Data/Process Analysis for Advanced Interoperable Cyber Ranges

Giuseppe Salerno[1,*]

[1]*University of Calabria, Italy*

Abstract

Cyber Ranges (CR) are strategic assets for cyber security that can be used by a wide range of users and for many purposes including cybersecurity education, testing, and research. The main focus of my research includes: exploring new domains and cross-domain scenarios by studying assets, potential weaknesses and vulnerabilities, and specific attack and defense techniques; investigating new enabling technologies and paradigms by leveraging the Digital Twins paradigm; and studying a new model for Attack Graph within the context of Cyber Ranges.

Keywords

Cyber Range, Digital Twin, Knowledge Graph, Attack Graph, Kill Chain

1. Introduction

Recent global security incidents highlight a significant increase in the complexity and impact of cybersecurity threats. Attackers are becoming increasingly sophisticated, utilizing advanced and automated methods in their operations. This situation necessitates enhanced protection for assets and comprehensive training for personnel to counteract these evolving threats. Cyber Ranges (CRs), as described by the National Institute of Standards and Technology (NIST), are "interactive, simulated representations of an organization's local network, systems, tools, and applications." These environments serve as secure spaces for training Information and Communication Technology (ICT) professionals, preparing them to address a wide array of cyber-attacks and scenarios.

A critical factor in the success of Cyber Ranges is their ability to accurately simulate complex, real-world environments. To this end, my research explores the integration of the Digital Twin (DT) paradigm to augment the functionality and scope of Cyber Ranges.

In more detail, Section 2 of this document, introduces a novel approach that utilizes Digital Twins technology, aimed at enhancing the capabilities and effectiveness of Cyber Ranges. This method facilitates ongoing monitoring of physical systems during the operation of Digital Twins and addresses the absence of specialized passthrough connectors typically unavailable in conventional Cyber Ranges. Moreover, it enables the accurate emulation of complex environmental behaviors. This is accomplished either through executable models that abstract real-world assets or, for software systems, by generating virtual replicas. In the contex of

SEBD 2024: 32nd Symposium on Advanced Database Systems, June 23-26, 2024, Villasimius, Sardinia, Italy

*Corresponding author.

giuseppe.salerno@dimes.unical.it (G. Salerno)

0009-0004-2502-9667 (G. Salerno)

virtualized environments such as Cyber Ranges, particularly those applied to Industrial Control Systems (ICS) and Internet of Things (IoT) devices, my research has extended into the areas of Penetration Testing and Vulnerability Assessment methodologies. This exploration led to the investigation of Attack Graph-based approaches. These methodologies model the process by which an external attacker sequentially executes attacks to progressively gain privileges in any computer system until achieving their ultimate goal of system compromise. In Section 3, I will discuss a novel Attack Graph model and provide an example scenario to demonstrate its practical application.

2. Enhancing Cyber Ranges with Digital Twin Integration via Knowledge Graph

2.1. Knowledge graph-based digital twin

Digital Twins are defined as dynamic, virtual replicas of physical systems, continuously synchronized to mirror the real system's performance and health status throughout its lifecycle [1]. A Knowledge Graph (KG) is a graph-based data structure designed to enhance contextual understanding by interconnecting metadata [2]. It proves particularly well-suited for applications in scenarios demanding the integration, management, and extraction of value from diverse sources on a large scale. Knowledge graphs offer numerous advantages over traditional data models, facilitating the modeling, structuring, management, and analysis of heterogeneous and complex data with dynamic relationships.

A dynamic knowledge graph proves to be an ideal basis for digital twins [3]. This knowledge graph integrates ontologies and autonomous agents that consistently engage with it. Utilizing ontologies facilitates standardized data use, promoting reuse and interoperability [4]. The multi-domain aspect enables the incorporation of new ontologies and the establishment of relationships between related terms, thereby enhancing connectivity. The interlinking of concepts and instances in knowledge graphs, complemented by dynamic updates from computational agents and real-time data feeds, facilitates numerous interactions among participants within a given digital twin.

The representation in the form of a Knowledge Graph of an entire environment is advantageous from a security standpoint, because through a unified graph view, it is possible to analyze every possible entry point and study the entire attack surface effectively.

It is therefore a fundamental starting point for subsequent studies and analyses to have a good, robust and well-defined knowledge base.

Hence, the first step in this research project will be to finalize the definition of a model for representing Knowledge Graph-Based Digital Twins. Each individual digital twin can be interconnected to other entities. The relationships between DTs will culminate in a Knowledge Graph, which will serve as the representation of the ultimate Cyber Range.

2.2. Digital Twin for ICS and Security

Historically, the idea behind the development of digital twin was to monitor and manage the performance of physical systems in the context of Industry 4.0 and smart manufacturing. The

construction of a digital twin for a physical item involves three key aspects: (1) identifying the components and parameters of the physical product in its real environment, (2) establishing a link between the physical and virtual versions of the product, and (3) integrating data and information to bridge the virtual and real worlds [5].

In the context of cybersecurity, digital twin applications have become increasingly significant [6]. For instance, they can be integrated with cyber ranges to analyze system behavior under different cyber attack scenarios. Indeed, digital twins can be used in attack emulations and simulations in order to evaluate resilience metrics, ultimately aiding in the design of security and safety mechanisms for cyber (physical) systems.

Notably, digital twins can also act effectively as honeypots, offering a proactive strategy for uncovering attack vectors within a network [7]. The advantage of using a digital twin as a honeypot is its ability to enhance both the level of interaction and attraction of the "twin" [8].

In this context, my research concentrates on addressing cybersecurity challenges in ICSs, with the goal of developing knowledge-based attack graphs that are also physics-aware. This approach aims to enable the orchestration of targeted attacks, that extends beyond mere denial of service [9]. Furthermore, I will be exploring the development remediation tactics, defensive mechanisms, and strategies to protect intelligent infrastructures against such targeted threats [10, 11, 12]. To this end, I will study the synergy between process mining techniques for detecting physics-aware attacks and the application of game theory models [13, 14, 15]. Moreover, to build models of malware behavior that are not only precise but also fast to discover and interpretable by humans, I intend to investigate effective log encoding [16] for advanced process mining methods [17, 18] paired with explainable AI [19] exploiting efficient computation schemes [20, 21]. An additional, promising avenue for future research involves addressing the challenges of maintaining anonymity in industrial IoT communication networks. Particularly, the principles of sender and relationship anonymity, similar to those applied in the Tor network, could substantially enhance the security of industrial communications [22]. By adapting protocols that offer sender anonymity against global passive adversaries, ICSs can be safeguarded against sophisticated adversaries monitoring critical network points [23]. Furthermore, incorporating privacy-preserving techniques from social networks and IoT, such as those for short communications and MQTT-anonymous protocols, could enhance the robustness of ICSs against advanced persistent threats [24].

3. Exploring and Developing an Attack Graph Approach in Cyber Range Environments

The ever-evolving capabilities of cyber attackers force security administrators to prioritize the early detection of emerging threats. Targeted cyber attacks commonly progress through multiple stages, spanning from the initial reconnaissance of the network environment to the eventual impact on objectives. Multi-step attacks can be conceptualized using the military kill chain concept. The cyber kill chain conceptualizes attacks as sequences of steps. It assumes that the attacker initially identifies suitable targets, then prepares the necessary deliverables, and subsequently transmits them into the environment. Another threat model is provided by attack graphs, which illustrate the paths taken by attackers through the network. Typically,

attackers achieve a series of attack steps, where each step grants them certain privileges on protected assets. During my research, I investigated approaches related to Kill Chain Attack Graphs. I studied the approach proposed by Sadlek et al. [25], which combines the kill chain and the attack graph concepts. It allows representing chains of attacker's actions divided into kill chain phases. According to their definition a Kill Chain Attack Graph (KCAG) is an ordered triple $(\mathcal{G}, P, f)$ where $\mathcal{G} = (V, E)$ denotes a directed graph with vertices V and edges E. A set P contains kill chain phases, and a function f assigns kill chain phases to attack techniques. Whereas Sheyner et al. defined an attack graph as a tuple of states, transitions between the states, an initial state and success states [26]. Ou et al. introduced the concept of a logical attack graph, which is a bipartite directed graph consisting of fact and derivation nodes. Each fact node is labeled with a logical statement represented as a predicate applied to its arguments, whereas each derivation node is labeled with an interaction rule utilized in the derivation step. The edges within a logical attack graph denote a "depends on" relation [27].

In the initial phase of my research, I defined an attack graph as a directed graph $\mathcal{G} = (V, E)$, whose vertices V are entities and whose edges E denote specific relationships or actions. The vertices in this graph are classified into four categories:

- **Attacker**: This vertex represents the knowledge and control an attacker possesses over an asset, underlining the capabilities and potential strategies at their disposal.
- **Asset**: This refers to any component, be it a system, network, or resource, susceptible to cyber threats.
- **Vulnerabilities and Properties**: This category includes the exploitable weaknesses or characteristics of an asset.
- **Attack Goals**: This specifies the final aims or targets the attacker seeks to accomplish, which could range from compromising data integrity to system disruption or control.

Specifically, an attacker's control over assets is differentiated into three levels. Level zero indicates unawareness of the asset's existence. At level one, the attacker is aware of the asset but lacks any control or capability to breach its security. The highest level identifies the attacker's ability to violate the asset's security protocols. Next, we have five asset categories: hosts, processes, individuals, technologies, and data. Properties/Vulnerabilities of assets constitute the third type of vertices. They include information about network services, vulnerable applications, user accounts, etc. A vulnerability can be a known Common Vulnerabilities and Exposures (CVE) or a custom vulnerability/bug present in a host, application code, service misconfiguration, and so on.. Attack goals, the fourth vertex type, denote the attacker's end targets and feature only incoming edges, indicating their terminal nature within the graph.

Furthermore, we have the following types of edges:

1. Edges connecting the first and second vertex types representing steps in the attack progression.
2. Edges linking the second vertex type back to the first (or to an attack goal) illustrate the control level an attacker acquires over an asset post-attack, as part of the attack sequence.
3. Edges from the third to the second vertex type, labeled "hasProperty," associate an asset with its properties or vulnerabilities.

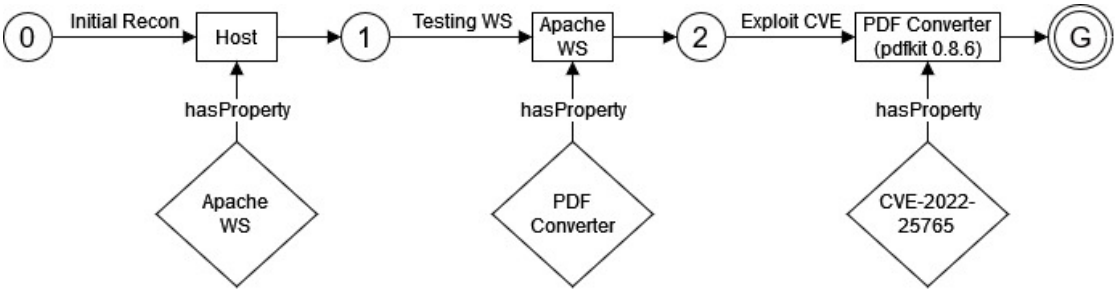

Figure 1: Attack Graph Example. Vertices corrispond to: attacker and their level of knowledge/control over an asset (cirle), asset (rectangle), property and/or vulnerability of an asset (rumble), attack goal (double-line circle).

An instance of a possible attack graph is depicted in Figure 1. In this example scenario, a server exposes a web application on port 80. At the first step (0), the attacker does not have knowledge of the services the server is exposing. After a reconnaissance phase, the attacker identifies the presence of a web server and initiates an analysis and testing phase on it (1). At the end of this phase, he has gained further knowledge and discovered that there is a "PDF Converter" functionality on the web server with a known CVE, which allows "remote command injection" and thus a reverse shell on the server. Consequently, by exploiting the CVE (2), the attacker gains access to a reverse shell and achieves Arbitrary Code Execution on the remote server (G). The model introduced aims to provide a comprehensive perspective on the specific attacker' strategies and processes over a scenario described by means of knowledge-graph.

Conclusion

The domain of Cyber Ranges within the cybersecurity landscape covers a vast range of challenges that can be approached from various perspectives. Currently, there is a lack of comprehensive Knowledge Graph-based models capable of representing any cyber-physical system or object as a Digital Twin. Many existing solutions are not suitable to specific devices such as Industrial Control Systems and IoT devices. Hence, the primary foundational step of this research involves finalizing the definition of the general model for representing Digital Twins. Additionally by proposing a new model for Attack Graph within the CR context, this research contributes to advancing the efficacy and versatility of cyber defense strategies. My definition of the attack graph has made it possible to represent threats intuitively and to clearly outline the potential phases of a cyber attack.

Acknowledgements

This work was partially supported by project SERICS (PE00000014) under the MUR National Recovery and Resilience Plan funded by the European Union – NextGenerationEU.

References

[1] A. M. Madni, C. C. Madni, S. D. Lucero, Leveraging digital twin technology in model-based systems engineering, Syst. 7 (2019) 7. URL: https://api.semanticscholar.org/CorpusID:86548244.

[2] A. Hogan, E. Blomqvist, M. Cochez, C. d'Amato, G. D. Melo, C. Gutierrez, S. Kirrane, J. E. L. Gayo, R. Navigli, S. Neumaier, et al., Knowledge graphs, ACM Computing Surveys (Csur) 54 (2021) 1–37.

[3] C. Ramonell, R. Chacón, H. Posada, Knowledge graph-based data integration system for digital twins of built assets, Automation in Construction 156 (2023) 105109. URL: https://www.sciencedirect.com/science/article/pii/S0926580523003692. doi:https://doi.org/10.1016/j.autcon.2023.105109.

[4] J. Akroyd, S. Mosbach, A. Bhave, M. Kraft, Universal digital twin - a dynamic knowledge graph, Data-Centric Engineering 2 (2021) e14. doi:10.1017/dce.2021.10.

[5] G. Epiphaniou, M. Hammoudeh, H. Yuan, C. Maple, U. Ani, Digital twins in cyber effects modelling of iot/cps points of low resilience, Simulation Modelling Practice and Theory 125 (2023) 102744.

[6] E. Russo, G. Costa, G. Longo, A. Armando, A. Merlo, Lidite: A full-fledged and feather-weight digital twin framework, IEEE Transactions on Dependable and Secure Computing 20 (2023) 4899–4912. doi:10.1109/TDSC.2023.3236798.

[7] M. Lucchese, F. Lupia, M. Merro, F. Paci, N. Zannone, A. Furfaro, HoneyICS: A High-interaction Physics-aware Honeynet for Industrial Control Systems, in: Proceedings of the 18th International Conference on Availability, Reliability and Security, ARES '23, Association for Computing Machinery, New York, NY, USA, 2023. URL: https://doi.org/10.1145/3600160.3604984. doi:10.1145/3600160.3604984.

[8] F. Lupia, M. Lucchese, M. Merro, N. Zannone, ICS Honeypot Interactions: A Latitudinal Study, in: 2023 IEEE International Conference on Big Data (BigData), 2023, pp. 3025–3034. doi:10.1109/BigData59044.2023.10386497.

[9] G. Longo, F. Lupia, A. Pugliese, E. Russo, Physics-aware targeted attacks against maritime industrial control systems, Journal of Information Security and Applications 82 (2024) 103724. doi:https://doi.org/10.1016/j.jisa.2024.103724.

[10] G. Longo, A. Orlich, A. Merlo, E. Russo, Enabling real-time remote monitoring of ships by lossless protocol transformations, IEEE Transactions on Intelligent Transportation Systems 24 (2023) 7285–7295. doi:10.1109/TITS.2023.3258365.

[11] G. Fortino, C. Greco, A. Guzzo, M. Ianni, Neural network based temporal point processes for attack detection in industrial control systems, in: 2022 IEEE International Conference on Cyber Security and Resilience (CSR), 2022, pp. 221–226. doi:10.1109/CSR54599.2022.9850333.

[12] G. Fortino, C. Greco, A. Guzzo, M. Ianni, Identification and prediction of attacks to industrial control systems using temporal point processes, Journal of Ambient Intelligence and Humanized Computing (2022). URL: https://doi.org/10.1007/s12652-022-04416-5. doi:10.1007/s12652-022-04416-5.

[13] G. Greco, F. Lupia, F. Scarcello, Coalitional games induced by matching problems: Complexity and islands of tractability for the Shapley value, Artif. Intell. 278 (2020).

[14] G. Greco, F. Lupia, F. Scarcello, Structural tractability of shapley and banzhaf values in allocation games, in: Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015, AAAI Press, 2015, pp. 547–553.

[15] S. Saraeian, B. Shirazi, Process mining-based anomaly detection of additive manufacturing process activities using a game theory modeling approach, Computers & Industrial Engineering 146 (2020) 106584.

[16] M. Ianni, E. Masciari, Scout: Security by computing outliers on activity logs, Computers & Security 132 (2023) 103355. URL: https://www.sciencedirect.com/science/article/pii/S0167404823002651. doi:https://doi.org/10.1016/j.cose.2023.103355.

[17] G. Greco, A. Guzzo, F. Lupia, L. Pontieri, Process Discovery under Precedence Constraints, ACM Trans. Knowl. Discov. Data 9 (2015) 32:1–32:39.

[18] M. L. Bernardi, M. Cimitile, F. M. Maggi, Data-aware process discovery for malware detection: an empirical study, Mach. Learn. 112 (2023) 1171–1199.

[19] C. Greco, M. Ianni, A. Guzzo, G. Fortino, Explaining binary obfuscation, in: 2023 IEEE International Conference on Cyber Security and Resilience (CSR), 2023, pp. 22–27. doi:10.1109/CSR57506.2023.10224825.

[20] F. Lupia, A. Mendicelli, A. Ribichini, F. Scarcello, M. Schaerf, Computing the Shapley value in allocation problems: approximations and bounds, with an application to the Italian VQR research assessment program, J. Exp. Theor. Artif. Intell. 30 (2018) 505–524.

[21] G. Greco, F. Lupia, F. Scarcello, The tractability of the shapley value over bounded treewidth matching games, in: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017, ijcai.org, 2017, pp. 1046–1052. URL: https://doi.org/10.24963/ijcai.2017/145. doi:10.24963/IJCAI.2017/145.

[22] F. Buccafurri, V. De Angelis, M. F. Idone, C. Labrini, A protocol for anonymous short communications in social networks and its application to proximity-based services, Online Social Networks and Media 31 (2022) 100221. URL: https://www.sciencedirect.com/science/article/pii/S2468696422000258. doi:https://doi.org/10.1016/j.osnem.2022.100221.

[23] F. Buccafurri, V. De Angelis, M. F. Idone, C. Labrini, S. Lazzaro, Achieving sender anonymity in tor against the global passive adversary, Applied Sciences 12 (2022). URL: https://www.mdpi.com/2076-3417/12/1/137. doi:10.3390/app12010137.

[24] F. Buccafurri, V. de Angelis, S. Lazzaro, Mqtt-a: A broker-bridging p2p architecture to achieve anonymity in mqtt, IEEE Internet of Things Journal 10 (2023) 15443–15463. doi:10.1109/JIOT.2023.3264019.

[25] L. Sadlek, P. Celeda, D. Tovarnak, Identification of attack paths using kill chain and attack graphs, in: NOMS 2022-2022 IEEE/IFIP Network Operations and Management Symposium, IEEE, 2022.

[26] O. Sheyner, J. Haines, S. Jha, R. Lippmann, J. M. Wing, Automated generation and analysis of attack graphs, in: Proceedings 2002 IEEE Symposium on Security and Privacy, IEEE, 2002, pp. 273–284.

[27] X. Ou, W. F. Boyer, M. A. McQueen, A scalable approach to attack graph generation, in: Proceedings of the 13th ACM Conference on Computer and Communications Security, CCS '06, Association for Computing Machinery, New York, NY, USA, 2006, p. 336–345.