Work in Progress: Evaluation of Security Standards through a Cyber Range using Hacker's Tactics, Techniques and Procedures

Abstract—We present a framework for the creation of a cyber range to test the effectiveness of security standards, policies and frameworks. These assets guide organisations on how to protect themselves from cyber threats, they have been created via a variety of methods including standards bodies, anecdotal evidence, findings from successful attacks and others. To date, however, there is not an agreed process for creating cyber ranges to conduct a practical assessment of the recommended controls. As a result, the ability of enterprises and standards bodies to judge the effectiveness of these measures is limited.
Utilising hackers tactics, techniques and procedures to evaluate security standards, should be an effective method for testing a lifelike cyber range which complies to a specific standard. We have started to produce the blueprint for such a laboratory, presented here to showcase our initial findings, using the Cyber Essentials framework as an inceptive use case.

Index Terms—vulnerability management, policies & standards evaluation, cyber ranges

1 Introduction

Organisations and individuals face a serious and increasing digital threat [2]. Cyber incidents, referred to as 'hacking', are becoming more common. [13]. These incidents cause suffering through business disruption and loss of data and the victims have to contend with the subsequent aftermath. This aftermath could be the loss of trust in their clients or a drop in share price for public companies [8] [9]. These organisations can also receive punitive fines if specific types of data were lost, costing up to hundreds of millions of pounds [12].

With this threat taking place every day, the cyber security industry is now worth £10.1 billion in the UK in 2021, rising 77% from £5.7 billion in 2017 [4]. Due to this ongoing threat, the security industry has created multiple standards and frameworks to follow. Whilst each of these provide advice on prevention and response, they may have multiple mitigation strategies, not all of which have empirical evidence of their effectiveness; with the communication and implementation so varied there is substantial room for different and varied levels of effectiveness. [15].

As standards and frameworks are updated at set intervals, which can be years, no peer-reviewed testing has been published to see if the standard keeps pace with the speed in which the malicious actors and scammers are changing their tactics to circumvent the recommended controls. Further to this, there has not been any comprehensive critical research into the evaluation of frameworks comparing them to hackers' Tactics, Techniques and Procedures (TTPs).

Using Cyber Essentials as the inital use case, this research reviews the feasibility of using a digital cyber range to evaluate security standards controls and to provide repeatable results through an exploit matrix. This research provides a practical gap analysis of hackers TTP's and provides an initial comparison to the Cyber Essentials recommended controls. Furthermore, the paper discusses the design decisions around the cyber range and how it will provide rigorous real-world results that can be fed back into the security standards for improvement.

It is the intention that the blueprint could be replicated in a cost-effective way, without the requirement for considerable technical expertise, and specifically designed to be used to test a multitude of different standards, tools and measures in academic and enterprise scenarios. The contributions of this paper will focus on the creation of a cyber range to evaluate security standards, utilising Hackers Tactics, Techniques and procedures for a real-world assessment, using only commodity level attacks.

The contributions of this paper are:

1) Feasibility of using of a cyber range to evaluate security standards.
2) Utilisation of Commodity level Attacks to build upon previous research.
3) Utilisation of Hackers Tactics, Techniques and Procedures for a Real-world assessment of security Standards.
4) Utilisation of Hackers Evasion Techniques to validate standards and framework controls effectiveness.

1.1 Frameworks and Standards

1.2 Approach

This research will investigate how practical assessments, using modern attack strategies in a cyber range environment, can empirically assess security standards. The intent of this research is to present a new approach to provide data and constructive feedback in the use of cyber ranges as a mechanism to evaluate security standards.

This base test case will use the Cyber Essentials framework to research commodity level attacks seen across various industries and investigate how the Cyber Essentials

recommended controls mitigate these attacks. The outcome of the research will see the creation of an exploit matrix to present the data discovered from this research. Whilst this paper explores the base test case, and therefore the effectiveness of Cyber Essentials, the approach will form a standard template that could be used to assess other security standards.

The cyber range will be based on a sand boxed environment where the recommended technical controls from Cyber Essentials framework will be applied during its creation. To ensure the robustness of this research and a more comparable study, we will draw a direct comparison to previous work from cyber security controls effectiveness paper [16], Where we will use similar network topology and technology designs. The hardened network and systems will then be attacked by ethical hackers and penetration testers, using only openly available hacking tools and solutions. If a system or control is compromised, it shows either that the standards have not been implemented correctly or there is a gap and weakness in the standards recommended controls.

2 Implementation

2.1 Cyber Range

A cyber range is a virtualised platform which provides a dedicated testbed allowing for a comprehensive and unbiased assessment. Typically containing multiple types of infrastructure, networks and computers, this enables security testing to be conducted in a real world cyber threat scenario [18] [20].

The literature shows cyber ranges focus has been on education, gaining a better understanding in attacking and defensive skills [20] [11].Whilst there are multiple cyber range designs that are used for cyber exercises worldwide [21] [20], there has been no research in the use of a cyber ranges to review security standards.

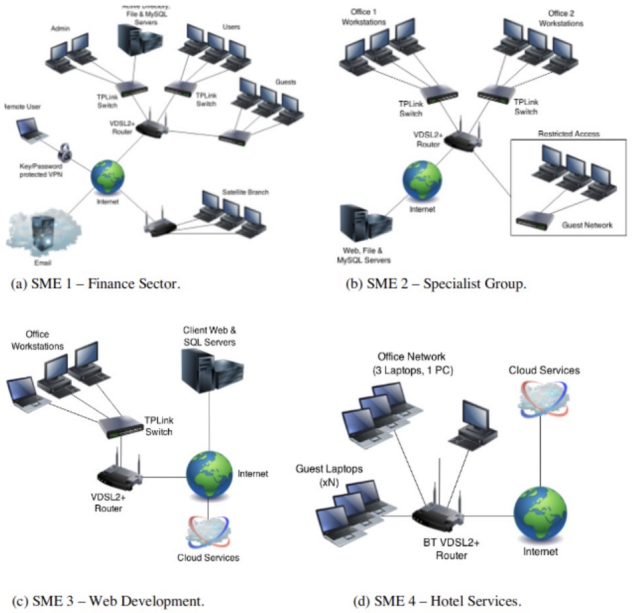
Cyber ranges have matured over time [20] [11], but there are still improvements that can be made [19]. Previous research has shown that cyber ranges are typically expensive, time consuming to construct and difficult to deploy and maintain [7].*Urias et al* wrote that previous cyber ranges used older technologies and suffered from slower networks and from licensing issues. This led to slow deployments which lacked automation [19].

Considering the above opportunity for improvements, as well as the issues with credential management and configuration management [19], this research will look to improve these factors by focusing on repeatability and robustness of the controls in place. In general, a cyber range should be able to provide real-time feedback, with an environment where teams can engage to support the range experiment, where hypotheses may be tested by various teams [19].

The creation of this cyber range will use the base test case to meet the standard requirements outlined above mentioned, where it has been designed to be repeatable to assess any other security standard. To allow for repeatable results with maximum scalability, ease of maintenance and security within a strict budget, the cyber range will be built using public cloud infrastructure. The key requirements were to be able to replicate SME infrastructure, as well as provide a secure unique portal for each hacker to be able to log into on any device without being able to influence any other testers environments. The infrastructure also had to allow for the administrative overhead of keeping the estate compliant to the assessed security standard throughout the testing cycle. With these key considerations Microsoft Azure, Google Cloud Platform, Amazon Web Services, Skytap, SnapLabs (by Immersive Labs) were considered.

Due to the key requirement repeatability and of ease of use from a maintenance and portal access perspective, Skytap was chosen to host the environment as it requires only a browser to access and administer the environment allowing all parties to focus on the critical outcomes of the assessment.

Figure 1 shows a peer reviewed topology design based on surveys of 20 Small to Medium Enterprises [16],



(a) SME 1 – Finance Sector. (b) SME 2 – Specialist Group.

(c) SME 3 – Web Development. (d) SME 4 – Hotel Services.

Fig. 1. Topologies of SME Networks [16]

This research has decided to use the SME1 Finance Sector topology, as it will prove a comparable study to the research conducted from cyber security Controls Effectiveness [16]. Of the four topologies evaluated, SME1 was chosen because it has a more varied technology stack and a more complex design. The network topology will provide more opportunity to investigate network design flaws due to its larger footprint in comparison to the other SME designs from figure 1. Using the same selection criteria for typologies, the design will draw a direct comparison from [16]. However, any technologies identified as end of life will be updated.

Including business functions is critical for a lifelike assessment, their importance to enterprises worldwide is why the protection of these systems make up a large proportion of most security standards. To make the cyber range as lifelike as possible, multiple business processes will be implemented such as File shares, SQL Databases, File Transfer Protocol Server, Security Monitoring, Web Applications and Domain Servers.

To ensure repeatable results for each assessment the cyber range will use preconfigured Lab templates for consistent targets. These machines use run time Scheduling to

VM	Do not run	Stage 1	Delay After stage 1	Stage 2	Delay After stage 2	Stage 3	Delay After stage 3	Final Stage
		All	1 minutes	All	1 minutes	All	0 minutes	All
Firewall	○	○		◉		○		○
Kali Linux - Adversary	○	○		○		○		◉
FILSB001 - File Server	○	○		◉		○		○
FINDT002 - Win10	○	○		○		◉		○
SB Domain Controller	○	◉		○		○		○
ITDT001 - Win10	○	○		○		◉		○
Splunk Monitoring	○	○		◉		○		○
DB001 - WinServ2019 MySQL DB	○	○		○		◉		○
Win10 - Adversary	○	○		○		○		◉
SafeBank - Website	○	○		◉		○		○
Test Server - Adversary	○	○		○		○		◉

Fig. 2. Environment Scheduling

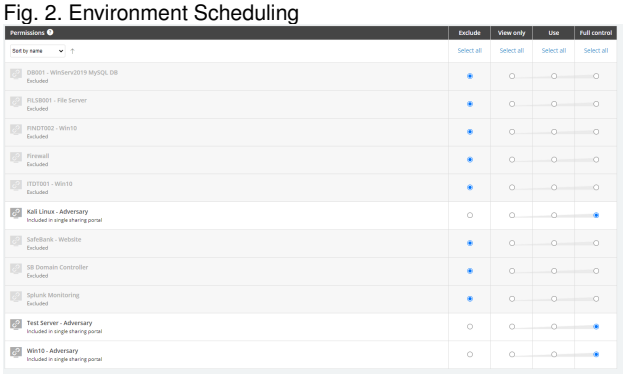


Fig. 3. Sharing Portal

ensure all services and operating systems have been enabled correctly. The environment is compliant to the standard, vulnerability and compliance scans will be conducted using automated vulnerability management and also manually checked for any updates before the assessment begins.

2.2 The Assessment

Once the cyber range has been completed and deemed to be compliant with the relevant standard, Access will be granted to the cyber range via private invitations called sharing portals. This will be practically assessed by penetration testers and ethical hackers. sharing portals are granular in nature and auditable, Penetration testers and ethical hackers will be recruited though private and public internet forums using challenges such as the SynAck Red Team Hacker Hang Out,0x00Sec Forum, and Hack The Box forums.

The assessments will be open to the tester for a period of fourteen days to allow the assessor plenty of time to conduct reconnaissance and enumeration of the target scope, whilst being under the patching window requirements of standards, to create a real life assessment.

The administrators will monitor the activities to ensure the assessor complies with the rules of engagement. To ensure repeatable results the assessors will need to provide the outcome of their tests in the form of a penetration test report. Testers will be encouraged to follow standard methodologies like Open Web Application Security Project (OWASP) Testing Guide or The Penetration Testing Execution Standard (PTES) [6] [17], to ensure a thorough assessment. Only reports which allow the attack to be replicated step-by-step will be accepted. Only commodity attacks may be used and zero day exploits will be out of scope.

The workflow outlines a similar process to the NCR Lifecycle [19] and only differs in the recruitment and communication to the testers:

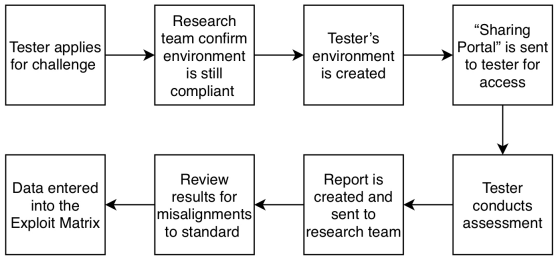


Fig. 4. Assessment Workflow

The testers will provide an accurate assessment of the recommended controls and will form the gap analysis for the Exploit Matrix. The assessments will also provide evidence that the cyber ranges have been implemented correctly, or whether a gap has been discovered in the framework.

2.3 Exploit Matrix

An Exploitation Matrix will be created containing the base test case technical controls and adding any gaps related to a specific control discovered by the ethical hacker. When a potential finding has been discovered on the cyber range, an assessment will then be made into whether the recommended controls mitigated the attack. If a gap has been discovered to a control it will be added to the exploit Column, with the control that was supposed to mitigate it. Further information will be provided to the matrix on whether modern evasion strategies were used to find the flaw.

Some controls are binary, such as default credentials, failure of which would render the lab noncompliant and out of scope. Other controls will be gradient, these will form the basis of the majority of the exploit matrix. Nontechnical and human controls, such as compliance training, will fall out of the scope of this assessment.

The base test case, will use the remediation advice from the penetration testing reports to create a quantitative score. Where any successful attack occurs, it will be assessed whether or not a control exists to mitigate the vulnerability. If there is not a control in place, or the recommended control fails to stop the attack, it will be noted in the exploit matrix as a gap, with an explanation and a proof of concept of the exploit.

The effectiveness of the standard will be attained by using a secure score. This value will be created by collecting the number of gaps in the framework and multiplied by the sum of the Common Vulnerability Scoring System (CVSS v3.1) [3] score.

The CVSS score will be calculated separately using the official CVSS v3.1 online calculator [3], before using the algorithm below. This is due to CVSS v3.1 having environmental factors depending on where the

vulnerability is in the process of discovery/patching and in regards how the impact unravels over time considering the confidentiality, integrity, and availability. Using CVSS provides a way to capture the principal characteristics of a vulnerability and produce a numerical score reflecting its severity. The numerical score can then be translated into a qualitative representation (such as low, medium, high, and critical) to help organizations properly assess and prioritize their vulnerability management processes [5].

Secure Score will be calculated by, Sum of CVSS v3.1 results (x) * amount of findings (n)

$$n.\sum_{i=1}^{n} x_i 0 \leq x_i \leq 10 \tag{1}$$

3 Initial Findings

Whilst creating the base test case some potential gaps were noticed that could provide interesting results within the Exploit Matrix. Hacking methodologies allow us to suggest gaps in the framework warranting further investigation.

Standards and frameworks insist that antivirus software is a requirement. However, they do not consider the various vendors' capabilities. An example of this can be seen when powershell is used to run malicious commands in memory, in an attempt to evade antivirus software through reflective expressions. This is typically detected by *Antimalware Scan Interface* (AMSI), but only approximately 81% of antivirus providers have this functionality enabled [10]. Whilst the adoption rate of this technology is concerning, a greater concern is that AMSI was released in 2015 but the majority of the antivirus providers took between three to six years to implement the interface standard.

Whilst guidance also exists for implementing updates for programming languages such as Java and .Net in the base test case. No secure coding practices or controls have been recommended, this could present a potential attack vector not covered by the standard. Whilst only providing two examples here from a list of sixty three controls, this shows the potential opportunities of using a cyber range to provide empirical evidence in evaluating security standards.

4 Discussion and Challenges

In the creation of the base test case there remains an outstanding discussion on how the cyber range will function and what other benefits could be leveraged from it. Whilst other cyber ranges have used a hybrid approach in their design using physical assets as well as digital, this led to multiple complications [20]. Whilst arguing this could add more realism to the Lab, we believe that the ability to increase automation and repeatability to our results outweighs these factors.

There are also other challenges that will need to be resolved, such as removing all ambiguity on the compliant nature of the cyber range to the standard that it is evaluating. Whilst there are multiple ways to achieve this goal, the most rigorous would be a manual assessment [26] from a 3rd party to validate its compliance, whilst an automated approach would be much faster and increase scalability through the use of vulnerability compliance assessment. Our intention is to use both approaches, Vulnerability and Compliance automated scans as well as utilise the testers results to ensure that any findings that should be mitigated by the standards are resolved before further testing continues.

Whilst one of the key contributions from this research is empirical evidence to substantiate the Security Standards effectiveness, this requires the cyber range to be as 'Real-world ' as possible and mimic a small to medium size enterprise. This comes with its own challenges as we will need a selection criterion which reviews how specific defensive controls such as antivirus and Firewalls are selected. We could implement best of breed security solutions based on Forrester and Gartner's magic quadrant [17], however this could be deemed unrealistic as it would suggest all SMEs will have the budget to purchase these systems. A more realistic approach could be to base the selection on market share research, with the trade off of potentially having limited functionality by using cheaper less mature security tools.

Whilst humans play a major part in cyber security, be it their daily task opening and replying to emails or interacting with productivity apps, the cyber range should look to include human behaviour where possible. This could be emulated via scripts opening of emails and attachments and clicking links to test auto run features as well as email and browser security. However, this would incorrectly suggest that 100% of employees would open suspicious links, but as this research is reviewing the effectiveness of Security Standards by allowing a 100% click rate allows for the ability to test other layered defence strategies, such as disabling auto run and enabling email attachment rules.

Computer hackers are not all equal in skill [12] and as they will form a major part of this assessment, in the evaluation of the controls, it is imperative that a diverse background of participants is found with the correct skills to conduct the assessments. As discussed in the assessment section in 2.2, These individuals will be recruited from various backgrounds and their skill level will be documented with an understanding of their background and also which certifications they might have obtained.

Whilst this paper is using Cyber Essentials as its initial use case, other considerations are being made regarding which standards would be most beneficial to evaluate using cyber ranges. ISO:27001 and NIST CSF have been selected as potential candidates due to their popularity and market share, as well as their different approaches in selecting security controls and framework design.

4.1 Related Work

Previous research suggests standards should be company specific [14], as well as questions why specific controls were chosen, proposing this was due to popular consensus and not empirical evidence. Evaluation tools can provide actionable recommendations to remediate gaps [1], which questions why specific controls were chosen, proposing this was due to popular consensus and not empirical evidence. Evaluation tools can provide actionable recommendations to remediate gaps [1] [14] [15]. With "Cyber Hygiene Does it work?"" [15], evaluating security standards in a similar

way to our research. This investigation reviewed the effectiveness of the Cyber Essentials controls in mitigating 'commodity-level' ("off the shelf") attacks attempting to exploit vulnerabilities in Small and Medium Enterprises (SME) networks. It identified that the resources required to establish and maintain cyber security was high and means that some enterprises were left unprotected.

The research randomly selected two hundred vulnerabilities and tested mitigations across four SME networks, with and without the Cyber Essentials controls in place. A hypothetical network was designed from survey responses was used to assess the typicality of the SME networks, as well as to develop a broader understanding of typical SME network configurations and security practices, showing that without the Cyber Essentials controls, none of the attacks assessed was mitigated on any network. In contrast, compliance with Cyber Essentials mitigated more than 99% of the vulnerabilities and of the exploits only partially mitigated a third relied on hardware or software vendors to release patches succinctly and effectively to combat any vulnerabilities. The investigation showed that a few vulnerabilities not mitigated by Cyber Essentials were due to a fundamental hard-coded flaws in hardware or software that are unable to be updated or patched to a secure state.

Four years on from this paper no studies have been carried out on reviewing the effectiveness of controls in standards from the perspective of an attacker.

5 Conclusion

We have presented a proposal for the creation of a safe, effective cyber range for the testing of security standards using Cyber Essentials an our initial base test case. We have designed a rigorous assessment program and Exploit Matrix presenting any potential gaps, as well as a Secure Score for the evaluation of security standards.

We have presented evidence that Security Standards have been created by a variety of methods including standards bodies policy discussions, anecdotal evidence, findings from successful attacks, market sentiment as well as others. However, they are still used as a mechanism to create a culture of trust between client and supplier, as well as used as a benchmark to show maturity.

Whilst this research has provided evidence that there isn't an agreed process for the testing of standards, or for the creation of a cyber ranges, it has highlighted the usefulness in utilising an offensive Security mindset in finding the potential gaps and the benefits of using a cyber range to evaluate security standards.

References

[1] Benz, M., Chatterjee, D.: Calculated risk? a cybersecurity evaluation tool for smes. Business Horizons **63**(4), 531–540 (2020)

[2] CE, N.: About cyber essentials - ncsc.gov.uk. https://www.ncsc.gov.uk/cyberessentials/overview (November 2021), (Accessed on 31/03/2023)

[3] DATABASE, N.V.: Cvss v3.1 calculator. https://nvd.nist.gov/vuln-metrics/cvss/v3-calculator (may 2022), (Accessed on 31/03/2023)

[4] Department of Digital, Culture, M., Sport: Cyber security sectoral analysis 2022. https://www.gov.uk/government/publications/cyber-security-sectoral-analysis-2022/cyber-security-sectoral-analysis-2022 (Feb 2022), (Accessed on 31/03/2023)

[5] Dugal, D., Rich, D.: Common vulnerability scoring system sig. https://www.first.org/cvss/ (May 2022), (Accessed on 31/03/2023)

[6] Elie Saad, R.M.: Web security testing guide v4.2 (December 2020)

[7] Fujimoto, R., Bock, C., Chen, W., Page, E., Panchal, J.H.: Research challenges in modeling and simulation for engineering complex systems. Springer (2017)

[8] Gordon, Loeb, Sohail: Market value of voluntary disclosures concerning information security. MIS Quarterly **34**(3), 567 (2010). https://doi.org/10.2307/25750692

[9] Hinz, O., Nofer, M., Schiereck, D., Trillig, J.: The influence of data theft on the share prices and systematic risk of consumer electronics companies. Information Management **52**(3), 337–347 (2015). https://doi.org/10.1016/j.im.2014.12.006

[10] Holmes, L., PyroTek3: whoamsi. https://github.com/subat0mik/whoamsi (February 2022), (Accessed on 31/03/2023)

[11] NICE: Cyber ranges (2018)

[12] Office, I.C.: Ico fines british airways £20m for data breach affecting more than 400,000 customers (2022), (Accessed on 31/03/2023)

[13] Rennie, L., Shore, M.: An advanced model of hacking. Security Journal **20**(4), 236–251 (2007)

[14] Siponen, M., Willison, R.: Information security management standards: Problems and solutions. Information & management **46**(5), 267–270 (2009)

[15] Such, J.M., Ciholas, P., Rashid, A., Vidler, J., Seabrook, T.: Basic cyber hygiene: Does it work? Computer **52** (2019)

[16] Such, J.M., Vidler, J., Seabrook, T., Rashid, A.: Cyber security controls effectiveness: a qualitative assessment of cyber essentials (2015)

[17] Team, T.P.: The penetration testing execution standard documentation release 1.1 (April 2022)

[18] Tian, Z., Cui, Y., An, L., Su, S., Yin, X., Yin, L., Cui, X.: A real-time correlation of host-level events in cyber range service for smart campus. IEEE Access **6**, 35355–35364 (2018)

[19] Urias, V.E., Stout, W.M., Van Leeuwen, B., Lin, H.: 2018 international carnahan conference on security technology iccst. Cyber range infrastructure limitations and needs of tomorrow: A position paper pp. 1–5 (2018)

[20] Yamin, M.M., Katt, B., Gkioulos, V.: Cyber ranges and security testbeds: Scenarios, functions, tools and architecture. Computers Security **88**, 101636 (2020)

[21] ÄŒeleda, P., ÄŒegan, J., Vykopal, J., TovarÅˆÃ¡k, D.: Kypoâ€"a platform for cyber defence exercises. MS Support to Operational Tasks Including War Gaming, Logistics, Cyber Defence. NATO Science and Technology Organization (2015)