, P. Zdzichowski. ‘Locked Shields 2015 Preliminary Report.’ 2015, May. Available from NATO CCDCOE.

[26] S. Chevaliar, H. Dang, K. Kent, T. Grance, “Guide for Integrating Forensics Techniques into Incident Response,” National Institute of Standards and Technology (2006).

[27] M. Sadlon, T. Svensson, P. Zdzichowski. ‘Forensics Challenge Wiki’ (2015). Available from NATO CCDCOE.

[28] “LS15_BT_systems.png,” Locked Shields Collaboration Portal (2015).

[29] “Exploring the Nature of a Cyber Attack,” E-Estonia (2012), from https://e estonia.com/exploring-nature-cyber-attack/

[30] M. Sadlon, Personal Communication (March 2016).

[31] “CCDC Mission,” National Collegiate Cyber Defense Competition (2016), from http://www.nationalccdc.org/index.php/competition/about-ccdc

[32] L. Lightner, C. O'brien, P. Pusey, “Preparing for the Collegiate Cyber Defense Competition (CCDC),” National Cyber Watch Center (2014), from https://scout.wisc.edu/cyberwatch/downloads/62

[33] G. Kessler, “Anti-Forensics and the Digital Investigator,” Champlain College (2007), from http://www.garykessler.net/library/2007_ADFC_anti-forensics.pdf

[34] R. Ottis, Private Communication (March 2016).

[35] Z. Bu, P. Bueno, R. Kashyap, A. Wosotoesky, “The New Era of Botnets,” McAfee Labs (2010), from http://www.mcafee.com/us/resources/white-papers/wp-new-era-of-botnets.pdf

[36] A. Howitt, R. Pangi, *Countering Terrorism: Dimensions of Preparedness,* MIT Press (2003).

[37] P. Doshi, S. Siddharth, “Five Common Web Application Vulnerabilities,” Symantec (2006), from http://www.symantec.com/connect/articles/five-common-web-application-vulnerabilities

[38] P. Zdzichowski, Private Communication (March 2016).

[39] “Recognizing and Avoiding Email Scams,” US-CERT (2009), from https://www.us-cert.gov/sites/default/files/publications/emailscams_0905.pdf

[40] “Spear-Phishing Email: Most Favoried APT Attack Bait,” Techno Mocro Inc (2012), from http://www.trendmicro.com/cloud-content/us/pdfs/security-intelligence/white-papers/wp-spear- phishing-email-most-favored-apt-attack-bait.pdf

[41] J. Kick, “Cyber Exercise Playbook,” MITRE (2014). Available: https://www.mitre.org/sites/default/files/publications/pr_14-3929-cyber-exercise playbook.pdf

[42] S. Sjouwerman, “Defeating Phishing and Spear Phishing Attacks,” Educause Review (2015), from http://er.educause.edu/blogs/2015/10/defeating-phishing-and-spear phishing-tactics

[43] T. Väisänen, M. Sadlon, P. Zdzichowski, “Anti-Forensic Study,” NATO Cyber Defense Centre of Excellence (2015), from https://ccdcoe.org/sites/default/files/multimedia/pdf/AF_wit%20intro.pdf

[46] R. Lee, “SIFT Workstation 2.0 Distro Version,” SANS Institute (2010), from http://www.jwgoerlich.us/files/sift-tool-listing.pdf

[47] “Package: plaso (1.4.0+dfsg-2),” SPI Inc (2016), from https://packages.debian.org/sid/admin/plaso

[48] N. Wilhelmson, T. Svensson, “Handbook for planning, running and evaluating information technology and cybersecurity exercises,” Center for Asymmetric Threat Studies(2011), from https://www.fhs.se/Documents/Externwebben/forskning/centrumbildningar/CATS/publikationer/Handbook%20for%20planning,% 20running%20an%20evaluating%20information%20technology%20and%20cyber%20security%20exercises.pdf

[49] P. Zdzichowski, Private Communication (March 2016).

[50] “Feedback Surveys,” Private Communications (April 2016). Available from the author.

[51] R. Rattas, Private Communication (April 2016).

[52] T. Lepik, Private Communication (April 2016).

[53] S. Garfinkel, “Anti-Forensics: Techniques, Detection, and Countermeasures,” Navel Postgraduate School (2006), from http://citeseerx.ist.psu.edu/viewdoc/download doi=10.1.1.109.5063&rep=rep1&type=pdf

[54] M. Olivier, S. Shenoi, *Advances in Digital Forensics II* (Springer: 2010).

[55] S. Cobb, “Are Your USB Flash Drives and Infection Malware Delivery System?” W Live Security (2012), from http://www.welivesecurity.com/2012/12/11/are-your usb-flash-drives-an-infectious-malware-delivery-system/

[56] “USB Rubber Ducky Deluxe,” Hakshop (2016), from http://hakshop.myshopify.com/products/usb-rubber-ducky-deluxe?variant=353378649

Appendix

I. Forensics Report Template

Forensics Analysis Report

Blue Team:

Evidence Item:

I. Overview

The overview must include the reason for the analysis examination and a summary of the most important findings of the analysis.

II. Items

List here all hardware and software used during the acquisition and analysis, including version number.

III. Acquisition

The acquisition section must include all steps taken to preserve the evidence (e.g. write blocker) and hash verification (e.g. SHA-1, MD5) of the images. For network analysis (PCAP) include file download and verification.

IV. Findings

The findings section must include all analysis types performed on the evidence item, important items uncovered (e.g. deleted files, suspicious running programs) and pertinent results of those findings.

V. Timeline

Construct here a basic timeline of events based on the results of the analysis, from system compromise to acquisition point. Show how time was obtained (e.g. system log time, file metadata).

VI. Appendices

Add here any additional screen shots, hashes, references important to the analysis.

II. Forensics Report Example

Forensics Analysis Report

Blue Team: Example Team

Evidence Item: 003

I. Overview

The overview must include the reason for the analysis examination and a summary of the most important findings of the analysis.

On 01FEB2016 an image was uploaded to the company webserver containing javascript to send session cookies to the email address badGuy001@gmail.com. The file was uploaded from an internal ip address. Employee X was seen inserting a USB, against company policy, to his machine minutes before the image was uploaded.

II. Items

List here all hardware and software used during the acquisition and analysis, including version number.

Hardware:

- Lenovo Thinkpad Yoga 14
- Sandisc Cruzer Blade CZ50 16MB

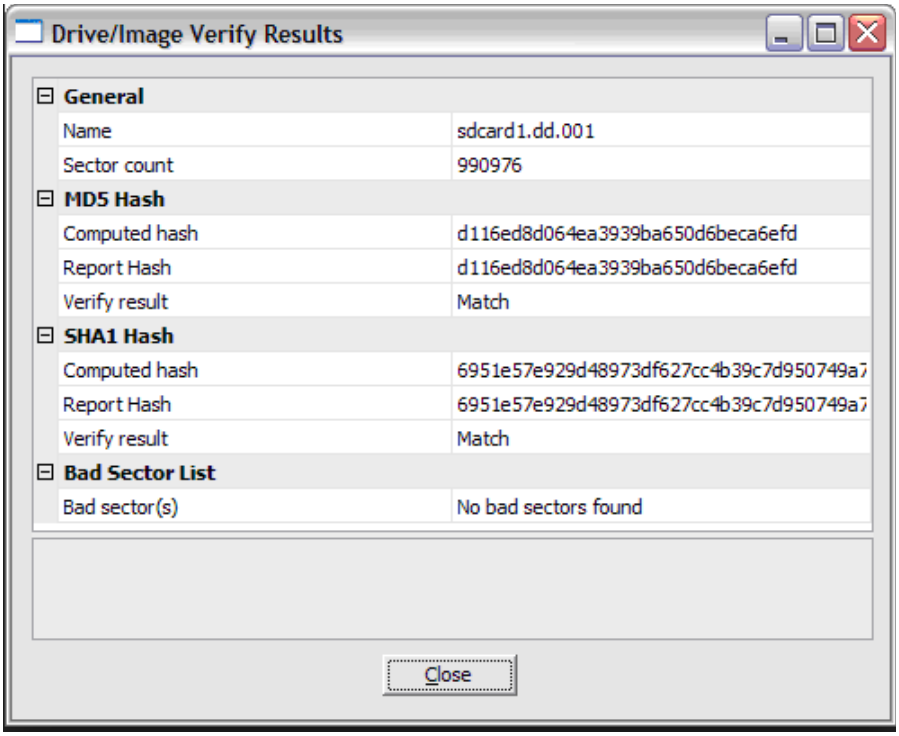
Software:

- Ftk Imager Lite 3.1.1
- Autopsy 2.0
- FileFormat.info/tool/hash.htm

III. Acquisition

The acquisition section must include all steps taken to preserve the evidence (e.g. write blocker) and hash verification (e.g. SHA-1, MD5) of the images. For network analysis (PCAP) include file download and verification.

1. Using the software FTK Imager Lite 3.1.1, analyst acquired an image and applied a write blocker. Selected options to compute the hash prior and after acquisition.



IV. Findings

The findings section must include all analysis types performed on the evidence item, important items uncovered (e.g. deleted files, suspicious running programs) and pertinent results of those findings.

1. Added the USB Image to Autopsy 2.0 and ran all default analysis modules.

2. Navigated to image folder and selected the “show thumbnails” tab.

    Result: no images found

3. Navigated to the deleted files folder

    Result: image from the webserver deface.PNG is located

3. recovered the image to the hard drive of the analysis computer and used online hash calculator to compare hash value of the file to that of the uploaded image file:

Results	
Original text	*(binary only)*
Original bytes	ffd8ffe000104a46494600010100000100010000ffdb008400... (length=10637)
Adler32	853d60cf
CRC32	cd8cd31e
Haval	c36a4ba78b8b7228da766c5e7c8b92a4
MD2	15ca7dc34bfe7085a52655a08cc4a9eb
MD4	8c423ee74c1ebeb49ce25c1f7bf8b411
MD5	0591906129205c82df23a04b877b8b54
RipeMD128	61c9a1c35109da4e6546b7481b381d82
RipeMD160	b3f42b7d04578f61e490b3c3d622e9dc2db2de49
SHA-1	bdad400e59d15bed686d75d05cea0e7add452c7f
SHA-256	274ae73215b88d1ea6dfde92e2f8ebea4fac807bc01e168e13e17218940679ee
SHA-384	f484bbc34627ed91dc0cb373e68abdfe623d3ec1617f03a9d9326998c7599d983873c08d9be9aeb9b1f4f71ef36e7802
SHA-512	6b7d443c1386ed9d9c794697d0be296e9bf8ff45ace2de12f981a5f09da0c93aa48215dd0341e66e416c59dd821e082627112152 81b00e41ca254895a66a4cef

V. Timeline

Construct here a basic timeline of events based on the results of the analysis, from system compromise to acquisition point. Show how time was obtained (e.g. system log time, file metadata).

01FEB2016

- 0830, employee X logs into his computer
    - source: system log shows userX sign in
- 0833, employee Y notices employee X plug a USB into his machine
- 0835, file is uploaded to the company home page
    - source: network logs show file upload of 900 KB from ip 192.168.10.24
- 0840, employee Y reports employee X's actions to Mr. Joe Shmoe to security desk
- 0845, employee X is asked to produce the USB, which he does
- 0900, acquisition of the USB image is made

VI. Appendices

Add here any additional screenshots, hashes, references important to the case:

- See evidence report 001 for analysis of image uploaded to the company webserver
- See evidence report 002 for network traffic analysis

III. Hint Sheets

Your Five Hints-- Instructions

During Day 1 and 2 you may request hint sheets for pcap analysis, memory dump analysis, and/or hard disk analysis. These sheets each contain five hints to help you direct your analysis efforts. They are not answer sheets. When you request an answer sheet, the total number of points you can earn for that category is reduced by 50%. For example, if you could earn up to 500 points in the category before requesting the sheet, after the request you can only earn up to 250 points. You can request one of these 3 sheets-- pcap, memory, or hard disk, through the chat channel.

Your Five Hints-- PCAP

1. Look for an abnormal TCP port connection

2. Search for connections to the web server address 129.42.65.189

3. Look at posts made from the victim machine to the web server

4. Look for file downloads to the victim machine 10.1.10.17

5. Look at image EXIF data

Your Five Hints-- Memory Analysis

1. Look at registry autorun executables

2. Look for mutants in suspicious processes

3. String search process dumps

4. Startup Manager

5. Netscan

Your Five Hints-- Hard Disk Anaysis

1. Search for email attachments

2. Look at prefetch files for file executions

3. Look for hidden folders

4. Look for signs of Spear Phishing

5. Search Windows Event Logs for suspicious connections

IV. License

Non-exclusive licence to reproduce thesis and make thesis public

I, Allyson Hauptman

(*author's name*)

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to:

   1.1. reproduce, for the purpose of preservation and making available to the public, including for addition to the DSpace digital archives until expiry of the term of validity of the copyright, and

   1.2. make available to the public via the web environment of the University of Tartu, including via the DSpace digital archives until expiry of the term of validity of the copyright,

of my thesis

Designing Digital Forensics Challenges for Multinational Cyber Defense Exercises

(title of thesis)

supervised by Patrycjusz Zdzichowski, Rain Ottis, and Raimundas Matulevicius ,

(supervisor's name)

2. I am aware of the fact that the author retains these rights.

3. I certify that granting the non-exclusive licence does not infringe the intellectual property rights or rights arising from the Personal Data Protection Act.

Tartu, **26.05.2016**