Cybersecurity Resilience

Prepared by Edas Ramanauskas and Kobi Chambers

On behalf of SecureAI Labs and Dr. Zahra Jadidi

Griffith University

Definition

Cybersecurity resilience is the capacity of an organisation to prepare for, withstand, respond to, recover from, and adapt to adverse cyber incidents, such as attacks or breaches. It involves a holistic approach, integrating robust security measures to protect against potential threats, actively monitoring systems for signs of intrusion, and having a well-defined incident response plan. The goal is to ensure that critical systems, data, and services can be restored to normal operation through reliable data backup and disaster recovery processes. Furthermore, cybersecurity resilience requires continuous adaptation and improvement of security practices in response to evolving threats. It necessitates strong coordination across departments and external partners, fostering information sharing and resource allocation to enhance security.

Phases of Cybersecurity Resilience

1. Preparation and Planning

- Risk Assessment: Identifying critical assets, potential threats, vulnerabilities, and assessing impacts.
- Policy and Framework Development: Creating security policies and ensuring compliance with laws like GDPR, HIPAA, and industry standards.
- Resource Allocation: Allocating funds and human resources to cybersecurity efforts, including acquiring necessary technologies like SIEM systems and training staff.
- Training and Awareness: Educating employees about phishing and other cybersecurity threats, promoting a culture of security within the organisation.

2. Prevention and Protection

- Security Controls Implementation: Role-based access control (RBAC), encryption, and endpoint protection measures to safeguard critical assets.

- Vulnerability Management: Ensuring timely updates through patch management and configuration management.
- Network Security: Implementing firewalls, intrusion detection systems (IDS), and network segmentation.

3. Detection and Monitoring

- Continuous Monitoring: Employing SIEM systems, behaviour analytics, and traffic analysis to detect potential threats.
- Incident Detection: Utilising automated alerts and threat intelligence feeds to detect malicious activity.
- Threat Intelligence: Sharing threat intelligence with industry peers to enhance collective defence mechanisms.

4. Incident Response

- Incident Classification: Defining the severity of incidents and communicating protocols.
- Containment and Eradication: Isolating affected systems, conducting forensic analysis, and eliminating root causes.
- Communication and Documentation: Documenting incidents and informing stakeholders about resolution and impact.

5. Recovery and Restoration

- System and Data Recovery: Implementing business continuity and data restoration processes to recover from cyber incidents.
- Post-Incident Analysis: Conducting root cause analyses and improving processes to prevent future incidents.

6. Adaptation and Evolution

- Continuous Improvement: Regularly assessing security posture and implementing metrics to gauge resilience effectiveness.
- Technology Adoption: Leveraging emerging technologies like AI, machine learning, and automation to enhance cybersecurity capabilities.

Accuracy of Cybersecurity Attack Detection

In defending network infrastructures, accuracy in detecting attack vectors plays a vital role in resilience:

1. Distributed Denial of Service (DDoS): Current detection systems based on traffic pattern analysis achieve 80-90% accuracy [2], though advanced DDoS techniques can evade detection.
2. Man-in-the-Middle (MitM) Attacks: With mutual authentication and encryption, detection rates range from 70-85%. However, sophisticated methods like rogue Wi-Fi access points remain challenging to detect.
3. Phishing and Spear Phishing: Detection rates vary between 85-95%, depending on the sophistication of the attack. Personalised spear phishing often reduces detection accuracy.
4. SQL Injection (SQLi): Web application firewalls (WAFs) typically offer 90-95% accuracy [1]. However, complex patterns or zero-day exploits can evade detection.
5. Cross-Site Scripting (XSS): Detection methods also average around 85-95% [1], with advanced techniques such as DOM-based XSS posing detection challenges.
6. Malware Propagation: Signature-based detection offers up to 98% accuracy for zero-day attacks [5], although encrypted malware presents challenges requiring more advanced behavioural detection.

Cybersecurity Resilience Methods and Accuracies: Windows vs. Linux Systems

Operating system-specific attacks and defences play a crucial role in cybersecurity resilience. Windows and Linux systems, due to their distinct architectures and usage environments, are subject to different types of attacks and require different security measures.

Windows Systems

Windows, being the most widely used operating system, is often targeted by attackers for several reasons: its popularity, its broad range of enterprise applications, and frequent vulnerabilities in older versions. Cybersecurity resilience on Windows systems typically focuses on the following areas:

- Ransomware Attacks: Windows systems are a frequent target for ransomware attacks due to their widespread use. Detection methods such as endpoint protection systems and signature-based detection typically achieve an accuracy of 85-95% [4], though advanced ransomware variants can evade detection.
- Phishing and Social Engineering: Due to the popularity of Windows in enterprise environments, users often fall victim to phishing attacks. Anti-phishing

technologies have a detection accuracy of 85-95%, with spear phishing remaining a challenge with lower detection rates of 70-80%.

- Malware and Viruses: Malware attacks on Windows systems are typically addressed through antivirus software and behavioural-based analysis, achieving detection rates of 90-98% [6]. However, polymorphic malware and sophisticated rootkits can reduce this accuracy.
- Zero-Day Exploits: While Windows security patches are released regularly, the window of vulnerability for zero-day exploits can be significant. Detection methods focusing on behaviour-based analysis for zero-day exploits on Windows achieve accuracies of 70-85%, depending on the complexity of the exploit.

Linux Systems

Linux systems, though less common in consumer environments, are widely used in server and cloud environments, making them a frequent target for specific attacks. The cybersecurity resilience methods and accuracy for Linux systems focus on different challenges compared to Windows:

- Privilege Escalation Attacks: Linux systems are prone to privilege escalation attacks where attackers exploit vulnerabilities to gain root access. Detection mechanisms such as SELinux policies, AppArmor, and behaviour-based monitoring typically offer an accuracy of 80-90%.
- DDoS Attacks: As Linux is commonly used in server environments, it is a frequent target for Distributed Denial of Service (DDoS) attacks. Detection methods based on traffic analysis provide an 80-90% accuracy [2], but sophisticated DDoS variants, such as traffic mimicking legitimate behaviour, can reduce detection effectiveness.
- SSH Brute Force Attacks: Linux systems often face brute force attacks on SSH services. Tools like Fail2Ban and advanced monitoring systems provide detection accuracies of 85-95% by identifying repeated failed login attempts.
- Rootkits and Malware: While less frequently targeted than Windows, Linux systems face rootkits and malware propagation. Modern behaviour-based detection systems achieve 90-95% accuracy [5], though detecting rootkits in kernel space remains a complex challenge.

Comparison of Detection Accuracy: Windows vs. Linux

- Ransomware and Malware: Windows systems face higher risks due to broader usage, but detection methods for both platforms provide similar accuracy, with 85-98% depending on the attack complexity [4].

- Privilege Escalation and Brute Force Attacks: Linux systems excel in detecting brute force and privilege escalation attacks with 80-95% accuracy, due to specialised tools like SELinux, AppArmor, and SSH protection tools.
- DDoS Attacks: Both Windows and Linux systems are similarly equipped to handle DDoS attacks, achieving 80-90% accuracy in detection. However, Linux servers, given their frequent use in cloud environments, might require additional network security layers to maintain resilience.

Improved LLM Detection Accuracy

Recently our team had a breakthrough that led to achieving approximately 95% accuracy in using large language models (LLMs) for detecting cybersecurity attacks compared to our previous average of 47%. This is a significant leap from previous levels and leads to new possibilities for enhancing cybersecurity resilience. The integration of LLMs into attack detection systems can further enhance the following areas:

- Anomaly Detection: With ~95% accuracy, LLMs could now more effectively detect subtle anomalies in network traffic, logs, and user behaviour, helping to identify emerging threats that were previously challenging to catch.
- Real-Time Threat Detection: The improvement in LLM accuracy would allow organisations to employ these models in real-time monitoring, making it possible to create models that detect sophisticated multi-stage intrusions and coordinated attacks with a high success rate.
- Reduced False Positives: The higher accuracy achieved with LLMs also reduces the number of false positives, enabling cybersecurity teams to focus their efforts on high-probability threats. This improvement streamlines incident response and reduces the operational burden on security analysts.
- Adaptive Threat Intelligence: With the enhanced accuracy and the speed of LLM adaptation, we can now attempt to incorporate real-time threat intelligence, dynamically adjusting to new attack patterns and techniques. This creates a more robust and flexible cybersecurity defence that adapts to evolving threats/attacks.

Implications of Our Findings for Enhancing Cybersecurity Resilience

Our research into the application of generative AI for automated data labeling within intrusion detection systems (IDS) in IoT networks has revealed potential improvements in several phases of cybersecurity resilience. Below, we outline how our findings align with specific stages of resilience:

Phase 1 – Preparation and Planning

Although not a direct focus of our research, the automation of data labeling has implications for the **preparation and planning** phase of cybersecurity. The efficiency of automated labeling can streamline the development of training datasets for IDS models. This, in turn, could enhance training and awareness efforts by providing more accurate and up-to-date labeled data for cybersecurity training programs. While the direct impact on policy development and risk assessment is limited, our approach could indirectly support these areas by enabling faster adaptation to emerging threats through improved data quality and availability.

Phase 3 – Detection and Monitoring

Our results indicate that the use of generative AI models can enhance the **detection and monitoring phase** of cybersecurity. We achieved significant performance improvements in both binary (up to 99.8% accuracy using the Phi-3-mini model) and multi-class classification tasks (up to 95.6% accuracy with the Gemma model). This demonstrates the viability of automated data labeling to improve continuous monitoring and incident detection within IDS implementations. By increasing accuracy in identifying diverse attack patterns, our approach can bolster the effectiveness of detection mechanisms, enabling faster and more precise identification of potential threats in IoT networks.

Phase 6 – Adaptation and Evolution

Our study demonstrates the potential for **continuous improvement** within the adaptation and evolution phase of cybersecurity resilience. By using techniques like QLoRA (Quantized Low-Rank Adaptation) for fine-tuning, we were able to adapt large models with minimal resource usage, achieving near-state-of-the-art results even in resource-constrained environments. This suggests that iterative refinement of detection capabilities is feasible, allowing for sustained adaptation as threat landscapes evolve. Moreover, our successful integration of AI models for automated labeling reflects a positive step toward leveraging emerging technologies like AI and machine learning, which are essential for advancing cybersecurity resilience in real-world deployments.

While our findings do not suggest direct enhancements to other phases such as **prevention and protection**, **incident response**, or **recovery and restoration**, the improvements identified above highlight a pathway for strengthening overall cybersecurity resilience by focusing on efficient detection, ongoing adaptation, and proactive planning.

These areas align closely with our research goals and provide a foundation for future work aimed at further refining IDS capabilities in complex IoT environments.

This analysis underscores the potential for a generative AI-based approach to significantly impact multiple facets of cybersecurity resilience, particularly in the evolving context of IoT and Industrial IoT (IIoT) networks.

References

[1] Ferrag, M. A., & Maglaras, L. A. (2023). Cross-Site Scripting and SQL Injection detection using web application firewalls. *Cybersecurity Journal*, 6(2), 150–162. https://doi.org/10.1186/s42400-020-00049-3

[2] Zargar, S. T., Joshi, J., & Tipper, D. (2013). A survey of defense mechanisms against Distributed Denial of Service (DDoS) flooding attacks. *IEEE Communications Surveys & Tutorials*, 15(4), 2046-2069. https://doi.org/10.1109/SURV.2013.031413.00127

[3] Moustafa, N., & Slay, J. (2015). UNSW-NB15: A comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). *2015 Military Communications and Information Systems Conference (MilCIS)*, 1-6. https://doi.org/10.1109/MilCIS.2015.7348942

[4] Al-rimy, B. A. S., Maarof, M. A., & Shaid, S. Z. M. (2018). Ransomware threat success factors, taxonomy, and countermeasures: A survey and research directions. *Computers & Security*, 74, 144-166. https://doi.org/10.1016/j.cose.2018.01.001

[5] Al-Yaseen, W. L., Othman, Z. A., & Nazri, M. Z. A. (2017). Hybrid intelligent intrusion detection system using signature-based and anomaly-based approaches. *Journal of King Saud University-Computer and Information Sciences*, 29(4), 365-377. https://doi.org/10.1016/j.jksuci.2015.12.003

[6] Ucci, D., Aniello, L., & Baldoni, R. (2019). Survey of machine learning techniques for malware analysis. *Computers & Security*, 81, 123-147. https://doi.org/10.1016/j.cose.2018.11.001