Automating Cyber Offensive Operations for Cyber Challenges

Conference Paper · March 2016

Ivan Daniel Burke
Council for Scientific and Industrial Resear...

Renier van Heerden
Council for Scientific and Industrial Resear...

Automating Cyber Offensive Operations for Cyber Challenges

Ivan Burke and R.P. van Heerden
CSIR, Pretoria, South Africa
iburke@csir.co.za
rvheerden@csir.co.za

Abstract: Cyber awareness training has become a growing industry, with more and more organisations starting to focus on training personnel on how to behave in a secure manner when engaging in cyber operations. Cyber challenges place participants in realistic cyber defence scenarios in order to provide training under fire. This paper documents steps taken to develop an automated attack capability for use within a cyber challenge environment. The challenges discussed within this paper focuses on cyber challenges conducted within developing countries, such as South Africa, but the principles discussed within this paper aim to be applicable to be applicable to all cyber challenges in general. The researchers based their work on prior publications covering threat modelling, construction of cyber security testbeds and planning of offensive cyber operations. The work presented in this paper is a practical application of an ontological model for cyber attack scenarios.

Keywords: cyber challenges, ontology, attack automation

1. Introduction

Security professionals can no longer with certainty state that, their critical systems cannot be breached. Hence it is of vital importance to constantly and repeatedly test and hone their skills. Cyber challenges or capture the flag (CTF) exercises are but some of the possible training tools available to security professionals to improve their practical experience with regards to mitigating a cyber-attack.

In this paper the authors will discuss the steps that were taken to construct an automated offensive cyber capability for use in a cyber gaming environment. In previous work, a cyber attack ontology was developed to model the core aspects of a cyber attack. This ontological model is used to define offensive behaviour of cyber attackers at various skill levels. By means of automation a methodical and repeatable automated attack mechanisms were created to represent cyber attacks. These automated attacks were then placed within a Cyber challenge environment to validate the theoretical models.

This paper provides a brief history of CTF and cyber challenges in Section 2, as well as, why CTF challenges are of particular importance to developing countries, such as South Africa. In Section 3, an overview of previously published work by the authors is presented. In Section 4, the current state of the Council for Scientific and Industrial Research's Cyber Challenge infrastructure is presented, as well as, the enhancements made with regards to attack automation.

2. Background on cyber games

CTF exercises are offensive and defensive cyber training exercises whereby teams compete against one another to obtain flags or tokens hidden on various servers. In some cases these flags are located on an opponent's servers, which results in teams being required to attack adversary servers while maintaining defences on their own services (The National Cyber League 2013). These challenges are often time based and victors are determined based on the number of flags acquired. The first formalized CTF contest was held during Defcon 4 in 1996 (DEF Con 2011). Since then several variants of the concept has emerged such Social Engineering Capture the Flag (Fincher & Hadnagy 2014)and Wireless Capture the Flag (Wireless Village 2014). CTF is by no means the only means of providing practical experience for cyber security training but it has proven to be one of the most common forms of training. In this article the more general term Cyber Games will be used as a blanket term to incorporate all forms of gamification of cyber security training.

The term gamification refers to combining basic gaming design elements with mundane or repetitive tasks to improve user experience when performing these tasks (Marczewski 2013). Cyber Games provide an engaging environment for security professionals to practice and improve their craft by means of active defensive and offensive exercises. These team exercises also improve group cohesion and communication skills by forcing the team to respond to time critical cyber events in an efficient manner. Cyber Security Incident Response Teams (CSIRT) can use these exercises to practice responses to real life scenarios in a controlled environment. These practice runs may asset CSIRT teams to identify shortcomings in current response plans, as well as, help to identify possible shortcomings of response actions.

Boopathi et al. (Boopathi, Sreejith & Bithin 2015) provides a compelling argument for cyber security training through gamification. In their paper, Boopathi et al. proposes a three tier CTF challenge aimed at graduate and under graduate level to address the country's growing need for security professionals within India. The proposed CTF challenge was called InCTF and its structure is depicted in Figure 1: InCTF phased structure. The three tiers are: Learning, Jeopardy and interactive training. During the learning phase focuses on conveying important security concepts and tutorials. The lessons learnt during the learning phase will be tested in the follow phases. The Jeopardy phase was subdivided into four challenge levels each testing specific concepts which were taught during the Learning phase. Participants need to complete at least three out of the five challenges of each level before being able to access the next level of challenges. The first level tests the participant's basic coding puzzles such as bug fixing and error correction. The second level focuses web application security, such as, SQL inject, Cross Site Scripting (XSS) and Cross Site Request Forgery (CSRF). Level three focused on application security, such as, buffer overflow, attacking the stack and string format attacks.

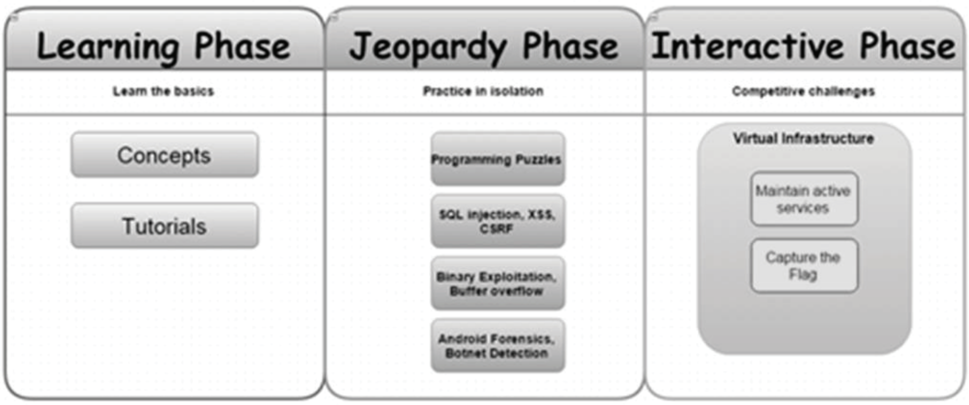

Figure 1: InCTF phased structure

Boopathi et al.'s work is of particular importance to the authors because of the striking similarity to the current state of cyber security training and awareness training available within South Africa. The ICT industry, within South Africa, has grown rapidly over the past ten years and the current curriculum at under graduate and graduate level is not adapting fast enough to keep current with the growing need for security professionals. Since 2012, the Council for Scientific and Industrial Research (CSIR) of South Africa has conducted its own annual Cyber Game challenges during the Cyber Security Awareness Month. Initially the challenges were aimed at interns and new employees as a means of training them in basic cyber security concepts but since then, the challenges have been expanded and offered the training to graduate and undergraduate students at various South African universities and even some corporate professionals. The results and observations of these cyber games will be discussed throughout the following section.

3. Previous work by authors

This paper builds on previous research outputs created by the authors. The previous work primarily focused on threat modelling, constructing cyber test environments and offensive cyber operations. This section provides a brief summary of previous outputs, as well as, an overview of how these outputs have been combined to develop an automated offensive capability for future CSIR cyber challenges.

3.1 Previous work on threat modelling

In previous work, an attack taxonomy and ontology was developed to classify types of cyber attacks (Van Heerden et al. 2012). This initial ontological model attempted to model the relationship between Attack Actors, Attack Targets and Attack Goals. The attacks were classified based on attack mechanisms used and automation levels. In this paper six network attack phases were identified. This attack model is depicted in Figure 2.

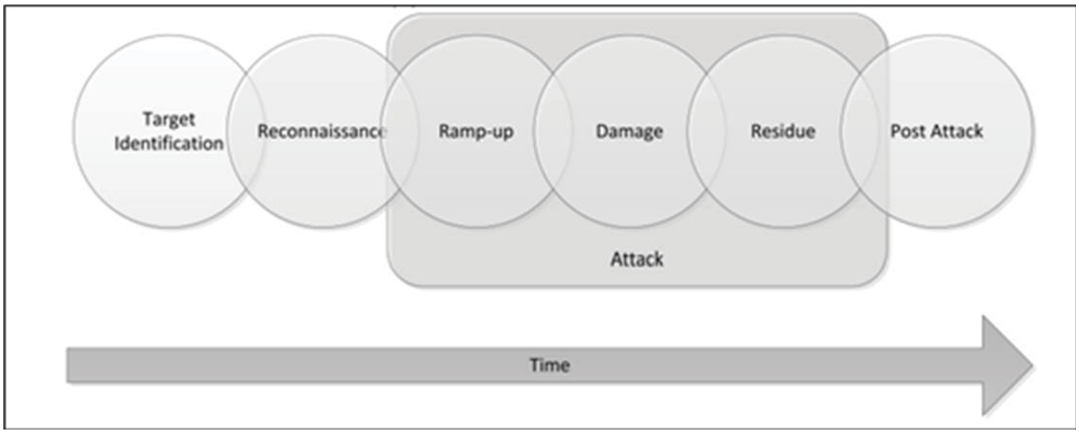

Figure 2: Non-discrete attack model

This classification of attacks and attack phases lead the development of the network attack prediction system, discussed in (Van Heeren, Pieterse & Irwin 2012) and expanded upon in Van Heerden's doctoral thesis (Van Heerden 2014). This attack prediction system aims to predict the end goal of a network attack by detecting attacks during the early phases of an attack. The system uses sensors to detect precursors of a network attacks during the reconnaissance and ramp-up phases of an attack. Sensors are single purpose attack detectors, which have specialized detection parameters. For example, the port_scan sensor simply listens on commonly unused ports for any ICMP, TCP or UDP scan attempts. If a sensor is triggered, data relating to the source of the attack is sent to the Network Attack Prediction Data Collection System, where all data relating to various sensors and attacks are collated to form a more coherent picture of the potential attack. The Network Attack Prediction System then performs various queries on the collected data, to classify attacks according to attack scenario and phase. This process is shown in Figure 3: The attack prediction process

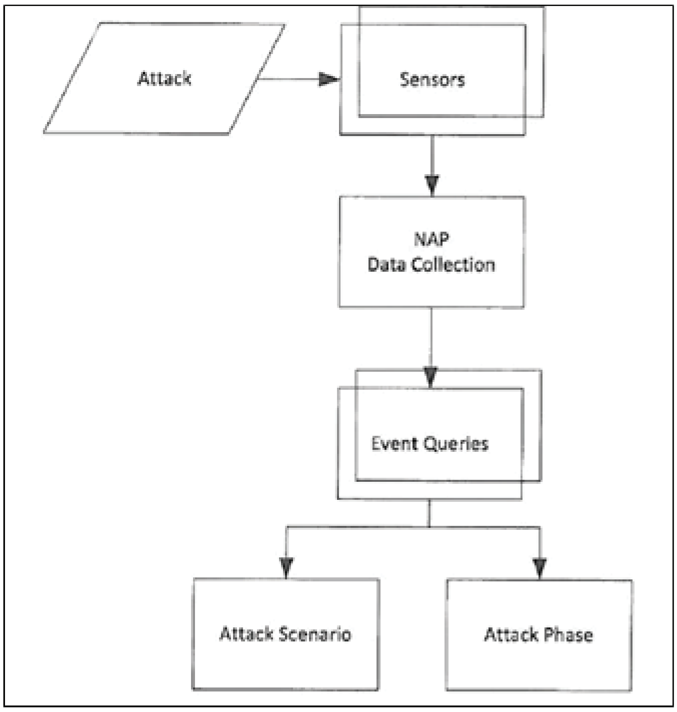

Figure 3: The attack prediction process

The predictive capability of the Network Attack Prediction system is in the mapping of sensors to potential attack scenarios and phases. A sample of such a mapping is shown in Table 1. Using this mapping one can predict possible next phases of an attack by detecting and classifying attacks early on. It should be noted that some overlap of sensors and phases are possible, hence the system cannot produce a 1:1 mapping of sensor event triggers to attack classification. For example, in Table 1, System Compromise and Resource Theft have similar Reconnaissance and Ramp-up triggers; yet their damage phases differ.

Table 1: Mapping of temporal phases of an attack to sensors

	Reconnaissance	Ramp-up	Damage
Web Defacement Attack	Unusual_Web_Activity	Vulnerability_Scan_Web_detector	Web_defacment-detector
System Compromise	Port_Scan, Failed_login_attempt	Vulnerability_Scan_Detector	Unauthorised_super_user_access
Phishing Attack	Fake_Email_Received	Fake_Email_Accessed	
Denial of Service		Network_Traffic_Ramp_Up	Server_Uptime_Sensor
Resource Theft	Port_Scan	Vulnerability_Scan_Detector	Trip_Wire, Directory_Watcher_trigger

In Table 1, all the sensors that are triggered during each phase of the attack are mapped. For example, to detect a web defacement event, The Network Attack Prediction system would deploy:

- an Unusual Web Activity sensor. This sensor attempts to detect abnormal web activity, such as, website crawling, directory discovery attempts and URL injection.
- a Web Vulnerability Scan Detector. This sensor attempts to detect common web vulnerability scanners, such as, JoomScan (Khant 2013), SQLMap (Damele & Stampar 2012) and WordPress Scan (Mehlmauer & Van Der Lier 2013).
- A Web Defacement detector. This sensor detects changes made to a web page based on a comparison between current web page and last known verified version of the page.

This example describes only a small subset of attack and sensor mappings. The full set of sensors and triggers for each sensor are discussed fully in Van Heerden (Van Heerden 2014). The use of the cyber attack ontology was practically demonstrated by mapping various high profile network attacks to a temporal computer attack model to identify key phase of a network attack (Van Heeren, Pieterse & Irwin 2012).

This work was extended upon by Chan (Chan et al. 2015). This subsection aims to highlight some of the key findings of the previous work. Chan et al. extended the ontological model to be used as a planning tool for cyber operations. A comparison of the ontological model to other methods of modelling cyber attacks was conducted by Grant et al. (Grant, Burke & Van Heerden 2012).

4. Implementation of automate offensive attacks

The offensive capability developed during this research will be deployed at two points within the cyber games infrastructure as depicted in Figure 4: CSIR CTF network infrastructure. The practice servers, as well as, the production server, which the teams will need to defend.

4.1 Current infrastructure

This sub-section will provide a basic overview of the current CSIR cyber challenge infrastructure. The infrastructure is illustrated in Figure 4: CSIR CTF network infrastructure.

Participants are invited to participate remotely or from the onsite libratory environment. A team consist of groups between three and five participants. Each team will require Virtual Private Network (VPN) connection to the CSIR virtual infrastructure, as well as, an internet connection.

The Internet connection is needed to allow participants access to the online challenges, flag submission system and the scoring system. The scoring system is simply a website containing the current leader board and any notifications sent by the organising comity.

The flag submission system is a separate website independent from the scoring system. The participants can submit flags to the scoring system. A flag is a unique twelve character alphanumeric string associated to each challenge. A flag is a unique twelve character alphanumeric string associated to each challenge. Participants are awarded a flag for performing specific challenges or tasks during the cyber games. For example, unmasking a shadow file and then cracking the passwords within, yields one flag per password cracked. Each flag has a score associated to it based on the difficulty of the challenge presented to acquire the flag. Easy tasks are plentiful and can be completed in rapid succession whereas, other tasks may require substantial effort and time to obtain, but they yield a greater reward. This variation in challenge difficulty provides easy access to novice participants whilst still providing challenging content for professionals.

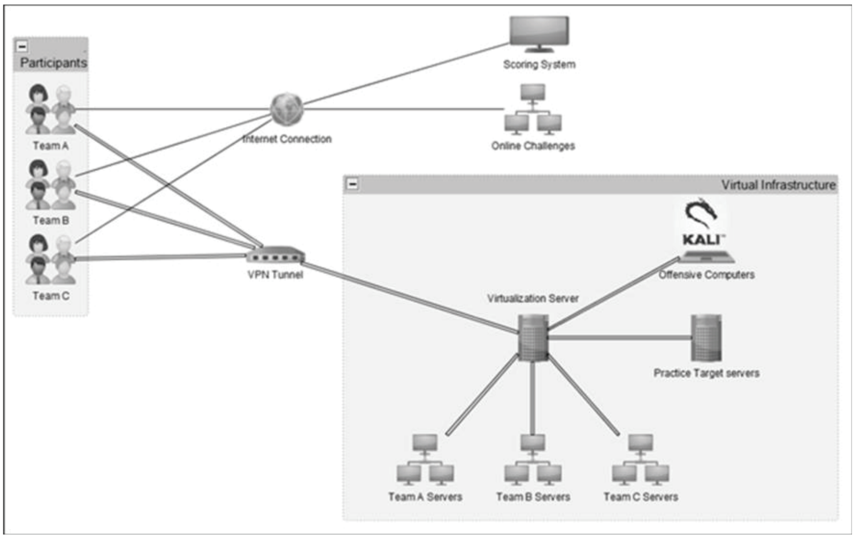

Figure 4: CSIR CTF network infrastructure

Each team has their own flag submission website. This is to prevent teams from attacking the challenge infrastructure to prevent people from submitting flags. This is a lesson learnt from previous cyber challenges where teams would build up a lead on the leader board and then launch a Denial of Servicing (DoS) attack on the scoring system, preventing other teams from submitting flags. This challenge is done online and attribution cannot be assured hence, the offending team cannot simply be penalised or have their connection severed. Each team's flag submission Uniform Resource Locator (URL) is kept secret and is hosted by various public web hosting companies. This additional precaution has lead to a sharp decline in attacks against cyber challenge infrastructure. This decline is shown in Figure 5: Offensive attacks observed during CSIR CTF challenges, which will be discussed in the following subsection.

The final online resource for the CSIR cyber challenge is the online challenge system. This system is a website containing entry level cyber challenges for the participants to solve. These challenges include cryptography, encoding, packet traces, steganography, forensic challenges, as well as, some basic puzzle and coding challenges. These challenges require basic security skills and are meant to be the introductory challenges. Several basic concepts required to solve these challenges will be expanded upon in later challenges. None of these challenges require any hacking tools or offensive attacks to be launched, since the challenge strictly prohibits attacks against the public infrastructure hosting the cyber challenges.

For the more advance challenges teams will need to connect to the Virtual infrastructure hosted by the CSIR. Teams can connect to this infrastructure by either establishing a Virtual Network Computing (VNC) or Virtual Private Network (VPN) connection to the Virtualization infrastructure. The virtualization infrastructure can be any assortment of virtual containers and or hypervisor components connected through virtual or physical networks. In the CSIR cyber challenge scenario a cluster of ESXi virtualization servers are interconnected to construct several practice targets and virtual corporate infrastructures. Each team is given the following virtual nodes to control in the virtual network:

- A number of offensive virtual machines with Kali Linux and MobiSec pre-installed. These virtual machines have all the offensive, analysis and forensic tools required for the challenges pre-configured. No internet access is given to these machines to prevent accidental attacks on none virtualized infrastructure. The guidelines set forth in Van Heerden et al. (Van Heerden et al. 2013) and (Burke & Pieterse 2014).
- A set of servers on which to practice offensive network attacks, such as, denial of service, data interception, web site vulnerability exploitation, executable reverse engineering and Android reverse engineering and malware analysis.

- Several production servers, running various services which need to be maintained, patched and protected from attacks to ensure service uptime. Each team has their own set of production servers to protect. These servers have several known vulnerabilities, back doors and root kits installed which needs to be identified and removed. At first these servers will not be accessible by other teams but as the challenges progresses teams will be granted routing access to opponent servers and tasked with attacking each other's servers.

Unlike the InCTF challenges, the CSIR CTF challenges are not unlocked by achieving a certain level of proficiency in previous challenges, but rather based on temporal events. Initially teams are only given access to their servers they need to protect, as well as, the online challenges, scoring system and flag submission system. After a set period of time the offensive virtual machines and practice servers will be activated; allowing participants to take on advance challenges. The final phase of the CSIR cyber games is a free-for-all style matchup where all participants are given access to each other's production servers and given the task to attack each other's servers.

4.2 Observations from previous CTF challenges

This sub-section provides a brief summary of attacks launched during previous CSIR cyber challenges.

The most cumbersome issue for the CSIR CTF challenges has been the implantation of the interactive attack phase. Figure 5 contains a graph summary of offensive attacks launched during the CSIR CTF challenges over the past four years. The Attacks are grouped into three categories: Attacks against winning teams, Attacks against Losing teams and Attacks against the CTF system. Winning and Losing teams are defined as the top (or bottom) two teams according to the scoring system at the time of the attack. The CTF system entails the entire CTF infrastructure, scoring system and interconnecting networks. An attack is defined as a disruptive, destructive or system compromise attack launched against an advisory. General recognisance and probing of advisory systems was not considered as attacks for this tally nor was repeated attacks. Each attack was only counted once even if it was launched repeatedly.

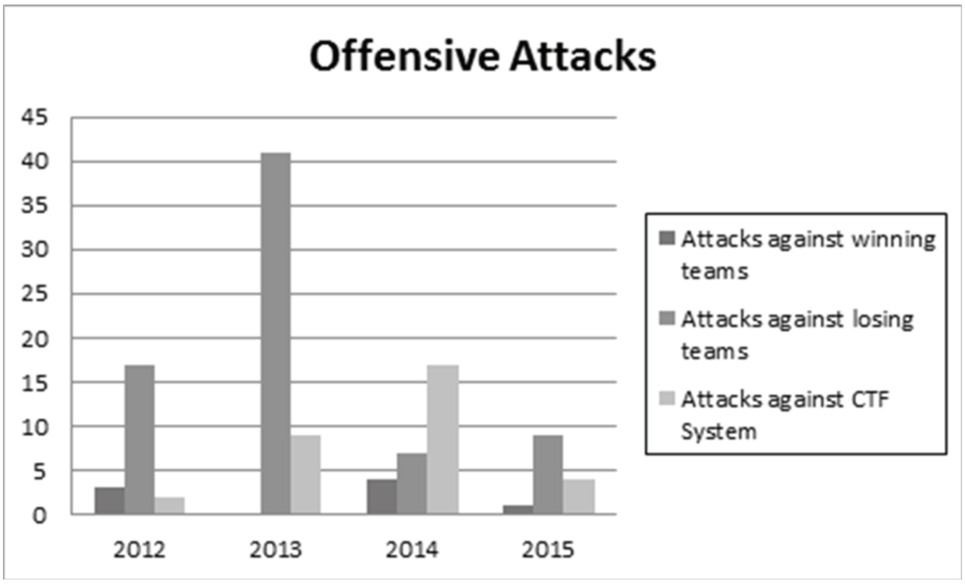

Figure 5: Offensive attacks observed during CSIR CTF challenges

The second year of the CSIR Cyber Challenge was the first year undergraduates and external companies could partake in the CSIR Cyber Challenge. During the 2013 CSIR Cyber Challenge, weaker teams were under constant attack from more advance advisories and in some cases teams were completely locked out of their systems for extended periods of time. This caused a sharp decline in novice participation in the following year. After concluding the 2013 CSIR Cyber Challenge researchers found that the winning teams actually had the majority of their systems left unpatched and with several of the back doors still active.

In 2014, with the decline in novice participants challengers started focussing more on attacking the CTF system than other advanced participants. The most common attack was DoS variants and interception attacks via Domain Name System (DNS) poisoning. These attacks were reduced by the Challenge organisers by creating separate flag submission systems and decoupling team into dedicated ESXi clusters.

In 2015, the attacks launched against the CTF infrastructure were still present but no longer disruptive to other teams. The overall attack count also went down substantially. A post analysis of the production servers revealed that production servers where still unpatched. In a post interview with the challengers it was revealed that certain teams would rather not attack, due to of fear of retaliation from other teams.

The current lack of offensive and defensive actions by participants is problematic since; the end goal of the Cyber Challenges is to prepare cyber security practitioners for real life defence scenarios. To address this issue the cyber challenges have been enhanced by introducing an automated offensive capability.

4.3 Offensive capability of practice servers

The practice servers will be equipped with several attack prediction sensors. These sensors act as traps for participants and emulate retaliation behaviour from practice servers. If a participant launches a particularly noisy or verbose attack on the target system, the participant will be penalised a few points but also presented with a new challenge to recuperate some of the lost points. The penalty for being detected also varies based on the skill level presented by the user. This perceived skill level is derived from the ontological model presented in Van Heerden et al. (Van Heerden et al. 2012). The ontological model is queried to determine the level of automation, skill threshold required, technical difficulty of attack and phase of attack detection to determine an appropriate penalty value.

For example, if a participant performs an intensive full port sweep using NMAP (Lyon 2011) to detect open ports or services, the port scan sensor will trigger and classify the reconnaissance operation as: an automated, low skilled and basic attack type. This would result in a high penalty given to the participant. However if the attacker used a NMAP scan for specific ports, in paranoid mode or even none standard scan operation, such as a Xmas scan (Lyon 2011), the sensor will classify the attack as: a crafted, medium skilled and basic attack, in the reconnaissance phase of an attack. This would result in a lower penalty. If the participant could avoid detection during reconnaissance phase and only get detected during the ramp-up or damage phase of the attack a even lower penalty would be assigned.

Once detected and penalised, participants will be informed of the detection, as well as, be presented with a brief summary of why the attack was detected, how to avoid future detections and the additional challenge to recuperate the lost points. These notifications allow participants to learn during the execution of the cyber challenge event, as well as, provide them with potential hints to protect their systems in the final stage of the cyber challenge.

4.4 Automated offensive attacks against production servers

In this phase of the CSIR Cyber challenge, participants are tasked with maintaining services on the production servers given to them. These servers are unpatched, infected with backdoors and have several well known vulnerabilities. Teams will need to patch and secure these servers before the automated offensive system starts assaulting the system. Teams are scored based on the uptime of these services. If at any point, during the final phase, these servers go off-line, the team will be penalised. To monitor service uptime, automated users are created to access and consume services. The network traffic automation is fully documented in (Van Heerden et al. 2013). For the purpose of this paper, the automated traffic generator emulates normal user activities, such as, website access and FTP access.

The automated offensive traffic is generated by altering the behaviour of these users act in a malicious manner. These malicious users aim to blend into normal traffic by acting like normal users until a pre-specified temporal event occurs. Participants are given the capability to block access to users via a dashboard. If the participant blocks a legitimate user an off-line event is triggered and the team will be penalised. If a malicious user is denied access, the participant will prevent the malicious user from continuing the offensive attack and the participant is awarded points for stopping the attack. Participants score higher points for preventing an attack during early phases of the attack. If the attack is allowed to proceed to its damage phase, the service will be disrupted and off-line events will be triggered by all users. The complete logic flow of an unauthorised user access attack generator is depicted in Figure 6.

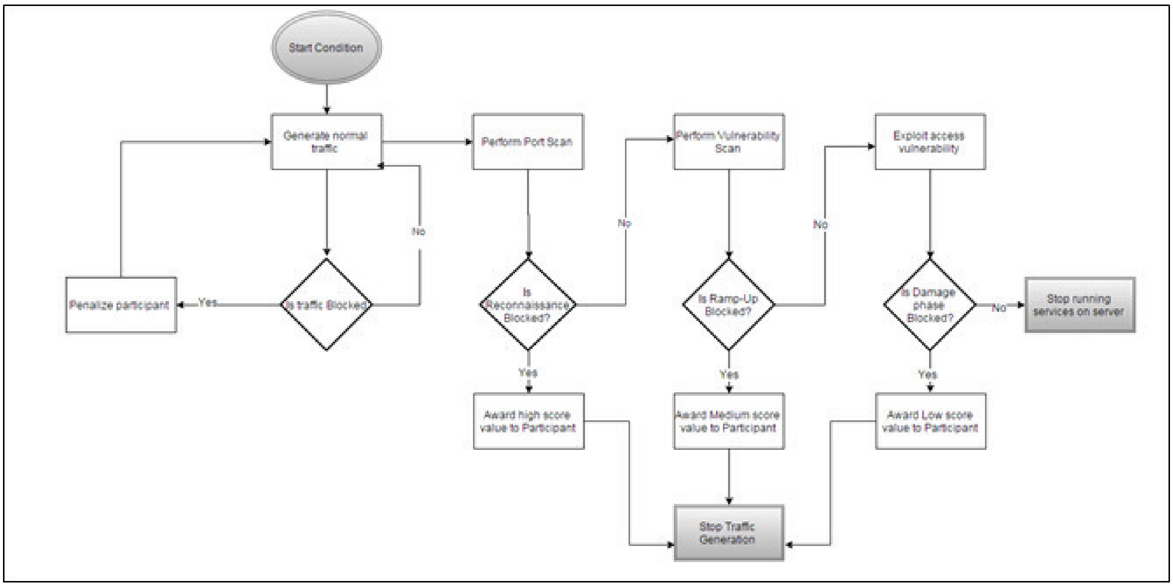
Start Condition
Generate normal traffic
Perform Port Scan
Perform Vulnerability Scan
Exploit access vulnerability
No
No
No
Penalize participant
Yes
Is traffic Blocked
Is Reconnaissance Blocked?
Is Ramp-Up Blocked?
Is Damage phase Blocked?
No
Stop running services on server
Yes
Yes
Yes
Award high score value to Participant
Award Medium score value to Participant
Award Low score value to Participant
Stop Traffic Generation

Figure 6: Logic flow of an unauthorised access attack traffic generator

The difficulty of thwarting an attack before it reaches the damage phase of an attack should increases over time. Initial offensive agents will use fully automated, noisy and basic attacks, as defined by Chan et al.'s ontology (Chan et al. 2015). Over time more sophisticated attack agents will be activated. Before closing off of the CSIR Cyber Challenge event, an attack on all known vulnerabilities will be launched by the challenge organisers to determine how effectively the participants have patched their production servers.

By providing the same level of offensive attacks against all participants a more realistic measurement of the participant's skills can be assessed. In prior Cyber Challenges, a strong offensive capability lead to reduced attacks launched against aggressors. Due to this, aggressors often ended the cyber challenges, having a far weaker defensive capability than other teams. The cyber challenges are meant to be holistic challenges which strive to improve awareness of both offensive, as well as, defensive measures that can be taken during a cyber engagement. By providing a more fair and balanced offensive and defensive scenarios the CSIR challenges aims to provide a more holistic training experience.

5. Conclusion and future work

In this paper, the network attack ontology defined by Van Heerden et al. was used to model the behaviour of automated attacks (Van Heerden et al. 2012). The attack ontology defined the phases of various attack scenarios and mapped then=m to potential sensors to detect the phases of a cyber attack. By using these sensors, as triggers and input for cyber attack and retaliation, one can emulate offensive cyber attack behaviour. This behaviour was scripted as a logic flow engine, to act as both a defensive and offensive agents in a cyber attack environment. To validate the attack model and logical flow of attacks, these agents were placed in a Cyber Challenge environment to interact with security professionals.

Beyond the validation of previous work, the paper also proposes a methodical means by which cyber attack agents can be derived and scripted for automated attacks during cyber challenges. The authors intend to deploy these agents during the upcoming 2016 CSIR Cyber Challenge, to test the hypothesis that having an automated attack mechanism will encourage a balanced challenge environment. Allowing new entrants into the cyber challenge environment to participate without fear of retaliation of attacks, as well as, providing an adequate challenge to proficient participants by emulating advance offensive attacks against production servers. These cyber games will be held in o-operation with other Southern African Development Community (SADC) countries. The research team will present the findings in future work.

References

Boopathi, K, Sreejith, S & Bithin, A 2015, 'Learning Cyber Security Through Gamification', *Indian Journal of Science and Technology* , vol 8, no. 7, pp. 642-649.

Burke, ID & Pieterse, H 2014, 'How to tame your Android Malware', *International Conference on Cyber Warfare and Security* , pp. 54-65.

Chan, P, Theron, J, Van Heerden, R & Leenen, L 2015, 'An Ontology Knowledge Base for Cyber Attack Planning ', *International Conference on Cyber Warfare and Security* , pp. 69-85.

Damele, B & Stampar, M 2012, *SQLmap : Automatic SQL injection and database takeover tool*, viewed 26 October 2015, <sqlmap.org>.

DEF Con 2011, *A history of Capture the Flag at DEF CON*, viewed 5 October 2015, <https://www.defcon.org/html/links/dc-ctf-history.html>.

Fincher, M & Hadnagy, C 2014, *Security through eductaion*, viewed 27 October 2015, <http://www.social-engineer.org/wp-content/uploads/2014/10/SocialEngineerCaptureTheFlag_DEFCON22-2014.pdf>.

Grant, T, Burke, ID & Van Heerden, RP 2012, 'Comparing Models of Offensive Cyber Operations', *Proceedings of the 7th International Warfare and Security*, pp. 108-121.

Khant, A 2013, *OWASP Joomla! Security Scanner*, viewed 26 October 2015, <http://sourceforge.net/projects/joomscan/>.

Lyon, G 2011, *Port Scanning Techniques*, viewed 5 October 2015, <https://nmap.org/book/man-port-scanning-techniques.html>.

Marczewski, A 2013, *Gamification: A Simple Introduction*, 2nd edn, Andrzej Marczewski.

Mehlmauer, C & Van Der Lier, P 2013, *Word Press Scan*, viewed 25 October 2015, <http://wpscan.org/>.

The National Cyber League 2013, *The National Cyber League - Capture the Flag*, viewed 29 September 2015, <http://www.nationalcyberleague.org/games.shtml>.

Van Heerden, RP 2014, *A Formalised Ontology for Network Attack Classification*, viewed 26 October 2015, <http://contentpro.seals.ac.za/iii/cpro/DigitalItemViewPage.external?lang=eng&sp=1011603&sp=T&suite=def>.

Van Heerden, RP, Irwin, B, Burke, ID & Leenen, L 2012, 'A computer network attack taxonomy and ontology', *International Journal of Cyber Warfare and Terrorism*, vol 2, no. 3, pp. 12-25.

Van Heerden, RP, Pieterse, H, Burke, ID & Irwin, B 2013, 'Developing a Virtualised Testbed Environment in Preparation for testing of Network Based Attacks', *International Conference on Adaptive Science and Technology (ICAST)*, pp. 1-8.

Van Heeren, R, Pieterse, H & Irwin, B 2012, 'Mapping the Most Significant Computer Hacking Events', *ICT Critical Infrastructures and Society*, pp. 226-236.

Wireless Village 2014, *Wireless CTF*, viewed 25 October 2015, <http://www.wirelessvillage.ninja/index.html>