Security Attacks and Solutions for Digital Twins

Sabah Suhail, Sherali Zeadally, Raja Jurdak, *Senior Member, IEEE*, Rasheed Hussain, *Senior Member, IEEE*, Raimundas Matulevičius, and Davor Svetinovic, *Senior Member, IEEE*

***Abstract*—Digital twins, being the virtual replicas of their physical counterparts, share the expected functional requirements and operational behavior of the underlying systems. Due to this fact, digital twins may act as a potential source of data breaches. Attackers may exploit the valuable knowledge about the system accessible through digital twins to put digital twins into a malicious state. We focus on potential attack strategies on digital twins ranging from digital twin's design to the dismissal phase. Next, we discuss defensive strategies to thwart the identified attacks on digital twins. Finally, we outline future research challenges that must be addressed to deal with malicious digital twins.**

***Index Terms*—Blockchain, Cyber-Physical System (CPS), Digital Twin, Industrial Control System (ICS), Intrusion Detection.**

I. Introduction

The convergence of Information Technology (IT) and Operational Technology (OT) in Industrial Control Systems (ICSs) enable the realization of the Industry 4.0 vision [1]. At the core of Industry 4.0 are the Cyber-Physical Systems (CPSs) that connects physical (OT) and cyber (IT) components through computational and networking capabilities [2]. ICSs are a subset of CPSs. ICSs provide promising solutions to various industrial ecosystems but they substantially expand the attack surface [3]. By exploiting different attack vectors (i.e., cyber and physical), attackers can launch Advanced Persistent Threats (APTs) through which they covertly reside in the system to continually exfiltrate information or undermine the critical processes. To ensure that the system operates securely and safely, we need essential measures to secure CPSs: (1) evaluating the operational behavior of the system, and (2) conducting penetration testing on the system to identify vulnerabilities or threats [4]. As CPSs cannot be deactivated for carrying out such analysis, assessing the system's security level requires online solutions that accurately reflect the actual CPS operations while avoiding any interference or side-effects of testing on the live systems [4]. Digital Twins are one such promising solution [3] which addresses this constraint.

Digital twin is a virtual replica of a physical asset (device or process) that enables analyzing, predicting, and optimizing operations by utilizing real-time and historical data [4]. In the information security domain, digital twins strengthen the security of CPS through various security-enhancing use cases, including system training and testing, security testing, and detecting system misconfigurations [5]. To do so, digital twins run synchronously with their physical counterparts where the goal is to track *data inconsistencies* between the physical and the virtual entity [6]. Digital twins collect and integrate data from multiple sources, such as sensory data from the physical environment, historical data from CPS lifecycle phases, and domain knowledge from experts to learn the behavior of the physical environment, which, in turn, serves as a valuable input for detecting anomalies. Then, following a closed-loop, we feed the optimized data to the physical entity to adapt operations to latest state operations.

Despite various security-enhancing use cases of digital twins (as Fig. 1 shows) in systems engineering or during the operation phase of CPS [2], the emergence of stealthy threats allow attackers to exploit digital twins to launch attacks on the CPS. Digital twins, being the virtual replicas of their physical counterparts, share functional requirements and operational behavior of the underlying systems. Therefore, digital twins may act as a potential source of data breaches, leading to the abuse case of digital twins [4]. Attackers may exploit the deep knowledge about the physical process and corresponding control devices accessible through digital twins with a two-stage strategy: use the key input data source namely, digital twins into a malicious state, and then through that state manipulate the underlying physical system's behavior covertly [4]. For example, manipulating the behavior of digital twins by modifying their defined states which would correspond to a direct attack on field devices, particularly when automated feedback loops are enabled between the physical objects and their digital counterparts [2]. It is necessary to ensure the trustworthiness of digital twins for timely corrections because ignoring such pre-emptive measures may lead to a feedback loop of erroneous data into the system resulting in the Garbage In Garbage Out (GIGO) problem [6]. Moreover, in human-machine collaboration scenarios, a slight system dysfunction caused by mirroring of malicious replicas may pose a severe threat to human safety. The repercussions of exploiting digital twins may have severe consequences within the *digital thread* that links data throughout various phases of the CPS lifecycle (as Fig. 2 shows). These links are an attractive target for attacks because the entire product lifecycle can be targeted in a data breach [3] such as when manipulating high-valued design artifacts. Furthermore, attack on digital thread may affect the next generation CPS system where digital twins data can be used as historical data. The authors of [2] have showed how digital twins can be exploited for launching attacks as one

S. Suhail and R. Matulevičius are with University of Tartu, Estonia (e-mail: (sabah.suhail, raimundas.matulevicius)@ut.ee).

S. Zeadally is with University of Kentucky, US (e-mail: szeadally@uky.edu).

R. Jurdak is with Queensland University of Technology, Australia (e-mail: r.jurdak@qut.edu.au).

R. Hussain is with the Bristol Digital Futures Institute (BDFI) and Smart Internet Lab, University of Bristol, UK (e-mail: rasheed.hussain@bristol.ac.uk).

D. Svetinovic is with the Research Institute for Cryptoeconomics, Vienna University of Economics and Business, Austria, and the Center for Cyber-Physical Systems, Department of Electrical Engineering and Computer Science, Khalifa University of Science and Technology, UAE (e-mail: davor.svetinovic@wu.ac.at).

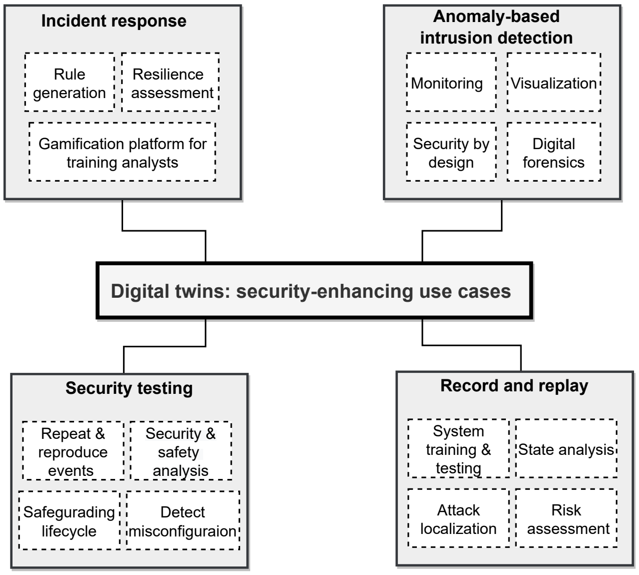

Fig. 1. Security-enhancing use cases of digital twins.

of the open research directions. However, to the best of our knowledge, there are no previous studies on the abuse case of digital twins. Our main research contributions are as follows:

- We discuss possible attacks on digital twins where the attacker defeats the security of digital twins by either manipulating the benign behavior of digital twins or by exploiting the cyclic state update from the physical process to digital twins to steer the CPS into an insecure state.
- We discuss potential security solutions that can mitigate the possible attacks (we have identified) on digital twins.
- In this context, we propose a digital twin-based gamification approach that can access the security level of the digital twins. Furthermore, the gamification approach provides security analysts with a controlled, supportive virtual training environment.

II. Digital Twins Abuse Case : An Attacker's Perspective

To understand the anatomy of a cyberattack, we need to understand the adversary tactics (as Table I shows). The following section discusses different attackers' strategies on digital twins.

A. *Reconnaissance attacks*

Reconnaissance involves intelligence gathering. This is achieved through activities such as network scanning, exploiting zero-day vulnerabilities, and enumerating services to identify security loopholes in the underlying system. For instance, the Triton malware targeted a petrochemical plant in Saudi Arabia and gained a foothold in the IT/OT networks to target Safety Instrumented Systems (SIS). However, the attacker may go beyond conventional network reconnaissance in industrial ecosystems to achieve the desired objectives. Sophisticated malware can defeat isolation mechanisms, including air gaps, sandboxes, virtualization, and so on. For example, the Stuxnet malware aimed at a Uranium enrichment plant demonstrates how to overcome air gaps [7]. Thus, beginning with reconnaissance scans, the attacker may gather information about the loopholes in the infrastructure and then use Stuxnet- or Triton-inspired malware strategies to launch attacks on digital twins.

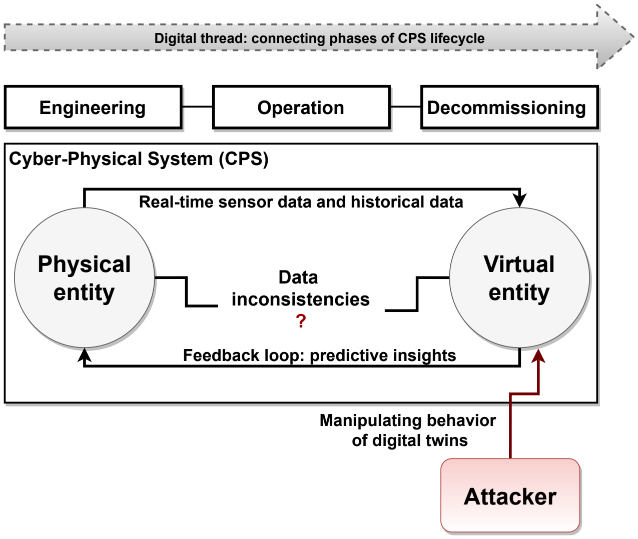
Digital thread: connecting phases of CPS lifecycle
Engineering
Operation
Decommissioning
Cyber-Physical System (CPS)
Real-time sensor data and historical data
Physical entity
Data inconsistencies
?
Virtual entity
Feedback loop: predictive insights
Manipulating behavior of digital twins
Attacker

Fig. 2. Exploiting digital twins to launch attacks.

B. *Which digital twins mode could cause more damage to the system?*

Digital twins do not need to replicate the CPS in its entirety. Given that the virtual representation in digital twins mimics the functionality of corresponding processes or devices with enough details reasonable feature generalizations or simplifications can occur if they stay context-aware [8]. More precisely, the goal of building digital twins is to provide

TABLE I
ATTACK ON DIGITAL TWINS: AN ATTACKER'S PERSPECTIVE.

Attacker's artifacts	Goals
Product lifecycle	• Manipulate benign behavior of digital twins to steer the CPS into an insecure state • Exploit digital thread as it links data throughout the entire product lifecycle
Replication mode	Run direct cyclic state updates by replicating the virtual behavior of digital twins to the corresponding program states of physical devices
Simulation mode	• Learn system behavior by re-running test simulations • Manipulate simulation parameters or system specifications' data during security tests
Design phase	Exploit specification-based or machine learning-based process knowledge of digital twins
Decommissioning phase	• Retain knowledge about the system's life for reuse due to improper disposal of digital twins • Use data security breach such as unauthorized access to gain access to archived digital twins' data
Lateral movement	• Gain control over high-value assets such as design artifacts • Manipulate sensor readings or simulation parameters at random intervals while ensuring that the new values do not deviate significantly from the real process values

a cost-effective solution to test the physical system rather than replicating (in terms of simulation or emulation) the system. Nevertheless, accurate representation of digital twins also contributes to the likelihood of a successful attack. In this case, the attacker considers the operation modes of digital twins. Next, we discuss the operation modes of digital twins which could be exploited by attackers. In replication mode, we record the event in a real system and then replay it while emulating the system behavior [1]. To do so, twins and their physical counterparts must be synchronized through sensor measurements, network communication, or log files [4]. We maintain a constant connection with the physical counterpart by integrating the system specification and the current state's data. The replication mode may initiate direct cyclic updates to and from the digital twins. However, to use the replication mode, the attacker needs to stay active to avoid the problem of time-dependent state synchronization and consistency of information between the physical entity and its different replicas (as Fig. 3 shows).

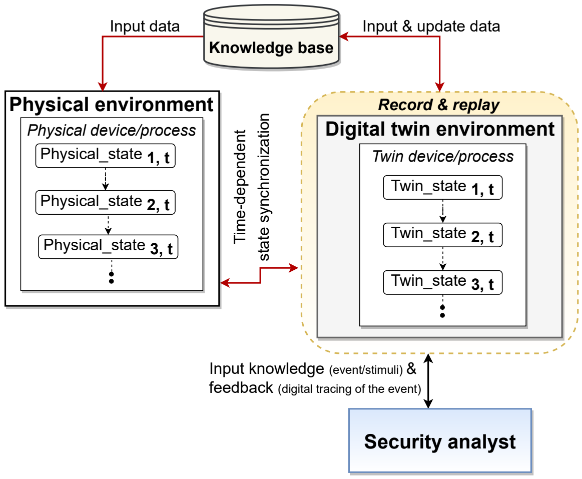

Fig. 3. Replication mode.

Simulation mode operates in an isolated virtual environment without having a direct connection to the live systems. It requires user-specified settings and parameters as input (as Fig. 4 shows). In simulation mode, being reproducible and repeatable with a broad range of trial-and-error learning mechanisms, can be directly used or tailored according to the attacker's needs. The attacker can reveal emergent system behaviors by resetting and re-running the simulation until he/she achieves his/her insidious goals. Even worse, it can target the theme of simulation mode - security by design by reversing the defined configurations during security tests within the virtual environment. Furthermore, the attacker can learn the system state passively. However, since the simulation mode runs independently of its physical counterpart, the attacker cannot trigger automated attacks on the system due to the absence of a direct feedback loop.

C. *Victimizing physical system or twin?*

1) Targeting the physical system: Integrating general-purpose IT systems with ICSs introduces novel attack vectors [3]. Usually operational functionality outweighs security, therefore loopholes in the system infrastructure allow attackers to launch advanced covert attacks (e.g., APTs). ICS-tailored malware (for instance, Stuxnet), is one such example wherein a

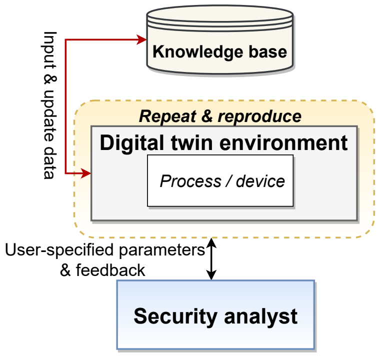

Fig. 4. Simulation mode.

malicious code self-protects and self-updates itself while intercepting and modifying the data sent to and from Programmable Logic Controllers (PLCs) to compromise its target covertly [4]. Despite the high level of effort needed, attackers may choose to directly interact with the physical system (through industrial espionage or a cyberwarfare weapon) with the objective to damage or destroy it.

2) Targeting the digital twin: When attackers choose digital twins as a target they can ultimately destroy the physical asset. This is because launching attacks on digital twins is like launching them on the physical system because malicious code can intercept and modify the simulation parameters in the digital twins of PLCs.

Next, we discuss motivations behind attacks on digital twins:

- Digital twins can serve as an anomaly-based intrusion detection tool wherein through time and state, users (both benign and malicious) can continuously monitor the ongoing processes to observe the expected behavior of the system [4]. For instance, analyzing the relationships among dynamic variables (of the physical process) and historical variables (of the virtual process) facilitates the detection of Safety and Security (S&S) rule violations. On the one hand, a benign user uses digital twins to spot deviations from a defined or learned baseline and alert security analysts. On the other hand, a malicious user can exploit the correlation of variables to disrupt the digital twins' behavior such that the twins do not exhibit the defined pattern of misbehavior (knowledge-based or behavior-based). As the attacker follows the "living off the land" attack strategy (i.e., without using any of the illegitimate software and functions to perform malicious actions), no intrusion can be detected and thus, it is hard to spot long-term deviations.
- Digital twins, from design to dismissal, are among the key input data sources for the physical systems. For example, the asset prototype is designed and tailored based on the simulation mode in the engineering phase. Then during the operation phase, the physical asset further evolves and optimizes its functionalities based on the simulation or replication modes of digital twins. With these chains of data inputs/outputs, exploiting digital twins may provide insights into attack vectors that can be used to plot long-term attacks on next generation CPS systems.

Digital twins must exhibit sufficient fidelity in terms of functionality and time-sensitive operational behavior of the physical component to protect against failures of the real system [6]. Besides, to maintain backward compatibility, a corresponding evolution of twin is needed as the physical object evolves over time [9]. With such challenges, attackers need to adapt their attack techniques to the new requirements of digital twins.

3) Could the hybrid approach be worse than attacking either the physical or the digital entity?: Considering the complexity and the ever-changing threat landscape, it is possible for the attacker covertly reside at both places, i.e., the physical system and its virtual counterpart. The proliferation of such sophisticated attacks enables cyber-attackers to conceal themselves within enterprise network traffic while actively hunting for valuable data. Even if they get caught while feeding adversarial data into the physical system, attackers may still exploit vulnerable entry points into digital twins.

D. Morphing digital twins through lifecycle phases

1) Engineering phase: The concept of building the process knowledge of digital twins can be achieved in two ways. First, by utilizing the CPS specification (such as the network and/or logic layer) to model the physical counterpart [5]. Second, by utilizing machine learning to learn security-related aspects based on sensor data [10] without obtaining process knowledge from DTs [4]. The process knowledge acquired through the specification of digital twins is less favorable compared to when it is based on machine learning because the former emulates the behavior of the system more closely. But complete transparency into the inner working of Artificial Intelligence (AI) models may expose them to adversarial attacks by allowing cyber attackers to make inferences from live cyber data or perform model poisoning into the training workflows.

2) Decommissioning phase: Usually, due to obsolescence or replacement of outdated hardware/software, the asset and its digital twin are destroyed during the dismissal phase. Nevertheless, as the knowledge about the predecessor system can be rehashed, digital twin data could be backed up and procured by similar objects or domain experts to optimize the next generation of the system. No matter whether the digital twins are destroyed or retained for future usage, attackers might exploit them. For instance, digital twins can be attractive targets of data breach incidents if they are not carefully disposed while complying with proper media sanitization guidelines. Similarly, if digital twins are archived for future usage without complying with adequate data security measures, they can be easy targets of security breaches.

E. Lateral movement

By moving deeper into the system in search of sensitive information or gaining control over high-value assets, attackers strategically target specific sensors and manipulate readings at different points in time, depending on the process dynamics, i.e., time dependent behavior of a process in response to data input.

F. Drawbacks of desirable features of digital twins

During the operation phase of CPSs, cyclic state updates can be allowed from the digital twins to the physical process or vice versa. Although such actions aim to optimize the underlying operations, however, doing so, attacks initiated through digital twins may have similar repercussions as those attacks which are launched directly on real field devices. Similarly, ranging from low to high, the precise representation of digital twins, i.e., fidelity, is essential in the deployment of digital twins. Usually, the desired fidelity depends on the use case of the digital twins. For instance, virtual honeypots [2] need to be more realistic to lure attackers, thus requiring high fidelity. However, cross-checking the engineering knowledge (such as verifying device data against threshold values, identifying unidentified connections or unknown devices, and so on [11]) based on the design specifications of the underlying CPS at the network and logic layer that includes physical devices [5] can be achieved with low-to-medium fidelity digital twins. The design specifications of the underlying CPS at the network and logic layer [5], [11] may provide medium-level fidelity, however, real-time sensor data provides high-level fidelity digital twins. Simply put, the more the digital twins accurately reflect their physical counterpart, the easier it is for the attacker to understand the system behavior.

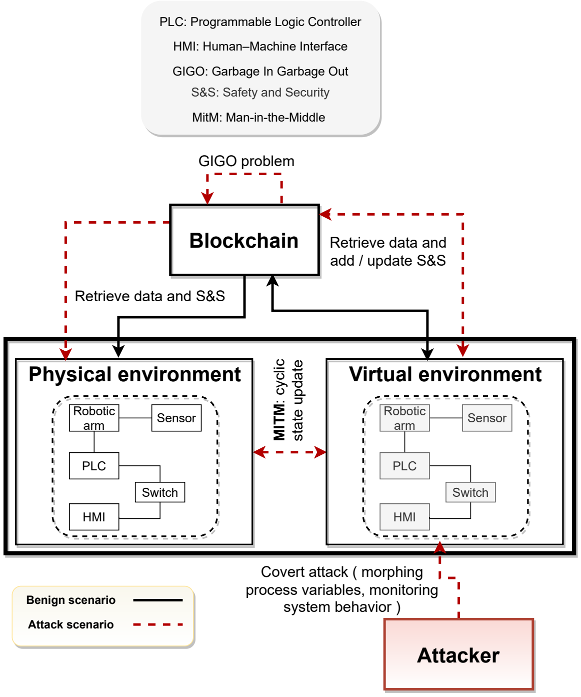

Fig. 5. Garbage In Garbage Out (GIGO) problem in blockchain-based digital twins.

III. Countermeasures

The following section discusses the countermeasures to thwart attacks discussed above on digital twins.

A. Blockchain-based digital twins

Given that the digital twin data is used as an input source to the CPS physical processes, the digital twin must be built on trusted data [4]. In this context, empowering digital twins with blockchain allow industries to manage data on a distributed ledger while ensuring trusted digital twin data coordination across multiple stakeholders [3]. Next, we discuss possible solutions that can mitigate the attacks we have identified earlier on digital twins.

1) Orchestrating provenance: In ICSs, the controllers usually focus on code syntactic disregard for changes originating from authorized engineering stations mistakenly or maliciously [7]. The inability to track changes in the system opens up opportunities for the wrongdoers to compromise targets covertly (such as reconnaissance attacks discussed in section II-A) before launching an overt attack. Similarly, infrastructure vulnerabilities (such as missing or weak authentication and authorization credentials) lead to exploitation of process knowledge of digital twins (during the engineering phase II-D1) and the decommissioning phase (mainly when digital twins are archived for future usage). Therefore, we need to enforce mechanisms that ensure trusted digital twins calls to keep track of activities such as granting access privileges to entity, modifying simulation parameters or state data, adding/updating S&S rules, and so on.

To better understand about the current state of a data object such as *why*, *where*, and *how*, we need to be aware of the complete lineage of process chain (i.e., set of actions performed on data) [6]. In this context, provenance-enabled blockchain-based digital twins assure the traceability and integrity of the data, thereby resulting in more informed decisions made by the underlying systems [3]. Supporting the digital twin engineering phase through provenance-aware blockchain-based solutions allows monitoring of the transitions in the process knowledge (both specification-based and machine learning-based) through time, outliers, and changes. Moreover, introducing an access control model such as Role-Based Access Control (RBAC) that integrates S&S rules (as proposed in [11]) at the digital twin engineering phase lowers security and incident response costs, thereby making subsequent lifecycle phases (such as operation and decommissioning (archived digital twins)) less prone to errors. It is worth noting that it is important to dismantle digital twins or the digital thread while complying with proper media sanitization guidelines. Similarly, recording provenance can help resolve issues during the replication mode cyclic state updates as we have discussed in section II-F. However, the challenges associated with the fidelity of digital twins can be partially addressed through access control mechanisms but cannot be handled entirely through a provenance-based solution and is an area of future research.

2) Securing lifecycle data: Fig. 2 depicts the lifecycle of a digital twin which spans different phases and includes multiple stakeholders who perform various tasks on it. Thus, multiparty use of a digital twin affects confidentiality, integrity, availability, and access control. By establishing a distributed infrastructure, blockchain solves the critical problem of data dissemination across multiple participating entities. Blockchain can manage enterprise policies and rules subject to access rights, i.e., only authorized entities can access, read, and write to the digital twin. Defining access controls may mitigate attacks discussed in section II-A and section II-D. Moreover, using its cryptographic strength, blockchain maintains an irrevocable history of digital twin access transactions. It can therefore circumvent the problem of unauthorized data modifications that may invoke other data security-related problems.

3) Role of smart contracts: Smart contracts allow the execution of code inside a blockchain to automate application-dependent scenarios. When deployed on digital twins, smart contracts can be used to store authorization information for all participating entities [12], track data sharing mechanism [13], represent twin-creation transactions [14] as surveyed in [3]. Additionally, smart contracts, aligned with the predefined conditions, are preferable for scenarios that require automation caused by a change of state, for instance, triggering S&S rules, invoking PLC functions or due to changing conditions of physical processes or modification of simulation setup parameters. Auditing digital twins by actively or retroactively monitoring smart contract transactions [3] can further strengthen the rationale for using smart contracts in digital twins. In this context, auditing helps identify the cause of changes (normal or abnormal). For instance, allowing automation of some of the transaction logic (such as triggering S&S rules) through smart contracts can help to invoke the appropriate defense mechanisms during the engineering and operation phases.

Considering the dynamism and complexity of ICSs, the impact of autonomous smart contracts might exacerbate before humans understand the situation, validate conditions, and control events. For instance, the system cannot adapt to changing conditions if the rules coded in smart contracts are not dynamically updated. Even dynamic modifications (such as those enabled through AI) may result in the risk of AI making unethical decisions.

4) Why may blockchain fail?: Blockchain mechanisms do not guarantee the trustworthiness of data at the source of the information [4]. Thus, any weak link in the process chain, either from a physical or virtual environment, can let the attacker enter and carry out malicious activities in the system. For instance, consider a virtual environment that represents IT/OT components of an industrial robotic arm as Fig. 5 shows. To command and control the robotic arm (such as turning on/off, speed, controlling joint movements) the PLC and the Human–Machine Interface (HMI) are used whereas the sensors collect the event logs. After gaining access to the virtual environment, an adversary can add or update bad safety and security practices that are then recorded in the blockchain to be used by subsequent processes. Such events may result in Garbage In Garbage Out problem (as Fig. 5 shows). Furthermore, even if the blockchain is not involved, the cyclic state updates to and from the digital twin environment to the physical process could be enough to cause the damage. To address the GIGO problem, one potential solution is to ensure the trustworthiness of the sources generating the data. To this end, engineering knowledge describing the design specifications at the network/logic layer of the underlying system can be utilized. Generating the network setup of the virtual environment based on technical, topological, and control artefacts can help model the correct behaviour of the physical counterparts. Furthermore, engineering knowledge can also serve as a basis for implicit security rules such as defining a safe state based on device benign behavior, cross-validating device data against threshold values, detecting unknown devices or unidentified connections [11].

In an industrial system, a massive amount of real-time data is disseminated to and from digital twins in a continuum to create a digital factory. The reluctance to adopt blockchain in such a digital factory in practice remains an open question. To address this concern in blockchain-based digital twins requires further investigation to resolve challenging issues related to scalability, time-varying network delay, time-sensitive tasks (such as real-time remote manipulation), quantum resistance, energy consumption, and integration with legacy systems [3].

B. Bringing gamification to digital twins security

Although digital twins operate virtually in an environment distinct from the live system, are prone to attacks. To thwart attacks on digital twins, one potential solution is to assess the security level of digital twins by launching attacks on them. However, such assessment must be performed in an isolated environment without negatively affecting the operation of digital twin modes (especially replication). To this end, we propose a gamification approach that provides twin assessment and a learning environment for security analysts. The following section discusses how the gamification approach can help evaluate the security of digital twins against the attacks specifically discussed in section II-B, section II-C, and section II-E.

Gamification is the process of incorporating game mechanics into non-game environments. In cybersecurity, the gamification approach aims to provide security analysts with a controlled, supportive virtual training environment. To investigate the resilience of physical processes of the digital twins environment against attack, determining the potential loss incurred in terms of service degradation can be gamified. For instance, the red-blue team cybersecurity exercises [4], penetration testing [2], Capture-The-Flag (CTF) challenges [1], or using cyber range to provide hands-on cyber skills and security posture testing [15] are among well-known approaches in the existing literature. By simulating attack and defense scenarios, without risking critical infrastructures, such solutions reap the following benefits: (i) we can train security analysts by defining context, environment, and learning objectives to gain practical knowledge and skills during an exercise or challenge, and (ii) we can evaluate the security of digital twins and eventually the physical asset. Leveraging gamification for security-awareness training, can further complement automated security testing of digital twins through incident response which may benefit in lateral movement. Furthermore, it can help to identify attacks during the hybrid approach (discussed in section II-C3) due to the connection between the learning environment and CPS (as Fig 6 shows).

Fig. 6 showcases the high-level architecture which includes main components of the proposed digital twin-based gamification approach, where our goal is to get the potential

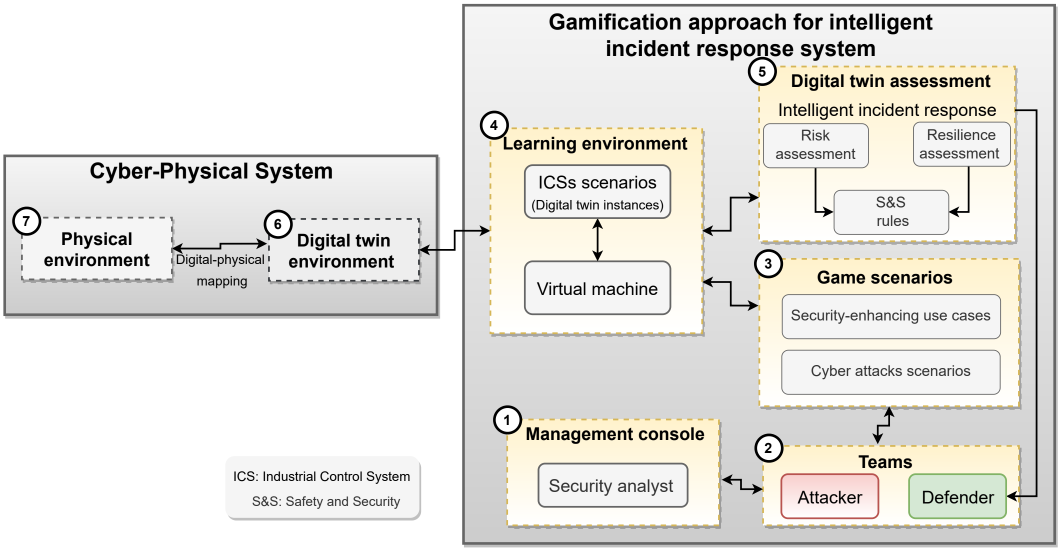

Fig. 6. Gamification approach for digital twins.

defensive solutions by analyzing the attack patterns in a simulated environment (also known as learning environment). Our proposed system comprises two main parts: 1) CPS and 2) gamification approach for intelligent incident response system. The CPS consists of (i) *physical environment* (such as a robotic arm) and (ii) *digital twin environment* (the virtual copy of a robotic arm). The gamification approach for intelligent incident response system consists of (i) *a management console* to assign roles and resources to security analysts, (ii) *teams* which include attacker and defender for respective scenarios, (iii) *game scenarios* with various attack scenarios, (iv) *learning environment* to replicate the digital twin environment for simulation purposes, and (v) *digital twins assessment* which analyzes the risks and resilience to automatically get the rules to improve the defensive mechanisms. Next, we describe the use case for the gamification approach. Initially, a security analyst (*step 1*) chooses the configuration of the teams (*step 2*) made up of attackers or defenders from the management console. Next, a security analyst chooses the game scenarios (*step 3*) that comprises security-enhancing use cases (for example, security testing, system testing and training) and cyber attacks scenarios (for example, Man-in-the-Middle (MitM), intrusion, altering device configurations or simulation parameters). Then, the learning environment (*step 4*) implements the ICS scenarios of different digital twin instances through the virtual machine. The purpose of using a virtual machine is to emulate the functionality of any of the desired digital twin instance without affecting ongoing processes (digital-physical mapping) in CPS. In a traditional approach, the learning material for the analyst is provided through videos or instructional texts. The simulated scenario produces log data documenting the operations. The digital twin assessment module (*step 5*) is then used to analyze the log data. Based on the incident response playbook, scenario-based learning can be reproduced in terms of the level of difficulty to guide the analyst through several training units (to and from step 4 and step 5).

Determining a viable response to a security incident is vital. Incidents can be a precursor to a future attack or it has already occurred or is currently underway. Investigating incidents provides an opportunity to learn and better prepare for similar incidents in the future. More specifically, it can expose activities (such as an attacker impersonating a legitimate user) associated with lateral movement. In this context, the incident response describes the action to be taken based on the type of incident. Responding to a security incident is indispensable to effectively minimize the damage and recovery time while finding and fixing the cause to prevent future attacks. To do so, we introduce an intelligent incident response module that can automate the incident response process. Generally, both the attacker and the defender can use AI agents [16]. The agents deploy machine/deep learning algorithms (such as Generative Adversarial Networks (GANs)) to execute tasks such as attacking and defending the system. Depending on the game scenario, the attacker launches various types of attacks on the system. In contrast, the defender must detect anomalies and determine the risks associated with the occurrence of anomalies. Both the attacker and defender agents can choose the best attack/response strategy to attack/protect the digital twin environment based on the trained data. For instance, an attacker can determine the probability of a successful attack. To improve the agents' capabilities, we can train our AI agents in the learning environment based on data used in actual attacks and simulated data. Moreover, an intelligent incident response can construct or update S&S rules depending on the log data input and resilience assessment sub-modules. As a result, we can develop the intelligent incident response for the CPS, where the intelligent agent (i.e., detect the anomaly and reconfigure the parameters to mitigate an attack or to reduce the damage) comes from the gamification process of the learning environment. Based on the intelligent incident response approach, the learning environment result can be directly replicated to the digital twin environment (*step 6*) and ultimately the physical environment (*step 7*).

To sum up, the proposed gamification approach is best suited for evaluating the security level of both digital modes, i.e., replication and simulation. Furthermore, the required ICS scenario can be virtualized to better understand specific types of attack. Since digital twins are the key source of input data, performing security assessments of digital twins can eventually secure the physical system.

IV. Conclusion

This article focuses on two aspects of digital twins, i.e., (i) various types of attacks on digital twins, and (ii) defensive strategies to thwart such attacks. As attackers are constantly improving their attack techniques, we could adopt the following techniques to limit the damage in the event of a compromise on digital twins.

- We need intelligence-driven solutions (such as data analytics and threat intelligence) to collect information on attackers' behaviors. Threat hunting can then use this intelligence to identify the forensic artifacts, such as Indicators Of Compromise (IOCs). Recent reports from incident response, for instance, sharing IOCs, i.e., knowledge of suspicious activities and artifacts with the community might also be useful. During the recovery process, switching off the affected device or service can limit the damage caused by cyberattacks.
- We should implement provenance-aware blockchain-based solutions to audit digital twins, i.e., track and trace the accountable entity which changed the simulation setup parameters or state data. Provenance data can help reconstruct the process chain to detect and localize the faulty node in the system.
- We need to develop a fault-tolerant system. Instead of disconnecting the entire system, we should enable a graceful degradation during which the system enters a fail-safe state and maintains an adequate level of control of the physical process when undesirable incidents occur. Depending on the priority or severity of the incident, contingency planning is required to identify the root cause of operational disruption or the fraudulent middleman. With short-term remedies and small-scale fixes, we must minimize the probability of incidents and their recovery time.

References

[1] M. Dietz and G. Pernul, "Unleashing the digital twin's potential for ICS security," *IEEE Security Privacy*, vol. 18, no. 4, pp. 20–27, 2020.

[2] M. Eckhart and A. Ekelhart, *Digital Twins for Cyber-Physical Systems Security: State of the Art and Outlook*. Cham: Springer International Publishing, 2019, pp. 383–412.

[3] S. Suhail, R. Hussain, R. Jurdak, A. Oracevic, K. Salah, C. S. Hong, and R. Matulevičius, "Blockchain-based digital twins: Research trends, issues, and future challenges," *ACM Comput. Surv.*, feb 2022, just Accepted. [Online]. Available: https://doi.org/10.1145/3517189

[4] S. Suhail, R. Jurdak, and R. Matulevicius, "Towards trusted and intelligent cyber-physical systems: A security-by-design approach," *CoRR*, vol. abs/2105.08886, 2021. [Online]. Available: https://arxiv.org/abs/2105.08886

[5] M. Eckhart and A. Ekelhart, "Towards security-aware virtual environments for digital twins," in *Proceedings of the 4th ACM Workshop on Cyber-Physical System Security*, ser. CPSS '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 61–72.

[6] S. Suhail, R. Hussain, R. Jurdak, and C. S. Hong, "Trustworthy digital twins in the industrial internet of things with blockchain," *IEEE Internet Computing*, pp. 1–10, 2021. [Online]. Available: http://dx.doi.org/10.1109/MIC.2021.3059320

[7] R. Langner, "Stuxnet: Dissecting a cyberwarfare weapon," *IEEE Security Privacy*, vol. 9, no. 3, pp. 49–51, 2011.

[8] R. Minerva and N. Crespi, "Digital twins: Properties, software frameworks, and application scenarios," *IT Professional*, vol. 23, no. 1, pp. 51–55, 2021.

[9] A. Rasheed, O. San, and T. Kvamsdal, "Digital twin: Values, challenges and enablers from a modeling perspective," *IEEE Access*, vol. 8, pp. 21 980–22 012, 2020.

[10] M. Groshev, C. Guimarães, J. Martín-Pérez, and A. de la Oliva, "Toward intelligent cyber-physical systems: Digital twin meets artificial intelligence," *IEEE Communications Magazine*, vol. 59, no. 8, pp. 14–20, 2021.

[11] S. Suhail, S. U. R. Malik, R. Jurdak, and R. Matulevičius, "Towards situational aware cyber-physical systems: A security-enhancing use case of blockchain-based digital twins," 2022, arXiv: 2201.07765. [Online]. Available: https://arxiv.org/abs/2201.07765.

[12] M. Dietz, B. Putz, and G. Pernul, "A distributed ledger approach to digital twin secure data sharing," in *IFIP Annual Conference on Data and Applications Security and Privacy*, vol. 11559. Cham: Springer, 2019, pp. 281–300.

[13] S. Huang, G. Wang, Y. Yan, and X. Fang, "Blockchain-based data management for digital twin of product," *Journal of Manufacturing Systems*, vol. 54, pp. 361–371, 2020.

[14] W. Shen, T. Hu, C. Zhang, and S. Ma, "Secure sharing of big digital twin data for smart manufacturing based on blockchain," *Journal of Manufacturing Systems*, vol. 61, pp. 338–350, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0278612521002004

[15] M. Vielberth, M. Glas, M. Dietz, S. Karagiannis, E. Magkos, and G. Pernul, "A digital twin-based cyber range for SOC analysts," in *Data and Applications Security and Privacy XXXV*, K. Barker and K. Ghazinour, Eds. Cham: Springer International Publishing, 2021, pp. 293–311.

[16] S. Zeadally, E. Adi, Z. Baig, and I. A. Khan, "Harnessing artificial intelligence capabilities to improve cybersecurity," *IEEE Access*, vol. 8, pp. 23 817–23 837, 2020.