World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(REVIEW ARTICLE)

Check for updates

A review on the effectiveness of red teaming exercises in modern cybersecurity

Muhammad Eizzat Abdul Razzak * and Mohammad Fadli Zolkipli

School of Computing, College of Arts and Science, Universiti Utara Malaysia (UUM), Sintok, Kedah, Malaysia.

World Journal of Advanced Research and Reviews, 2025, 26(03), 2592-2606

Publication history: Received on 15 May 2025; revised on 23 June 2025; accepted on 25 June 2025

Article DOI: https://doi.org/10.30574/wjarr.2025.26.3.2456

Abstract

Red teaming exercises have become an essential tool in modern cybersecurity, providing a proactive approach to assessing and enhancing defensive capabilities against sophisticated threats. This paper presents a comprehensive review of the effectiveness of red teaming by analysing its core methodologies, tools, and emerging adaptations. A detailed examination is provided regarding adversary simulation and emulation techniques, emphasizing the use of frameworks such as MITRE ATT&CK and Breach and Attack Simulation (BAS) platforms. Physical security assessments also play a significant role, with red teaming techniques continuously evolving to address new architectures like cloud computing, serverless environments, and microservices. Evaluating the effectiveness of red teaming exercises requires a robust framework, incorporating key performance indicators (KPIs) and metrics. However, challenges persist in measurement and attribution, necessitating accurate and reliable methods to truly assess the impact of these exercises. Common challenges are discussed, together with operational risks and ethical challenges. Looking towards the future, the influence of artificial intelligence (AI) and automation on red teaming is analysed, along with the rise of purple teaming and the application of red teaming strategies within Zero Trust architectures. These trends highlight the continuous adaptation required to address the dynamic threat landscape. This review aims to provide valuable insights into the evolving methodologies and future directions of red teaming exercises.

Keywords: Cybersecurity; Red Teaming; Ethical Hacking; Security Metrics; Cyber Resilience

1. Introduction

The contemporary cybersecurity threat landscape is characterized by an unceasing escalation in the volume, sophistication, and diversity of attacks [1] [2]. Malicious actors, ranging from individual hackers and organized criminal syndicates to nation-state sponsored groups, continually develop and deploy novel tactics, techniques, and procedures (TTPs) to compromise organizational assets [3]. Traditional, reactive security measures, which primarily focus on responding to incidents after they occur, are increasingly proving insufficient in the face of such agile and persistent adversaries.

The dynamic nature of modern IT environments - encompassing cloud adoption, the proliferation of Internet of Things (IoT) devices and remote workforces - further expands the attack surface and introduces new complexities for defenders [3]. This environment underscores an urgent need for organizations to shift towards proactive defence strategies. Proactive defence involves anticipating potential threats, identifying vulnerabilities before they can be exploited, and continuously validating the effectiveness of security controls in realistic scenarios. Red teaming exercises have emerged as a cornerstone of such proactive strategies, offering a robust methodology to simulate real-world attacks and assess an organization's true defensive capabilities [4].

The evolution of red teaming itself, from its origins in military strategy to its current sophisticated application in cybersecurity, mirrors the escalating complexity of the threats it seeks to address. As attackers become more adept and

* Corresponding author: Muhammad Eizzat Abdul Razzak

their methods more insidious, the necessity for equally sophisticated and adversarial validation techniques like red teaming becomes paramount, driving its adoption and refinement within the cybersecurity domain.

1.1. Defining Red Teaming in the Context of Modern Cybersecurity

Red teaming, in the context of cybersecurity, is a goal-oriented, adversarial assessment conducted by an authorized and organized group (the *"red team"*) to emulate the TTPs of real-world attackers [3]. The primary objective is to provide a comprehensive evaluation of an organization's security posture by actively attempting to bypass or compromise its defensive controls across people, processes, and technology. Unlike vulnerability assessments or standard penetration tests that might focus on identifying a broad range of potential weaknesses, red teaming is typically objective-driven, aiming to achieve specific goals that a genuine attacker might pursue, such as gaining access to critical data, compromising key systems, or disrupting operations.

The scope of a red team exercise can be broad, encompassing external and internal network penetration, web and application security testing, social engineering, and physical security assessments. The core principle is to simulate the full lifecycle of an attack, from initial reconnaissance and gaining a foothold to post-exploitation activities like lateral movement, privilege escalation, and data exfiltration, all while attempting to evade detection [6].

Originating from military exercises where a *"red team"* would simulate enemy tactics to test defensive strategies, the concept was adapted for cybersecurity to provide a realistic measure of an organization's resilience [5]. The U.S. National Institute of Standards and Technology (NIST) defines a red team as *"A group of people authorized and organized to emulate a potential adversary's attack"* [9]. This definition has evolved beyond purely technical attack simulation. Modern red teaming acknowledges that security effectiveness is a multifaceted interplay of technological controls, human awareness and behaviour, and organizational processes and procedures. Consequently, red teaming has matured into a holistic security capability assessment, reflecting a deeper understanding that attackers exploit not just software flaws but also human fallibility and procedural gaps.

1.2. Rationale for Evaluating Red Teaming Effectiveness

The mere execution of red teaming exercises, while beneficial, is insufficient to guarantee an improved security posture. A systematic and rigorous evaluation of their effectiveness is crucial for several reasons. Firstly, red teaming represents a significant investment in terms of time, resources, and cost. Organizations must be able to quantify the value derived from these exercises to justify continued investment and to ensure that engagements are meeting their intended security objectives. As noted by Mindgard AI, *"measuring success is just as important as running the test itself"* [7].

Secondly, evaluating effectiveness drives continuous improvement - both in the red teaming practices themselves and in the organization's defensive capabilities [8]. By analysing the outcomes, identifying what worked and what did not, and understanding why certain defences failed or succeeded, organizations can refine their security strategies, prioritize remediation efforts, and enhance their overall cyber readiness. This feedback loop is essential for adapting to an ever-evolving threat landscape.

Thirdly, evaluation allows for the benchmarking of an organization's security posture over time [6]. Tracking metrics and KPIs related to detection, response, and remediation following red team exercises can provide tangible evidence of security improvements (or lack thereof) and highlight areas requiring further attention. This ability to demonstrate measurable progress is vital for accountability and for maintaining stakeholder confidence.

The impetus to evaluate red teaming effectiveness is not solely about optimizing security expenditure; it also signals the maturation of red teaming into a strategic business function. When security leaders can demonstrate the tangible benefits of red teaming - such as quantifiable risk reduction, enhanced operational resilience, and validated compliance - they can more effectively communicate cybersecurity's value in terms understood by executive leadership and the board [6]. This elevates cybersecurity from a purely technical cost centre to a core business enabler, contributing to informed decision-making and strategic alignment.

1.3. Scope of the Review

This paper is structured to provide a systemic literature review of the effectiveness of red teaming exercises in modern cybersecurity. Section 2 delves into the diverse methodologies and tools employed, covering adversary simulation and emulation, including the role of the MITRE ATT&CK framework, Breach and Attack Simulation (BAS) platforms, and physical security assessments. It also examines how red teaming is adapting to emerging architectures like cloud, serverless, and microservices. Section 3 explores and addresses the critical aspects of evaluating red teaming

effectiveness, discussing framework, key performance indicators (KPIs) and metrics, and the challenges in measurement and attribution. Section 4 examines the common challenges, operational risks, and ethical considerations inherent in red teaming. Section 5 looks towards the future, analysing the evolving landscape influenced by AI and automation, the rise of purple teaming, and red teaming strategies for Zero Trust architectures. Finally, Section 6 offers concluding remarks on the strategic importance of red teaming in the contemporary cybersecurity paradigm.

2. Methodologies and Tools in Modern Red Teaming Exercises

Red teaming exercises employ a variety of methodologies and tools to simulate adversarial actions and assess an organization's defences. These approaches differ in their objectives, scope, level of automation, and the types of attack vectors they explore [10]. Understanding these distinctions is crucial for selecting the most appropriate red teaming strategy for a given organization and its specific security goals. The primary methodologies analysed in this section are Adversary Simulation and Emulation, Breach and Attack Simulation (BAS), and Physical Security Assessments.

2.1. Adversary Simulation and Emulation

Adversary simulation and emulation represent sophisticated red teaming approaches focused on replicating the behaviour of real-world cyber attackers. While sometimes used interchangeably, a distinction exists: adversary simulation generally involves mimicking the tactics, techniques, and procedures (TTPs) of a particular type of attacker (e.g. a ransomware group, an insider threat), whereas adversary emulation aims to precisely recreate the known TTPs of a specific threat actor or group [11]. Both approaches are critical for assessing an organization's preparedness against relevant threats.

2.1.1. Core Principles, Techniques, and Lifecycle

The foundational principles of adversary simulation and emulation are objective-based, intelligence-driven, and simulate the full attack lifecycle [11]. Engagements typically begin with clearly defined rules of engagement (ROE) and objectives, often centred on compromising specific assets or achieving particular outcomes.

Table 1 Full Attack Lifecycle

Reconnaissance	Gathering information about the target organization, its infrastructure, employees, and potential vulnerabilities using open-source intelligence (OSINT), network scanning, and other information-gathering techniques.
Weaponization and Delivery	Crafting payloads (e.g., custom malware, exploit kits) and choosing methods to deliver them to the target (e.g., phishing emails, exploiting external-facing services).
Exploitation	Leveraging identified vulnerabilities in systems, applications, or human behaviour to gain initial access.
Installation and Command and Control (C2)	Establishing persistence within the compromised environment and setting up covert communication channels with attacker-controlled infrastructure.
Actions on Objectives (Post-Exploitation)	Performing activities to achieve the engagement's goals, such as lateral movement to access other systems, privilege escalation to gain higher levels of access, data collection, and exfiltration.
Reporting	Documenting all activities, findings, successful TTPs, defensive strengths and weaknesses, and providing actionable recommendations for improvement.

Techniques employed span a wide array of attack vectors, including network service exploitation, web application vulnerabilities (e.g. SQL injection, XSS), client-side attacks, social engineering (phishing, vishing, pretexting), and physical intrusion where in scope [6]. The increasing emphasis on specific adversary emulation, rather than generic penetration testing, reflects a maturation in security testing. This intelligence-led approach allows organizations to focus their defensive efforts and resources on preparing for the threats most likely to target them, representing a more strategic and efficient use of security validation resources [3].

2.1.2. Leveraging the MITRE ATT&CK Framework for Realistic Emulation

The MITRE ATT&CK (Adversarial Tactics, Techniques, and Common Knowledge) framework has become an indispensable resource for conducting realistic adversary emulation. ATT&CK is a globally accessible, curated knowledge base of adversary tactics and techniques based on real-world observations. It categorizes attacker actions into tactical objectives, and under each tactic, lists specific techniques for how those objectives are achieved with deeper sub-techniques [12].

Red teams utilize the ATT&CK framework extensively to enhance their operations in multiple ways. By integrating threat intelligence and planning scenarios based on ATT&CK's documentation of known threat groups and their TTPs, red teams can accurately emulate adversary behaviours relevant to the target organization's industry or threat profile. They select and execute specific ATT&CK techniques during engagements to test defensive controls, ensuring comprehensive coverage of attacker's methodologies [6].

ATT&CK provides standardized communication through its common taxonomy, using technique IDs to facilitate clear communication between red teams, defenders, and stakeholders [6]. Additionally, mapping red team activities to ATT&CK techniques helps organizations identify gaps in their detection and prevention capabilities, highlighting specific areas for defensive improvement if techniques are successfully executed without detection or blockage.

The MITRE ATT&CK framework's value extends beyond being a mere catalogue of TTPs; it serves as a vital common operational picture that significantly enhances the effectiveness of *"purple teaming"*, which will be discussed later. By aligning offensive actions with ATT&CK techniques, red teams execute manoeuvres that blue teams are specifically trained to detect and have designed defences against. This common language and understanding streamlines the validation process, making the feedback loop for security improvement more direct and efficient, which is essential for productive collaborations between offensive and defensive teams.

We examined the features of two open-source tools that support adversary simulation and emulation activities based on ATT&CK framework. The selection of tools depends on the engagement objectives, the red team's expertise, and the target environment. We note that there are other commercially available tools used in adversary simulation, among them including C2 frameworks like Cobalt Strike, or Purple Team Exercise Framework (PTEF) with SCYTHE, however they are not covered in this review.

Table 2 Key Adversary Simulation/Emulation Tools

Tool Name	Key Features	Primary Use Case	Strengths	Limitations
MITRE Caldera™	ATT&CK-based automation, adversary profile creation, plugin architecture, post-exploitation support	Automated adversary emulation, continuous testing, research	ATT&CK alignment, extensible, free	Can be complex to set up/customize, UI may be less polished than commercial tools
Atomic Red Team™	Library of discrete ATT&CK technique tests, PowerShell-based execution framework	Detection validation, security control testing, blue team training	Simple, focused tests, easy to use, broad ATT&CK coverage, community-driven	Lacks built-in automation for complex campaigns, requires manual coordination

2.2. Breach and Attack Simulation (BAS)

Breach and Attack Simulation (BAS) has emerged as a significant methodology in modern red teaming, offering automated and continuous validation of an organization's security controls against a wide array of cyberattack TTPs [13]. Unlike traditional red teaming or penetration testing, which are often periodic, resource-intensive exercises, BAS tools provide ongoing, automated stress testing of security controls against the latest adversarial behaviours [14]. These platforms can execute full attack kill chains across various environments, including networks, endpoints, web applications, and email systems, assessing whether an identified exposure can genuinely be exploited within the organization's unique environment.

BAS solutions offer data-driven insights into the effectiveness of implemented security measures by continuously validating defences against a wide array of attack scenarios, such as network infiltration, malware execution,

ransomware downloads, and data exfiltration, and help prioritize remediation efforts based on actual exploitability. This continuous validation capability is particularly valuable in today's rapidly expanding and dynamic attack surfaces [13].

2.2.1. Functionalities, Automation, and Key Platforms

BAS platforms are designed to automatically and continuously simulate attacker behaviours within an organization's environment to test the efficacy of its security defences. Artificial Intelligence (AI) is specifically leveraged to automate and conduct penetration testing [14]. These platforms typically deploy lightweight agents or operate agentless to execute simulated attacks across various vectors, including network, endpoint, email, and web applications. BAS platforms work by mimicking the TTPs of real cybercriminals to test security controls, software vulnerabilities, and baseline weaknesses in a safe, non-destructive manner.

Key functionalities include continuous security control validation, where BAS tools repeatedly test whether security controls such as firewalls, Endpoint Detection and Response (EDR), Web Application Firewalls (WAF), email gateways, Data Loss Prevention (DLP) systems are correctly configured and effective in preventing, detecting, and mitigating simulated threats [15]. They come equipped with extensive libraries of predefined attack scenarios, often mapped to frameworks like MITRE ATT&CK, which cover known malware behaviours, Advanced Persistent Threat (APT) campaigns, and common exploitation techniques. By observing the outcomes of these simulations, BAS platforms can also identify security gaps, misconfigurations, and areas where controls are failing or underperforming [16].

Additionally, many BAS solutions provide prioritized remediation guidance, offering actionable insights and vendor-specific or neutral mitigation recommendations to help security teams efficiently address identified weaknesses. Emerging threat testing is another crucial functionality, with platforms like Picus Security offering rapid updates to their threat libraries, enabling organizations to test their defences against newly emerging threats and vulnerabilities, sometimes within a 24-hour SLA for threats with proof-of-concept exploits [17].

BAS offers a complementary approach by delivering breadth and frequency in testing, while manual exercises typically provide greater depth and creative exploration of vulnerabilities. The growth of BAS platforms addresses a critical need for continuous security validation, which traditional, resource-intensive, point-in-time red teaming engagements often struggle to provide. These platforms emphasize ease of use, broad coverage, and integration capabilities with the existing security ecosystem.

2.2.2. Comparative Analysis of BAS versus Manual Red Teaming

Table 3 BAS versus Manual Red Teaming

Aspect	**BAS**	**Red Teaming**
Automation vs. Human Creativity	Highly automated, relies on predefined scripts and attack scenarios	Driven by human experts, adapt TTPs dynamically, think creatively to bypass defenses, chain exploits in novel ways
Scope and Depth	Broad coverage, continuously testing a wide range of known TTPs across multiple attack vectors	More focused, aiming for specific objectives, achieving greater depth in exploring vulnerabilities and attack paths
Frequency and Duration	Continuous or frequent testing, providing ongoing visibility into security control effectiveness	Periodic engagements, often conducted annually or bi-annually, lasting weeks to months depending on scope
Resource Requirements	Reduces need for extensive human intervention, less resource-intensive once deployed	Requires highly skilled penetration testers and ethical hackers, more costly and time-consuming per engagement
Types of Vulnerabilities Uncovered	Excels at identifying misconfigurations, gaps in security control coverage against known TTPs, failures in detection logic	Better suited for uncovering complex, multi-stage attack paths, zero-day-like vulnerabilities, weaknesses related to human behavior and organizational processes

While both BAS and manual red teaming aim to improve security posture by simulating attacks, they differ significantly in their approach, scope, and objectives. The decision to use BAS, manual red teaming, or a combination often depends on an organization's security maturity. Organizations with less mature security programs may derive significant initial value from BAS by identifying and remediating common misconfigurations and control gaps. As maturity increases and defences become more hardened, the in-depth, creative assessments provided by manual red teaming become more critical for uncovering nuanced weaknesses. Ultimately, many organizations find that a hybrid approach, leveraging BAS for continuous validation and manual red teaming for deep-dive assessments, provides the most comprehensive security assurance.

2.3. Physical Security Assessments in Red Teaming

Physical security assessments are an integral, though sometimes underemphasized, component of comprehensive red teaming exercises. They recognize that a breach of physical security can often serve as a precursor to, or facilitator of, a significant cyberattack [18].

A holistic red teaming strategy must extend beyond the digital realm to encompass physical security assessments [6]. Physical red teaming evaluates an organization's physical security measures by simulating attacks against its physical infrastructure, access controls, and personnel [20]. Techniques employed can include lock picking, attempting to bypass electronic access systems, tailgating, and social engineering to manipulate employees into granting access or divulging sensitive information [19]. Such assessments are often crucial for regulatory compliance and are instrumental in strengthening physical security controls and enhancing employee awareness of physical security threats.

Modern red teaming exercises frequently integrate both physical and cyber-attack scenarios to provide a comprehensive evaluation of an organization's overall defensive posture. Tools used in physical red teaming can range from traditional lock-picking kits to more advanced technologies like drones for surveillance, and RFID cloning devices to duplicate access cards. This expansion of red teaming to include physical security reflects a growing understanding that true organizational resilience requires a multi-layered defense that addresses all potential attack vectors [19].

2.4. Adapting Red Teaming for Emerging Architectures and Threats

As technology landscapes evolve, red teaming methodologies must adapt to address new architectures and the unique threat vectors they introduce. In this section we cover the emergence of cloud environments, serverless architecture, and microservices architectures.

2.4.1. Cloud Environments: Challenges and Best Practices

Red teaming in cloud environments, such as AWS, Azure, and GCP, presents distinct challenges compared to traditional on-premises assessments. A breach into a cloud service does not affect one user, but a multitude of people, underscoring the interconnected and distributed nature of cloud services [22]. These challenges include complex and dynamic architectures where cloud environments are highly distributed, spanning multiple regions and services, with resources being provisioned and de-provisioned dynamically. The shared responsibility model divides security responsibilities between the cloud service provider (CSP) and the customer, requiring red teams to focus on customer-controlled configurations and services [21].

Limited visibility and control often restrict red teams' access to underlying infrastructure logs or network telemetry, especially in Platform as a Service (PaaS) and Software as a Service (SaaS) models. The API-centric attack surfaces of cloud services, which rely heavily on APIs for management and interaction, indicates that APIs are recognized as a key interface for interaction and potential attack vectors [23]. Factoring in the complexity of Identity and Access Management (IAM) roles and policies configuration, credential-based attacks such as SSH brute force attempts, indicate attack vectors executed exploiting the authentication weaknesses, highlighting the importance of robust protocols and password policies [14].

Best practices for cloud red teaming involve understanding cloud-specific threats, focusing on attack vectors described above such as misconfigured cloud storage (e.g. public AWS S3 buckets), IAM privilege escalation, insecure serverless functions, vulnerabilities in container orchestration (e.g. Kubernetes), and exploitation of cloud service APIs. Collaboration with CSPs is crucial, though could be limited, involving understanding the CSP’s terms of service for penetration testing and coordinating where necessary, even though many red team exercises aim for stealth without CSP notification [24]. Leveraging automation is essential, using cloud-specific security assessment tools and scripts with AI capability that's highly adaptable to detect and counter new and unforeseen threats effectively [25]. Simulating realistic cloud attack scenarios by emulating TTPs used by cloud-focused threat actors, such as credential harvesting

from metadata services, lateral movement across cloud accounts, and data exfiltration from cloud storage, is also key [21]. This represents a shift from traditional network perimeter testing to cloud-specific security assessments.

2.4.2. Serverless and Microservices Architectures: Attack Vectors and Testing Approaches

Serverless and microservices architectures introduce unique attack surfaces that necessitate specialized red teaming approaches. Microservices are recognized as a component of modern system architectures in the context of security analysis and penetration testing. These microservices, like other components, have *"interfaces"* with associated *"capabilities"* and potential *"vulnerabilities"* [26].

Serverless architectures, while simplifying infrastructure management, introduce new attack vectors for cybercriminals. Key vulnerabilities include insecure code, event injection attacks, overprivileged roles, and dependency vulnerabilities. Attackers can exploit these to gain unauthorized access, trigger malicious functions, and compromise the entire system [28]. Microservices architectures bring their own set of security concerns, often revolving around insecure inter-service communication and API vulnerabilities, such as authentication, authorization, input validation and rate limiting issues [27].

Red teams adapt their tactics, techniques, and procedures (TTPs) by rigorously testing APIs exposed by serverless functions and microservices for vulnerabilities outlined in the OWASP API Security Top 10. This involves assessing the security of serverless functions and the surrounding infrastructure to identify vulnerabilities. It aims to find weaknesses in the application, APIs, and underlying cloud resources that could be exploited by attackers [29]. With microservices architecture, red team can focus on simulating attacks to identify vulnerabilities in a distributed application architecture. This process helps organizations secure their microservices and prevent malicious actors from exploiting weaknesses. It's crucial because microservices have numerous entry points and APIs, making them potential attack surfaces [30].

Utilizing frameworks like the upcoming OWASP Serverless Top 10 and the established OWASP API Security Top 10, red teams guide their testing efforts with a focus on API security, identity propagation across distributed components, and the security of event-driven triggers and inter-service communication channels. The shift towards serverless and microservices architectures necessitates a red teaming focus on these areas, as traditional network-centric attack paths are either less prevalent or manifest very differently.

3. Evaluating the Effectiveness of Red Teaming Exercises

To maximize the value of red teaming and ensure continuous improvement, organizations must systematically evaluate the effectiveness of these exercises [13]. This involves employing established frameworks, tracking relevant Key Performance Indicators (KPIs) and metrics, and quantifying the business value derived from the findings.

3.1. Frameworks and Models for Evaluation

Several established cybersecurity frameworks and models can be adapted or directly used to evaluate the effectiveness of red teaming exercises and the resulting improvements in an organization's security maturity [31]. The development and adoption of such frameworks signify a broader trend towards standardizing red team operations and their evaluation. This formalization helps transition red teaming from ad-hoc, artful exercises to more scientific, repeatable, and measurable processes, thereby enhancing consistency, comparability, and quality assurance in the field.

One framework that goes in depth in assessing and improving the maturity of a Red Team within an organization is the Red Team Capability Maturity Model (RTCMM). Mirroring the well-known Capability Maturity Model (CMM), it is specifically designed to assess the maturity of an internal red team program across four domains: Processes, Technology, People, and Program. It defines five maturity levels, from occasional and disorganized (Level 1) to continuous and fully effective (Level 5), allowing organizations to benchmark their red team's capabilities and plan for improvements [32].

The RTCMM uses a matrix-based approach to assess various aspects of a Red Team's capabilities, including organization, process, and technology, thus providing roadmap for improvement, and set goals for future development.

Table 4 RTCMM Level Descriptor

Level 1	Occasional, Not Consistent, Not Planned, Disorganized, One-Size-Fits-All, Basic Technical Capability, No OPSEC Considerations
Level 2	Intuitive, Not Documented, Occurs Only When Necessary, Inconsistent Manual Processes, Somewhat Effective Capability, Limited OPSEC Considerations
Level 3	Documented, Predictable, Evaluated Occasionally, Understood, Custom Technical Solutions, Documented Manual Processes, Primary-Use Effectiveness, Best-Practice OPSEC Considerations
Level 4	Well-Managed, Formal, Often Automated, Evaluated Frequently, Majority-Effective Capability
Level 5	Continuous and Effective, Integrated, Proactive, Usually Automated, Easily Customized, Fully Effective Capability, Advanced OPSEC Considerations

3.2. Key Performance Indicators (KPIs) and Metrics

Effective evaluation relies on tracking specific KPIs and metrics that can quantify both the impact of red teaming on the organization's defences and the performance of the red team itself.

3.2.1. Quantifiable Improvements in Security Posture

The table below shows the essential metrics to be captured in each test, to improve ability to handle threats as a result of red team exercises [33].

Table 5 Quantifiable Improvement Metrics

Mean Time to Detect (MTTD)	The average time it takes for the defensive team to discover a simulated malicious activity or TTP initiated by the red team. A reduction in MTTD over successive exercises indicates improved detection capabilities.
Mean Time to Respond (MTTR)	The average time taken by the Blue Team to act or begin assessment after detecting red team's activity. This reflects the efficiency of alert triage and prioritization.
Mean Time to Initial Access (MTTIA)	The time it took for the Red Team to gain initial access to the systems after initiating the attack. This KPI indicates the strength of the security controls e.g. perimeter security or even end user's vulnerability to social engineering attacks.
Mean Time to Act (MTTA)	The duration of time between the moment the business (Blue Team) responded to the Read Team's TTP activity, to the time a solution to address the TTP is implemented. Lower MTTA indicates improvement in the defence posture of the business holistically.
Mean Time to Remediate (MTTR)	The measure of time for full remediation - to contain, eradicate, and recover or accept the vulnerabilities identified and exploited by the red team. This reflects the effectiveness of the containment actions and vulnerability management program. Lower MTTR indicates a more efficient overall incident response.

While a reduction in exploitation success rate in these time-based metrics is generally positive, their interpretation requires careful consideration of the context. For instance, if the red team employs significantly stealthier or more complex TTPs in a later exercise, MTTD might not decrease linearly but should be assessed against this increased challenge. The goal is to demonstrate improved resilience against relevant and evolving threats.

3.2.2. Assessing Red Team Performance

Evaluating the red team itself is crucial for ensuring the quality and relevance of the exercises. The objective achievement rate measures the extent to which the red team successfully achieved its predefined goals, such as accessing specific data, compromising designated critical systems, or disrupting a particular business process.

Another important metric is the mean time to initial access (MTTIA), which calculates the average time it took the red team to gain their first unauthorized foothold in the target environment. A very short MTTIA against hardened targets might indicate exceptional red team skill, or significant perimeter weaknesses [33].

TTP coverage assesses the breadth and depth of the techniques (MITRE ATT&CK or other relevant frameworks) simulated during the engagement. This can be measured by mapping executed TTPs against the ATT&CK matrix and comparing them to the planned scope [33]. Stealth and detection avoidance is another critical metric, evaluating how long the red team operated before being detected, the number of critical actions performed without triggering alerts, or the ratio of detected versus undetected activities. This measures the red team's sophistication and ability to emulate advanced adversaries [34].

Efficiency considers the time and resources, including personnel and tools, consumed by the red team to achieve their objectives. Although this can be a more subjective measure, it is important for resource allocation. The vulnerability discovery rate and criticality metric measure the number and severity of new, previously unknown vulnerabilities or unique attack paths uncovered by the red team.

There may be more metrics that can be used to evaluate red team performance, however it is worth noting that solely relying on whether Red Team *"got in"* is an oversimplification. A mature assessment considers how they achieved their objectives, including the TTPs employed, the stealth maintained, and the efficiency of their operations. This holistic view ensures that the red team is genuinely testing defences against plausible, sophisticated adversaries rather than merely exploiting easily found, low-hanging fruit.

3.2.3. Qualitative Assessment of Value and Impact

Beyond quantitative metrics, the qualitative value derived from red teaming can be factored in to maximise Red Team's engagement. While Red teaming should produce quality thinking and advice, these difficult-to-measure qualitative benefits, such as fostering a security-aware culture or improving collaborative defence, are vital for achieving long-term improvements in security posture and resilience [33] [35].

Table 6 Qualitative Improvement Metrics

Strategic Insights Gained	The exercise's contribution to a better understanding of the organization's true risk posture, the effectiveness of its security strategy, and areas for strategic investment.
Improvements in Security Awareness and Culture	Observable changes in employee behaviour regarding security practices, increased vigilance, and a more proactive security mindset across the organization.
Enhanced Inter-Team Collaboration	Improved communication and cooperation between security teams (e.g. SOC, IR, vulnerability management) and potentially IT and business units, particularly if purple teaming elements are incorporated.
Increased Stakeholder Confidence	Greater assurance for leadership and the board that security investments are being validated and that the organization is prepared to handle cyber threats.

3.3. Challenges in Measuring Effectiveness and Attributing Improvements

Despite the benefits, accurately measuring the effectiveness of red teaming and attributing specific security improvements solely to these exercises presents several challenges. One of the primary difficulties is the isolation of impact, as organizations often implement multiple security initiatives concurrently [36]. Isolating the precise contribution of red teaming to an improved metric, such as reduced Mean Time to Detect (MTTD), from the effects of new security tools, updated policies, or general staff training can be challenging.

Obtaining consistent metrics poses another issue. Qualitative improvements, like enhanced security culture or improved inter-team collaboration, are inherently difficult to quantify consistently. Even for quantitative metrics, variations in red team scope, objectives, and the sophistication of simulated attacks across different engagements can complicate direct comparisons [33].

Additionally, there's the *"proving a negative"* problem. A core value of red teaming is preventing breaches, but it is inherently difficult to definitively prove that a breach didn't happen because of improvements made due to a red team exercise. This issue is particularly pronounced in AI red teaming, where defining what *"good"* or *"safe"* means and measuring these variables can be highly subjective and context-dependent [36]. Red teaming can demonstrate a weakness exists, but it cannot guarantee that others do not.

Both the external threat landscape and the organization's internal environment, where systems, personnel, processes are constantly evolving. This moving target further complicates measurement. Security posture is not static, and improvements observed after a red team exercise might be influenced by ongoing changes. This necessitates a holistic, trend-based view of security posture improvement over time, rather than relying on single-point-in-time measurements.

Finally, the resource intensity of measurement is a notable challenge. red team to constantly update their knowledge of attacker TTPs, develop or acquire sophisticated tools, and adapt to increasingly complex IT environments. Thoroughly tracking metrics, conducting post-exercise analysis, and attempting to quantify ROI can itself be a resource-intensive process.

Addressing these challenges requires careful planning of evaluation strategies, establishing clear baselines before red team engagements, meticulously tracking remediations linked to red team findings, and employing a combination of quantitative and qualitative measures over the long term. By adopting these methods, organizations can better assess the true impact of red teaming on their security posture.

4. Challenges and Risks in Red Teaming Exercise

Red teaming exercises, while crucial for strengthening cybersecurity, are subject to various challenges, operational risks, and ethical considerations. These factors can significantly affect the effectiveness and overall success of these simulated adversarial attacks.

Among the pressing technical challenges include significant investment in time, skill, resources, and human capital [40]. It is often a manual process, meaning its quality depends heavily on the expertise, time, and dedication of the red team members [2]. Automated post-exploitation tools only test a subset of foundational cyber actions, and manually automating initial access can take hours of research. This is on top of scarcity of experienced red teaming services and personnel who can manage the complexity of advanced red teaming frameworks [40].

With ambiguity in red teaming scope and definition, there's a chance of conflict on the precise structure and assessment criteria for red teaming, which can lead to vague goals and inconsistent practices [5]. This lack of standardization extends to reporting, with no unified protocols for disclosing findings or resource costs, making it difficult to gauge true effectiveness. A cost-effective in-house red teaming however can suffer from *"familiarity blind spots,"* making objective performance difficult to maintain [35].

Beyond the challenges in execution, red teaming carries specific operational risks. Conducting exercises can pose a risk of disrupting day-to-day operations. Operational disruption can occur if red team activities inadvertently cause disruptions to business operations, affect system stability, or corrupt data. This could lead to an increased susceptibility to a real attack, if thorough post-exploitation analysis by red teams is not done, leaving organizations more vulnerable to actual offensive cyberspace operations [40]. Relying on conventional red teaming methodologies that don't fully replicate real threats can also lead to a false sense of security within an organization [6]. Thus, it is important to have a mitigation strategy before the exercise, include agreeing on a defined methodology to handle critical systems with care, establishing clear communication channels with regular checkpoints between the client and testers, and having emergency stop procedures.

The nature of simulating adversarial behaviour introduces several ethical considerations that must be carefully managed. The fundamental ethical dilemma in red teaming involves the question of trust between the organization and the ethical hackers. It is crucial to define clear rules of engagement, including what tools can be used, the scope and depth of testing, and how far testers can go, all agreed upon by the organization and the red team [22]. Consent for testing, especially for human elements like phishing simulations, is vital, noting that unexpected risks to other parties may arise.

Red teaming, especially when involving social engineering or physical intrusion, navigates a delicate ethical landscape. Red teams have an ethical duty to report identified vulnerabilities to the affected parties, however, challenges exist in responsible disclosure, particularly concerning sensitive findings that could inspire real attackers [22]. Thus, ethical hackers in the red team must operate within legal boundaries, adhere to agreed-upon scopes, and above all, *"do no harm"*. Employee privacy must be protected, ensuring activities do not unnecessarily infringe on individual rights and comply with data protection laws. Data should be minimized, anonymized, and handled with care. Psychological impact is another consideration; employees targeted by social engineering or witnessing simulated breaches may experience

stress or anxiety [22]. Exercises should be designed to minimize psychological harm, avoiding aggressive tactics and ensuring ethical justifiability.

5. The Evolving Cybersecurity Landscape and Future of Red Teaming Exercise

The field of red teaming is continuously evolving, driven by advancements in technology, changes in adversarial TTPs, and a deeper understanding of how to maximize the value of these exercises. Consequently, the red team will have to adopt the way to these trends shaping its future, to ensure its effectiveness of the exercise. include the integration of Artificial Intelligence (AI) and automation, the increasing sophistication of APT simulations, the rise of collaborative purple teaming, and the adaptation of strategies for Zero Trust architectures.

5.1. The Role of Artificial Intelligence (AI) and Automation

Artificial Intelligence (AI) and automation are poised to play a transformative role in the future of red teaming, enhancing red team capabilities and enabling simulations of AI-augmented adversaries. AI and machine learning (ML) are increasingly being incorporated into red teaming tools to augment human capabilities and improve efficiency. AI can automate and conduct penetration testing, a task traditionally time-intensive, high-cost, and requiring expert cybersecurity professionals. Automated penetration testing is increasingly important due to the difficulty in finding suitable cybersecurity professionals, thereby lessens the burden of manual testing, and enhancing the efficiency and scalability of penetration testing [14].

Intelligent vulnerability discovery using ML models can identify patterns indicative of code vulnerabilities or misconfigurations in complex systems, providing deeper insights into security gaps [37]. AI also revolutionizes the generation of sophisticated phishing campaigns, particularly through Large Language Models (LLMs), which can create highly convincing and personalized phishing emails or social engineering pretexts at scale [38]. This enhances red teams' ability to test the human element of security effectively.

The use of AI can create realistic cyber threat simulations by adapting attack strategies based on target system responses, making the behaviour more sophisticated and challenging [14]. This is particularly useful for replicating real-world scenarios and advanced persistent threats (APTs) that are designed to be stealthy [23].

In addition to these applications, AI can optimize attack paths by analysing network topologies and vulnerabilities to identify the most effective routes to achieve specific objectives, streamlining complex attack planning. Adaptive malware simulation is another promising area, where AI can create simulated malware that adapts its behaviour in response to defensive measures, mimicking advanced evasive threats and testing the robustness of security defences. Automated reporting is yet another domain where AI shines, assisting in collating findings, mapping them to frameworks like ATT&CK, and drafting initial report sections, saving time, and ensuring comprehensive documentation.

As AI technology continues to advance, its integration into red teaming is expected to shift the focus of human red teamers from repetitive, time-consuming tasks to more strategic endeavours. AI-driven red teaming enables adaptive learning and continual refinement of defence strategies by identifying weaknesses in defence mechanisms. The iterative process between blue and red teams helps build more robust and secure network environments [8]. Overall, AI and automation significantly enhance the effectiveness and impact of red teaming exercises, preparing organizations to defend against the next generation of cyber threats.

5.2. The Rise of Purple Teaming: Fostering Collaboration Between Offensive and Defensive Teams

Purple teaming represents a significant evolution in the adversarial testing paradigm, emphasizing collaboration and knowledge sharing between the red team (offensive) and the blue team (defensive). Unlike traditional approaches where teams operate in silos with a *"win/lose"* dynamic, purple team exercises foster a cooperative environment focused on collective improvement and enhanced security posture. Collaborative nature of red-teaming exercises promotes interdepartmental communication and coordination, which are critical for effective cybersecurity management [6].

Open communication is a hallmark of purple teaming, with red teams often sharing their tactics, techniques, and procedures (TTPs) as well as attack plans with the blue team, either before or during the execution of specific attack steps [33]. This transparency helps the blue team understand and anticipate potential threats.

Real-time feedback is another crucial component, where blue teams can observe red team actions and test their detection and response capabilities on the spot. This immediate interaction allows teams to identify gaps and areas for

improvement in real-time, enhancing the effectiveness of defensive measures. Joint analysis and remediation involve both teams working together to analyse the results of the exercises, that can be addressed in remediation plan [39]. They collaborate to understand why certain detections failed or succeeded and develop and implement remediation strategies collectively. This joint effort ensures that both teams contribute their expertise to the improvement process.

A structured red team exercises can explicitly facilitate purple teaming by including planned pauses for debriefing sessions or allowing the red team to trade stealth for efficiency, testing a wider range of blue team responses to specific TTPs. This collaborative model maximizes the value of red team findings by ensuring immediate knowledge transfer and fostering a more integrated, agile, and continuously learning security organization.

5.3. Red Teaming Strategies for Zero Trust Architectures

As organizations increasingly adopt Zero Trust architectures, which operate on the principle of *"never trust, always verify"* and designed to eliminate implicit trust based on physical or network location, red teaming methodologies must adapt to assess the effectiveness of these new security models [41]. While ZTA is a significant security measure, is not foolproof, and motivated adversaries will continue to find creative ways to conduct post-exploitation actions despite its implementation [40]. Therefore, among the many red teaming strategies that needs to be focused on against a ZTA, the area of focus would include assuming an initial breach to focus on post-exploitation actions, and challenges in continuous verification in ZTA's dynamic policies.

A critical aspect is validating micro-segmentation, a core tenet of Zero Trust that divides the network into small, isolated zones to limit lateral movement. Red teams should operate with the assumption that an initial perimeter breach has occurred, or that an insider threat is present within the network [40]. Red teams attempt to breach these microsegments, move between them, and test the effectiveness of granular network access controls and policies. This helps in understanding how well the network segmentation strategy prevents unauthorized lateral movement and protects critical assets.

The other key aspects of red teaming on ZTA is addressing ZTA's heavy reliance on continuous authentication, authorization, and real-time assessment of user behaviour and environmental conditions to detect anomalies and respond swiftly. Red teams focus on compromising identities, bypassing multi-factor authentication (MFA), exploiting vulnerabilities in identity providers (IdPs), and attempting privilege escalation through compromised accounts [21]. This approach ensures that the identity and access management systems are thoroughly evaluated for their strength and resilience.

Exploiting misconfigurations in Zero Trust enforcement points could also be part of the exercise. Red teams test the security of policy enforcement points, including Zero Trust Network Access (ZTNA) gateways, API gateways, and data access controls, for misconfigurations or vulnerabilities that could be bypassed [42]. Identifying and addressing these weaknesses is crucial to maintaining the integrity of the Zero Trust model.

Red teaming in a Zero Trust environment shifts the focus from breaching a traditional hardened perimeter to testing the granularity and efficacy of distributed micro-perimeters, the resilience of identity and access management systems under attack, and the organization's ability to detect and respond to lateral movement attempts within supposedly isolated segments. This comprehensive approach ensures that the Zero Trust architecture's foundational principles are rigorously tested and validated, enhancing the organization's security posture and its ability to withstand sophisticated attacks.

6. Conclusion

The red teaming exercises, when thoughtfully planned, expertly executed, and rigorously evaluated, are an indispensable component of a mature cybersecurity strategy. They provide unparalleled insights into an organization's true defensive capabilities, drive meaningful improvements in security posture, and ultimately enhance resilience against the persistent and evolving cyber threats of the modern era. Their effectiveness lies not just in finding flaws, but in fostering a proactive, adaptive, and robust security culture prepared to meet future challenges.

Compliance with ethical standards

Acknowledgments

The authors would like to thank all members of the School of Computing who are involved in this study. This study was carried out as part of the Hacking and Penetration Testing Project. This work was supported by Universiti Utara Malaysia. Special appreciation is extended to academic mentors and colleagues for their valuable feedback and guidance throughout the research process. The author is also thankful for the availability of open-access research and documentation, which played a critical role in shaping the analysis presented in this paper

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Gnanasekaran V, Bartnes M, Grotan TO, Heegaard PE. Cyber-incident Response in Industrial Control Systems: Practices and Challenges in the Petroleum Industry. In: Proceedings of the 2024 ACM/IEEE 4th International Workshop on Engineering and Cybersecurity of Critical Systems (EnCyCriS) and 2024 IEEE/ACM Second International Workshop on Software Vulnerability [Internet]. New York, NY, USA: ACM; 2024 [cited 2025 May 29]. p. 53–60. Available from: https://doi.org/10.1145/3643662.3643958

[2] Enoch SY, Huang Z, Moon CY, Lee D, Ahn MK, Kim DS. HARMer: Cyber-Attacks Automation and Evaluation. IEEE Access. 2020;8:129397–414.

[3] Ajmal AB, Khan S, Alam M, Mehbodniya A, Webber J, Waheed A. Toward Effective Evaluation of Cyber Defense: Threat Based Adversary Emulation Approach. IEEE Access. 2023;11:70443–58.

[4] Enoch SY, Huang Z, Moon CY, Lee D, Ahn MK, Kim DS. HARMer: Cyber-Attacks Automation and Evaluation. IEEE Access. 2020;8:129397–414.

[5] Feffer M, Sinha A, Deng WH, Lipton ZC, Heidari H. Red-Teaming for Generative AI: Silver Bullet or Security Theater? Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society. 2024 Oct 16;7:421–37.

[6] Yulianto S, Soewito B, Gaol FL, Kurniawan A. Enhancing cybersecurity resilience through advanced red-teaming exercises and MITRE ATT&CK framework integration: A paradigm shift in cybersecurity assessment. Cyber Security and Applications. 2025 Dec;3:100077.

[7] Glynn F. How To Measure the Effectiveness of a Red Teaming Assessment [Internet]. Mindgard. 2025 [cited 2025 May 30]. Available from: https://mindgard.ai/blog/how-to-measure-a-red-teaming-assessment

[8] Wang C, Redino C, Clark R, Rahman A, Aguinaga S, Murli S, et al. Leveraging Reinforcement Learning in Red Teaming for Advanced Ransomware Attack Simulations. In: 2024 IEEE International Conference on Cyber Security and Resilience (CSR) [Internet]. IEEE; 2024 [cited 2025 May 30]. p. 262–9. Available from: https://doi.org/10.1109/csr61664.2024.10679510

[9] NIST. CSRC. 2025 [cited 2025 May 30]. Red Team/Blue Team Approach. Available from: https://csrc.nist.gov/glossary/term/red_team_blue_team_approach

[10] Feffer M, Sinha A, Deng WH, Lipton ZC, Heidari H. Red-Teaming for Generative AI: Silver Bullet or Security Theater? Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society. 2024 Oct 16;7:421–37.

[11] Ajmal AB, Shah MA, Maple C, Asghar MN, Islam SU. Offensive Security: Towards Proactive Threat Hunting via Adversary Emulation. IEEE Access. 2021;9:126023–33.

[12] The MITRE Corporation. MITRE ATT&CK® Knowledge Base. [cited 2025 May 30]. MITRE ATT&CK®. Available from: https://attack.mitre.org/

[13] Ajmal AB, Khan S, Alam M, Mehbodniya A, Webber J, Waheed A. Toward Effective Evaluation of Cyber Defense: Threat Based Adversary Emulation Approach. IEEE Access. 2023;11:70443–58.

[14] Karagiannis S, Fusco C, Agathos L, Mallouli W, Casola V, Ntantogian C, et al. AI-Powered Penetration Testing using Shennina: From Simulation to Validation. In: Proceedings of the 19th International Conference on Availability, Reliability and Security [Internet]. New York, NY, USA: ACM; 2024 [cited 2025 May 30]. p. 1–7. Available from: https://doi.org/10.1145/3664476.3670452

[15] SentinelOne. SentinelOne. 2024 [cited 2025 Jun 12]. What is Breach and Attack Simulation (BAS)? Available from: https://www.sentinelone.com/cybersecurity-101/cybersecurity/breach-and-attack-simulation-bas/

[16] Glynn F. Breach and Attack Simulation (BAS) vs. Red Teaming: What's the Difference? [Internet]. Mindgard. 2025 [cited 2025 Jun 12]. Available from: https://mindgard.ai/blog/breach-and-attack-simulation-vs-red-teaming

[17] Picus Security. What Is Breach and Attack Simulation (BAS)? [Internet]. Picus Security. 2025 [cited 2025 Jun 12]. Available from: https://www.picussecurity.com/resource/glossary/what-is-breach-and-attack-simulation

[18] Almaarif A, Lubis M. Vulnerability Assessment and Penetration Testing (VAPT) Framework: Case Study of Government's Website. International Journal on Advanced Science, Engineering and Information Technology. 2020 Oct 15;10(5):1874–80.

[19] Yulianto S, Soewito B, Gaol FL, Kurniawan A. The Crucial Role of Red Teaming: Strengthening Indonesia's Cyber Defenses Through Cybersecurity Drill Tests. International Journal of Safety and Security Engineering. 2024 Aug 30;14(4):1231–42.

[20] Reconnaishawnce. GitHub. [cited 2025 Jun 12]. PhySec Red Teaming - Tools and Resources for Physical Security Red Teaming. Available from: https://github.com/Reconnaishawnce/Red-Team

[21] Nguyen H. Savvycom. 2025 [cited 2025 Jun 12]. Red Teaming in the Cloud: Challenges and Best Practices. Available from: https://savvycomsoftware.com/blog/red-teaming-in-the-cloud-challenges-and-best-practices/

[22] Pawlicka A, Pawlicki M, Kozik R, Choraś M. What Will the Future of Cybersecurity Bring Us, and Will It Be Ethical? The Hunt for the Black Swans of Cybersecurity Ethics. IEEE Access. 2023;11:58796–807.

[23] Walter MJ, Barrett A, Tam K. A Red Teaming Framework for Securing AI in Maritime Autonomous Systems. Applied Artificial Intelligence. 2024 Sep 4;38(1).

[24] FedRAMP. Penetration Test Guidance [Internet]. Federal Risk and Authorization Management Program; 2022. p. 1–3. Available from: https://www.fedramp.gov/assets/resources/documents/CSP_Penetration_Test_Guidance.pdf

[25] Narula S, Ghasemigol M, Carnerero-Cano J, Minnich A, Lupu E, Takabi D. Exploring Research and Tools in AI Security: A Systematic Mapping Study. IEEE Access. 2025;13:84057–80.

[26] Skandylas C, Asplund M. Automated penetration testing: Formalization and realization. Computers and Security. 2025 Aug;155:104454.

[27] Shaw B. CrowdStrike. [cited 2025 Jun 12]. What is Microservices Security? Available from: https://www.crowdstrike.com/en-us/cybersecurity-101/cloud-security/microservices-security/

[28] Palo Alto Networks. Palo Alto Networks. [cited 2025 Jun 12]. What Is Serverless Security? Available from: https://www.paloaltonetworks.com/cyberpedia/what-is-serverless-security

[29] Bipin Gajbhiye, Shalu Jain, Pandi Kirupa Gopalakrishna Pandian. Penetration Testing Methodologies for Serverless Cloud Architectures. Innovative Research Thoughts. 2022 Dec 30;8(4):347–59.

[30] Oyekunle Claudius Oyeniran, Adebunmi Okechukwu Adewusi, Adams Gbolahan Adeleke, Lucy Anthony Akwawa, Chidimma Francisca Azubuko. Microservices architecture in cloud-native applications: Design patterns and scalability. Computer Science and IT Research Journal. 2024 Sep 6;5(9):2107–24.

[31] McCammon K. KWM. 2024 [cited 2025 Jun 13]. Open Source Roundup of Cybersecurity Models. Available from: https://kwm.me/posts/cybersecurity-models/

[32] Harrell B, Stroup G. RTCMM. 2022 [cited 2025 Jun 13]. Red Team Capability Maturity Model. Available from: https://www.redteammaturity.com/

[33] Risk Crew. Guide to Essential RED Team KPIs and Metrics [Internet]. Red Team. 2024. Available from: https://www.riskcrew.com/wp-content/uploads/2023/05/Red-Team-KPIs-Metrics-Guide.pdf

[34] Shaffer A, Hembree D, Singh G. Obfuscation, Stealth, and Non-Attribution in Automated Red Team Tools. International Conference on Cyber Warfare and Security. 2025 Mar 24;20(1):132–41.

[35] Fortra. A Simple Guide to Successful Red Teaming [Internet]. Cobalt Strike. 2024. Available from: https://www.cobaltstrike.com/resources/guides/a-simple-guide-to-successful-red-teaming

[36] Ji J. How to Improve AI Red-Teaming: Challenges and Recommendations [Internet]. Center for Security and Emerging Technology. 2025 [cited 2025 Jun 13]. Available from: https://cset.georgetown.edu/article/how-to-improve-ai-red-teaming-challenges-and-recommendations/

[37] Al-Azzawi M, Doan D, Sipola T, Hautamäki J, Kokkonen T. Artificial Intelligence Cyberattacks in Red Teaming: A Scoping Review. In: Lecture Notes in Networks and Systems [Internet]. Cham: Springer Nature Switzerland; 2024 [cited 2025 Jun 13]. p. 129–38. Available from: https://doi.org/10.1007/978-3-031-60215-3_13

[38] Hazell J. arXiv.org. 2023. Spear Phishing With Large Language Models. Available from: https://arxiv.org/abs/2305.06972

[39] European Central Bank. Purple Teaming Guidance [Internet]. TIBER-EU. EUROSYSTEM; 2025. Available from: https://www.ecb.europa.eu/pub/pdf/annex/ecb.tiber_eu_purple_best_practices_2025.en.pdf

[40] Benito R, Shaffer A, Singh G. An Automated Post-Exploitation Model for Cyber Red Teaming. International Conference on Cyber Warfare and Security. 2023 Feb 28;18(1):25–34.

[41] Ogendi EG. Leveraging Advanced Cybersecurity Analytics to Reinforce Zero-Trust Architectures within Adaptive Security Frameworks. International Journal of Research Publication and Reviews. 2025 Feb;6(2):691–704.

[42] Exploring Effective Zero Trust Architecture for Defense Cybersecurity: A Study. KSII Transactions on Internet and Information Systems. 2024 Sep 30;18(9).