A Framework For Applying Digital Twins To Support Incident Response

Sabah Suhail[1] [0000-0002-2464-9790], Mubashar Iqbal[2] [0000-0003-0543-613X], Kieran McLaughlin[1] [0000-0002-1299-2364], Brian Lee[3] [0000-0002-8475-4074], and Babar Imtiaz[3] [0000-0003-4775-9033]

[1] Queen's University Belfast, UK
{s.suhail, kieran.mclaughlin}@qub.ac.uk
[2] University of Tartu, Estonia
mubashar.iqbal@ut.ee
[3] Technological University of the Shannon, Ireland
{brian.Lee, Muhammad.BabarImtiaz}@tus.ie

Abstract. The convergence of information technology (IT) and operational technology (OT) has made manufacturing industries an attractive target for cyberattacks, ranging from industrial espionage to sabotage. Existing security tools, operating alone, are not capable enough to manage cybersecurity operations effectively. Digital twin (DT), as a security-enhancing enabler, can support complementary security measures alongside existing incident response (IR) solutions. DTs have been proposed in different IR phases; however, a comprehensive solution covering the IR lifecycle has yet to be addressed. This paper presents a DT-based IR solution to guide plant operators in modeling security-enhancing DTs for manufacturing industries. Moreover, the DT-based IR solution integrates existing security tools to ensure the effective safeguarding of critical assets and prompt response to cyber incidents. With an automotive assembly line as a cyber-physical production system (CPPS) use case, we examine the applicability of a DT-based IR solution.

Keywords: Digital Twin (DT) · Incident Response (IR) · Cyber-Physical Production System (CPPS) · Industrial Control System (ICS) · Cybersecurity.

1 Introduction

In recent years there have been a number of reported cases of cyberattacks on manufacturing industries, where there have been direct effects on the operations of physical processes such as production lines. For instance, Emotet [17] was found to have compromised multiple automotive manufacturers, including Toyota in Japan, which exemplifies the need for securing an industrial control system (ICS) against potential attacks. Similarly, ransomware attacks on Renault-Nissan [15], Saint-Gobain [1], and Norsk Hydro [2] are among other notable cyberattacks on manufacturing ecosystems.

To address the aforesaid cyberattacks in a timely and effective manner, it is essential to adopt solutions that can analyze data and state inconsistencies without causing damage to the physical assets [12]. One such solution is digital twins (DTs). DTs are virtual replicas of their physical counterparts that must *(i)* exhibit sufficient fidelity in terms of attributes and services, *(ii)* maintain a continuous synchronized feedback loop, and *(iii)* provide objective-specific actionable insights while covering entire lifecycle [4, 13, 31]. Recently, DTs have gained significant attention as security-enhancing enablers to support various phases of incident response (IR) [14]. The existing works emphasizing DT for securing cyber-physical systems (CPSs) cover different IR phases [5]. Nevertheless, a comprehensive approach that covers the entire IR lifecycle is yet to be addressed. Furthermore, the integration of existing IR solutions or tools like security information and event management (SIEM) along with DT-based security solutions is not given due attention.

This work aims to answer the following research questions (RQs) to address IR phases in manufacturing industries using DTs. *(RQ1)* How can critical assets, processes, or services be identified as suitable candidates for developing a security-enhancing DT? *(RQ2)* Given resource constraints, including communication, computation, and storage costs, which DT operation modes can be leveraged to optimize cybersecurity value? *(RQ3)* How can DT-based IR solutions complement existing security solutions? These questions lead to our main contributions:

- To guide plant operators, such as in manufacturing industries, we present a framework for DT-based IR lifecycle phases, including design-and-engineering and operation-and-maintenance.
- As an example of a cyber-physical production system (CPPS), we explore an automotive assembly line as a use case and consider IR lifecycle phases as the underlying security objective. We investigate viable IR solutions applying different DT operation modes (simulation and replication) and consider existing IR solutions, such as SIEM, using assessment criteria, including cost (operational and maintenance), damage (financial or operational), and recovery time.

The paper is organized as follows. Section 2 provides background information and existing works to establish the context of the work. Section 3 showcases an automotive assembly line use case featuring an attack scenario. Section 4 discusses the modeling requirements of a DT-based IR solution, covering how to model and utilize DT-based IR. Finally, Section 5 presents the concluding remarks of the paper and future research directions.

2 Background and Related Work

This section provides an overview of essential topics necessary for comprehending the outcomes of this research, e.g., Section 2.1 discusses the DT operations modes, Section 2.2 provides an overview of IR phases, and Section 2.3 presents the related works concerning DT-based IR solutions in manufacturing industries.

2.1 DT operation modes

In manufacturing industries, introducing sustainable-by-design or security-by-design at the initial (design) phase can help to *(i)* lower defects such as parameter configuration, *(ii)* support agile product lifecycle management methods, and *(iii)* analyze assets in a virtual environment before real-world operations. A DT can investigate data inconsistencies through a simulation mode that does not connect to the physical asset or a replication mode that involves continuous mapping of the physical and twin environments. Being reproducible and repeatable, simulation mode can enable a plant operator to test and debug design artifacts by (re)running DT instances until the optimal operating conditions are met for a given production process [9, 35]. Replication mode allows plant operators to identify data inconsistencies by continuously synchronizing the state and data of the physical system with its twin counterparts [11].

2.2 Incident Response (IR)

IR refers to the structured approach taken by organizations to manage and address security incidents, cyber threats, or disruptions to minimize damage, reduce recovery time, and mitigate the impact on the business [30]. The IR lifecycle comprises the following phases.
Preparation: Creating an IR plan, establishing a response team, defining roles and responsibilities, and conducting training drills.
Identification: Identifying a security incident through various monitoring systems or reports to understand the nature and scope of the problem.
Containment and Eradication: Taking immediate actions to remove the threat and eliminate the incident's root cause.
Recovery: Restoring affected systems, data, and operations to normal functionality while ensuring the security and integrity of the environment.
Lessons learned: Conducting a post-incident review to analyze the incident and improve IR procedures for the future.

2.3 Related work

DTs as security-enhancing enablers have been proposed to enhance the performance of IR phases, such as staff training, anomaly detection, and system testing [13]. For example, [8, 37] utilize DTs as cyber ranges to provide hands-on cyber skills (preparation phase). Works including [8, 9, 12, 24, 33] utilize DTs for security posture testing and anomaly-based intrusion detection (identification phase). DTs can help to minimize the impact on live systems (containment phase) as discussed in [7, 16, 24, 26]. For example, virtual segmentation and isolation, access controls simulation, firewall configuration, simulation, and testing of different containment strategies. Using DTs, organizations can accelerate the restoration of affected systems (eradication and recovery phases) while ensuring security, minimizing downtime, and validating the recovery procedures in a controlled environment before implementing them on live systems [7, 16]. Additionally, DTs can refine IR strategies based on insights gathered from the recovery process (lesson learned phase), ultimately contributing to a more effective and secure incident resolution [7].

The existing works on DT-based IR focus on either one or two IR phases and lack a coherent view across the entire IR lifecycle. Furthermore, the existing works overlook the integration of security-enhancing DTs with established security tools/solutions. Therefore, the question is how to model DT-based IR and integrate them with existing security measures. In this regard, our work bridges the research gap by highlighting how to model DT-based IR (Section 4.1), which DT modes must be leveraged, and how to integrate DT-based IR solutions with existing security solutions (Section 4.2).

3 Use Case: An Automotive Assembly Line

This section discusses the physical infrastructure in a typical automotive assembly line (Section 3.1) and an attack scenario targeting the welding process in the body shop (Section 3.2).

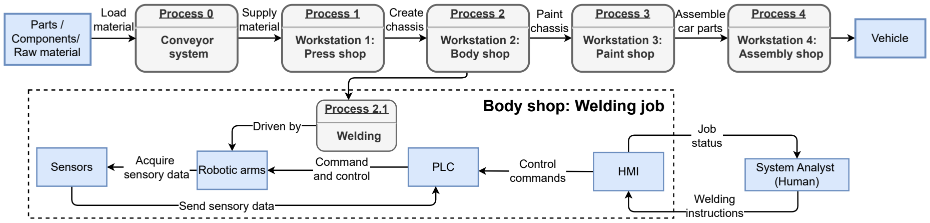

Fig. 1. Data Flow Diagram illustrating automotive manufacturing stages emphasizing the welding process in the body shop (represented by dashed lines).

3.1 Infrastructure in automotive assembly line

We present an automotive assembly line as a CPPS use case where we consider four automotive production stages (as Fig. 1 shows), namely the press shop, the body shop, the paint shop, and the final assembly shop [19]. Each production stage can be represented by a workstation operated on a conveyor system that moves the vehicle chassis or parts between workstations. Each workstation performs a specific set of tasks on the vehicle. To do so, workstations are equipped with physical assets, including *(i)* control systems such as programmable logic controllers (PLCs), to manage the sequence of operations in the assembly line, human-machine interfaces (HMIs), to provide interfaces for operators to interact with and control the assembly such as robotic arms (to perform tasks including welding, painting, and handling heavy components), feeders and hoppers (to store and supply raw materials, components, or parts to the assembly line), automated guided vehicles (AGVs), to transport heavy or bulk materials, and *(iii)* data acquisition sources (sensors, actuators, vision systems) to measure or monitor operations. Note that we have not considered auxiliary processes such as inventory management or manual inspection. Under such settings, we can achieve two objectives: *(a)* protect critical assets or processes by separating them with less protected/less critical (sub)zones and *(b)* identify physical assets or processes that require twin counterparts to track data inconsistencies.

To simplify the illustration of complex automotive production processes, we focus on the welding process within a body shop as an example (as Fig. 1 shows). The rationale for selecting this process is that the welding process demands precision and accuracy to ensure the alignment of various components in the assembly phase. More specifically, any flaws in the welding process can directly impact the structural integrity, durability, and safety of the manufactured vehicle [10]. In the body shop, the components, including frame sections, panels, or other structural parts, that require welding to create the chassis structure are obtained from the press shop. The HMI communicates voltage, current, wire feed speed, and arc length to a PLC, which then controls the welding robotic arms to weld at predefined joints or designated areas on the components to assemble the chassis [29]. Sensors, such as tactile, temperature, and arc, affixed to robotic arms operate based on predefined configurations for efficient and automated monitoring of the welding process [29]. Note that the welding process may differ based on a specific model, make, individualized parts, or welding approach; however, the fundamental steps remain constant across various welding methods.

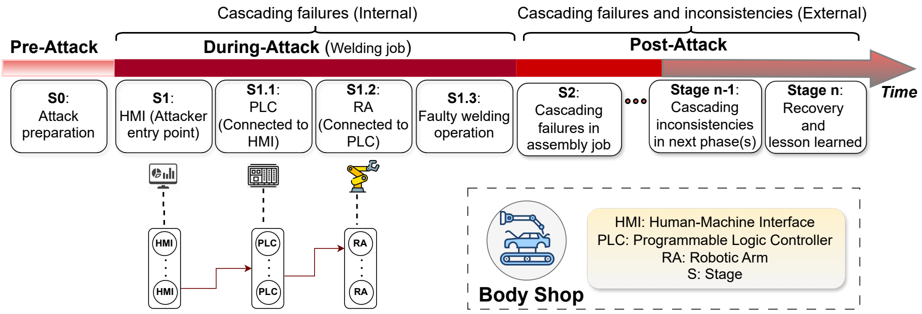
Cascading failures (Internal)
Cascading failures and inconsistencies (External)
Pre-Attack
During-Attack (Welding job)
Post-Attack
Time
S0: Attack preparation
S1: HMI (Attacker entry point)
S1.1: PLC (Connected to HMI)
S1.2: RA (Connected to PLC)
S1.3: Faulty welding operation
S2: Cascading failures in assembly job
Stage n-1: Cascading inconsistencies in next phase(s)
Stage n: Recovery and lesson learned
HMI
HMI
PLC
PLC
RA
RA
Body Shop
HMI: Human-Machine Interface
PLC: Programmable Logic Controller
RA: Robotic Arm
S: Stage

Fig. 2. Illustration of an attack scenario targeting the welding process, highlighting the attack stages and potential internal and external cascading failures.

3.2 Attack scenario

Visualization tampering at the HMI could be an attractive target for attackers, as evident from Stuxnet and Industroyer incidents [20]. Attackers may conceal critical information, show erroneous data to mislead operators or tamper with data [4]. Following the cyber kill chain (CKC), we have outlined the progression of a cyber attack by classifying the attacker's actions targeting the welding process into three stages: pre-attack, during-attack, and post-attack, as illustrated in Fig. 2. During the pre-attack stage, attackers collect intelligence regarding system infrastructure and operational processes, for example through reconnaissance or phishing techniques, or with the collaboration of insiders. During the attack stage, attackers may exploit vulnerabilities in the HMI to gain entry and manipulate welding parameters. Unauthorized access to the HMI due to default passwords or unpatched software could allow attackers to inject malware into the interconnected systems. The adulterated parameters are passed to the PLC that commands and controls welding robots (see Fig. 1). Improper voltage and current levels can lead to overheating or weak welds. For instance, overheating due to higher voltage or current levels may lead to burn-through or distortion of the chassis, whereas inadequate penetration into the base material due to lower voltage or current levels may create weak joints, thereby jeopardizing the chassis's strength and stability. Thus, during the post-attack phase, such attacks eventually lead to production delays in the automotive production phases (such as the assembly phase in the body shop) because detecting and rectifying poor-quality welds require additional time and resources. Furthermore, inappropriate settings of welding equipment may result in malfunctions of the welding gun or power source, leading to equipment damage, costly repairs, and production downtime. Thus, an attack on one shop or workstation can lead to external cascading failures or inconsistencies in the subsequent automotive production shops (as depicted in the extreme right of Fig. 2). We will revisit this attack scenario in Section 4.2, where we will explore effective solutions for resolving such attacks.

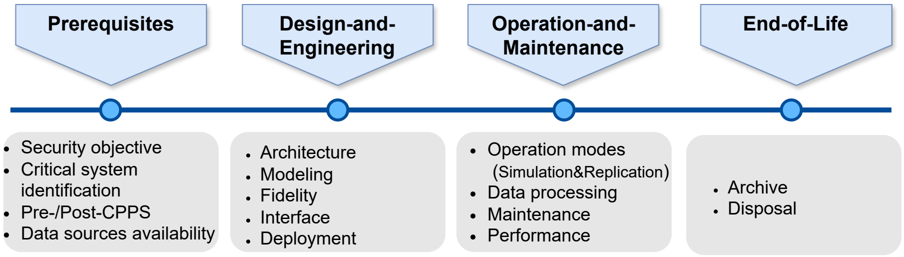

Fig. 3. Illustrating the key steps within security-enhancing DT lifecycle phases.

4 DT-based IR Framework

In this section, we propose a DT-based IR framework (Section 4.1), primarily focusing on the design and operation phases. The framework of DT-based IR provides abstract representations crucial for understanding various DT concepts, illustrating the system's structure and behavior. Following [36], our framework is developed using unified modeling language (UML) class diagrams. Then we analyze the derived IR-based DTs models (Section 4.2) based on an attack scenario (discussed in Section 3.2), followed by a discussion of insights (Section 4.3).

4.1 Designing and developing DT-based IR solutions

We propose that the lifecycle of a DT in tandem with its physical counterpart comprises four main phases: *(i)* prerequisites, *(ii)* design-and-engineering, *(iii)* operation-and-maintenance, and *(iv)* end-of-life (as Fig. 3 shows).

First, the security objective(s) must be determined, including intrusion prevention and detection, cyber deception, digital forensics, testing, and training [14], reflecting the subsequent CPPS-DT lifecycle phases. The reason for determining the security objective at the prerequisite phase is to configure the security-enhancing DT design parameters following the available resources.

Second, depending on the CPPS infrastructure, plant operators may choose to create DTs of a network, system, application, or a combination of these for an underlying use case and security objectives. Due to feasibility limitations in creating a DT for the entire infrastructure [13] as well as the challenge in identifying which CPS sub-systems can be mapped to a DT, we propose that the DT modeling process can be refined by: *(i)* breaking down the infrastructure into sub-components, *(ii)* constructing a data flow diagram to pinpoint critical assets, processes, or services, and *(iii)* selecting which of these critical elements should be modeled as a DT. This step addresses RQ1.

Third, DTs rely on data sources obtained from ongoing physical processes, assets state data, real-time sensory data, historical data, or other sources [35]. Data availability is crucial because DTs rely on the data to accurately represent the physical system they simulate [4,35]. Therefore, it is essential to identify the availability of data sources subject to the pre and post-existence of a specific CPPS instance [34]. For instance, if a CPPS is operational and requires security assessment during its operation-and-maintenance phase, modeling a DT from scratch (i.e., design phase) can utilize historical and real-time data. Conversely, if the CPPS is not yet in existence, DT modeling can draw upon alternative data sources, such as historical data from similar production or manufacturing processes. Upon establishing the prerequisites, the DT can be developed and managed based on CPPS-DT lifecycle phases.

To enforce security by design, DT modeling must cover the IR lifecycle so that the DT instances can be instantiated accurately, exhibiting the functionalities required during an incident. In the following, we mainly discuss the design-and-engineering (Fig. 4) and operation-and-maintenance (Fig. 5) lifecycle phases of DT-based IR.

Design-and-engineering phase Fig. 4 illustrates the DT design-and-engineering phase (marked with dotted lines) and its integration with IR phases (marked with dashed lines), including preparation and identification. For plant operators, Fig. 4 refers to design DT-based IR solutions. *First*, it provides the foundational

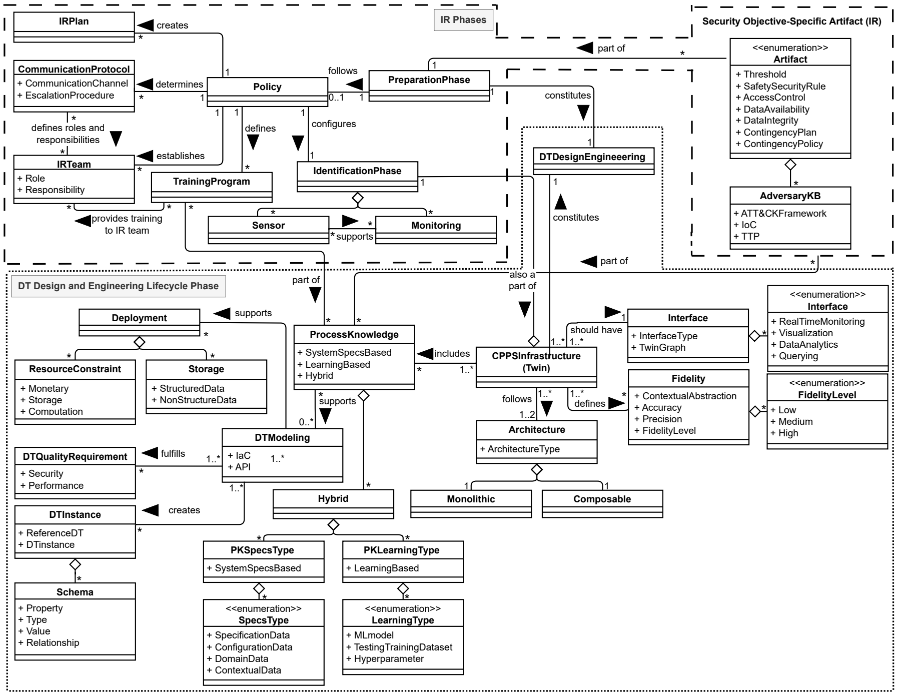

Fig. 4. Design-and-engineering phase of DT-based IR. For illustration purposes, the dashed and dotted lines delineate the IR and DT lifecycle phases, respectively.

building blocks necessary for modeling security-enhancing DTs. For instance, selecting a DT architecture, integrating process knowledge (PK), accessing DT artifacts via an interface, and determining sufficient fidelity depending on the underlying security objective. Moreover, it outlines the factors to consider when deploying DT-based IR solutions, such as resource constraints, storage requirements, and quality requirements. *Second*, it integrates IR phases into the DT model, ensuring that DT instances accurately reflect the functionalities needed during the preparation phase, such as establishing the plan for IR team training based on organizational policies, and the identification phase, such as monitoring CPPS twin infrastructure based on predefined thresholds and rules.

The DT-based IR preparation phase comprises plans, protocols, and resources heavily relying on organizational policies (refer to the upper left dashed part of Fig. 4). The policies are derived from security objective artifacts such as data security, access control, and existing adversary knowledge bases (KBs), e.g., tactics, techniques, and procedures (TTPs) and indicators of compromise (IoCs) (refer to the rightmost dashed part of Fig. 4). The preparation phase requirements, such as instantiating simulated training programs, can be translated as DT PK, i.e., data flows, behaviors, connections, and components within a physical system or asset. PK could be learning-based, specification-based, or hybrid [14]. PK is further subjected to fidelity, i.e., the degree of resemblance between the twin and its physical counterpart. Measuring the fidelity levels, such as low, medium, or high depends on DT operation modes, i.e., simulation and replication (as Fig. 5 shows). For instance, a high-fidelity DT model is required for replication mode. Furthermore, high-fidelity DT correlates with higher DT accuracy and precision, albeit at higher deployment costs. Moreover, high-fidelity DTs lead to information leakage [31]. Therefore, a trade-off is required between generalization and contextualization.

The CPPS twinned architecture can be defined as monolithic or composable (refer to the lower right dotted part of Fig. 4). Monolithic refers to asset, system, or process-level twin, whereas composable refers to component-level twin [23,28]. The diversity among these DT types stems from their granularity levels, ranging from singular representations of parts or components (such as a temperature sensor, which can be termed as a component twin) to more comprehensive twins that encompass entire systems (such as the overall welding process, which can be termed as a process twin). For DT-based IR, composable DTs, i.e., component twins, offer a flexible and modular strategy, thereby facilitating attack localization by analyzing the component-level DT instance. The deployment of DT is constrained by resource availability, including communication, computation, and storage costs (refer to the upper right dotted part of Fig. 4). For instance, long-term data storage may utilize cloud services as a scalable and cost-effective solution. The DT data and models can be stored as structured or non-structured data. The deployed DT models must exhibit security and performance requirements, including confidentiality, integrity, availability, access control, and scalability. A reference DT (blueprint) can be instantiated to generate DT instances with schema having properties, types, values, and relationships. For instance, in manufacturing custom-built luxury vehicles with distinctive chassis designs, DT instances can be tailored to accommodate individualized or customer-specific welding specifications while maintaining consistency with fundamental welding operations outlined in the reference DT blueprint.

To automate the deployment of system configurations, infrastructure as code (IaC) can be used to (re)generate DT instances. The DT modeling can use application programming interfaces (APIs) to support services specific to simulation tools used for DT modeling. For instance, Microsoft Azure DT provides Azure IoT Hub–a cloud-hosted service that facilitates communication between applications and devices [22]. CPPS twins can be accessed through an interface for querying and visualization; additionally, it supports twin graphs (based on DT models) that combine real-time monitoring and data analytics [22] (refer to the lower rightmost dotted part of Fig. 4). During the identification phase, the DT operation mode supports monitoring CPPS twin infrastructure to identify deviations in the replication mode during the digital-physical mapping based on the predefined thresholds or safety and security rules defined in policies or analyze DT instances through reconfigurable and reproducible simulation environments.

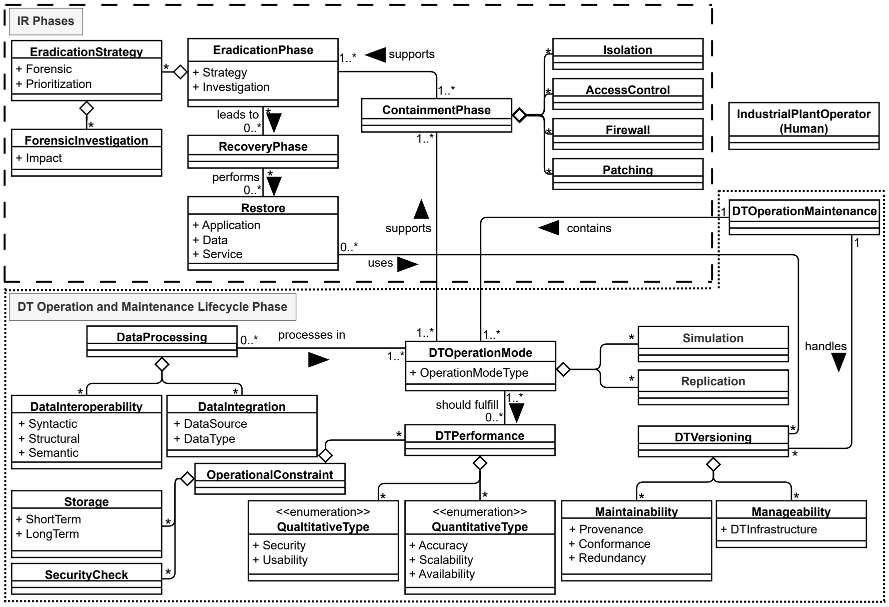

Fig. 5. Operation-and-maintenance phase of DT-based IR. For illustration purposes, the dashed and dotted lines delineate the IR and DT lifecycle phases, respectively.

Operation-and-maintenance phase Fig. 5 illustrates the DT operation-and-maintenance phase (marked with dotted lines) and its integration with IR phases (marked with dashed lines), including containment, eradication, and recovery. For plant operators, Fig. 5 serves as a reference to operate DT-based IR solutions. *First*, it outlines DT operation modes (simulation and replication). Moreover, it provides the operational requirements of DTs, such as DT performance requirements essential for the trustworthiness of decisions made by the DT. *Second*, it integrates IR phases into the DT model during the containment and eradication phases to leverage DT modes. For instance, using DT simulation to analyze the propagation of attacks and the effectiveness of various containment strategies. Additionally, DT versioning can aid in restoring compromised applications, data, or services in the recovery phase.

The DT-based IR containment phase limits the impact of an incident by using containment strategies, including isolation, access control, firewall, or patching (refer to the rightmost dashed part of Fig. 5). The eradication phase comprises forensic and prioritization, focusing on removing threats. Forensic investigation involves examining the affected systems, logs, and other digital artifacts to understand the incident or exploited vulnerabilities. Based on the urgency of securing critical assets or systems, prioritization defines which actions need immediate attention. DT simulation and replication modes can be leveraged to carry out containment and eradication phases (refer to the upper rightmost dotted part of Fig. 5). For example, with the simulation mode, plant operators can analyze the asset, process, or service under attack, in isolation from the physical system, to understand the attack scenario and its ramifications, or to localize the compromised node. The data processing during DT simulation and replication modes must comply with data integration, i.e., combining data from heterogeneous sources or formats into a unified format, and interoperability, i.e., enabling systems or devices to exchange and interpret data seamlessly. The recovery phase involves restoring the normal operations of applications, data, or services from backups, repairing compromised systems, and implementing security patches. To restore the system, DT instances managed and maintained through versioning (refer to the lower rightmost dotted part of Fig. 5), can be utilized during the backup process to ensure data integrity and system reliability. Maintenance of DT instances involves keeping track of provenance data, compliance with any relevant standards, and backup that serves as a fail-safe in case of primary system failure. The performance of a DT must adhere to quantitative and qualitative aspects. Quantitative assessments involve metrics, such as accuracy, scalability, and availability, whereas qualitative assessments involve security and usability. These factors may significantly effect the system's recovery time and the trustworthiness of decisions made by the DT.

4.2 Utilizing DT-based IR solutions during system attacks

Based on the attack scenario (Section 3.2), in the following, we explore a viable IR solution that leverages *(i)* an existing IR solution (such as SIEM) to address RQ3, *(ii)* DT operation modes (simulation and/or replication) to address RQ2, and/or *(iii)* a combination of both solutions while minimizing cost, reducing response time, and mitigating potential damage. The selection of a cybersecurity solution depends on various factors, including a company's size, security posture, financial constraints, regulatory obligations, or current industry adoption trends in the automotive manufacturing industry. Note that the proposed solution focuses on synergizing DT-based solutions with established IR solutions such as SIEM, considered a key cyber defense platform in the industry [18]. Table 1 summarizes our findings regarding integrating SIEM with DT-based IR solutions across cyber attack stages.

We consider the effectiveness of the IR solution based on two factors, i.e., cost and response time. The following discussion is based on the attack scenario on the welding process (Section 3.2). Various costs associated with cybersecurity solutions can be defined as *(i)* initial costs (covering implementation and deployment, licensing), *(ii)* operational costs (utilities, consumables, resource usage such as hardware, servers, or cloud infrastructure), *(iii)* maintenance and manageability costs (scaling and upgrading, debugging, hardware maintenance, software licensing renewals), and *(iv)* others such as training cost (programs and resources for educating employees) [25]. For instance, in the case of a DT-based IR solution, the initial cost may include DT modeling that involves creating multi-fidelity DT instances (ranging from low to high), integration and interoperability with physical system infrastructure, i.e., data sources, and DT infrastructure (computation, communication, and storage) and DT performance requirements. The operational cost of DT may include real-time monitoring tools, analytics platforms, and other allocated resources such as storage. The maintenance cost includes regular updates and security patches. The training cost includes personnel training and resources to educate employees about the DT-based IR solution. Considering initial DT modeling and deployment costs have already been incurred, we focus on operational and maintenance costs. While we have theoretically addressed the cost factor, it is worth noting [6] formulates the problem of deriving cost-effective DT for security tests as a 0–1 non-linear programming problem.

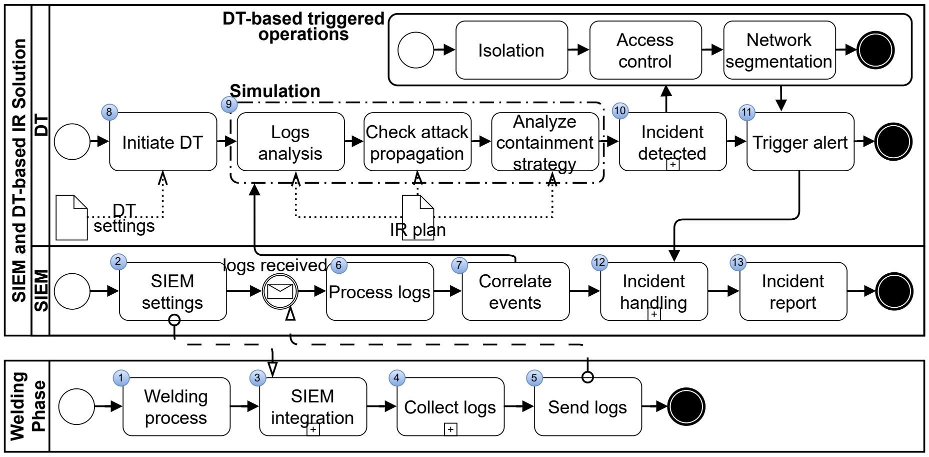

Fig. 6. Illustrating the response to a cyberattack (depicted in Fig. 2) during the containment and eradication phases.

Preparation *Effective solutions: SIEM and DT-based simulation and replication modes.*
Justification: During the preparation phase, the manufacturing industry must be equipped with the necessary strategies, including developing IR plans, establishing IR teams, conducting regular training programs, deploying tools, and reviewing policies to mitigate the impact of potential security incidents effectively (refer to the preparation phase in the dashed part of Fig. 4). For instance, the DT replication mode continuously synchronizes with physical welding processes to promptly detect data inconsistencies. Furthermore, DT simulation instances can be instantiated to facilitate training and awareness programs, enabling teams to practice response strategies in a risk-free, controlled environment (see Fig. 4). These training initiatives ensure that IR teams are equipped with the necessary skills to effectively respond to incidents, thereby enhancing their preparedness. SIEM can offer real-time monitoring and analysis of security events by collecting,

correlating, and logging data for compliance or auditing purposes. For instance, during the pre-attack stage, login attempts to HMI may be correlated with irregular network traffic patterns or unauthorized alterations to welding parameters. Additionally, it can be used to log access data for the HMI controlling welding robotic arms via PLCs.

Identification *Effective solution: SIEM.*
Justification: During the identification phase, SIEM can assist plant operators by analyzing network traffic and system logs, providing insights into potential insider threats or compromised credentials, to identifying suspicious activities or deviations from benign behavior (identification phase in the dashed part of Fig. 4). For example, adulteration of welding parameters, unauthorized access attempts, or unusual login patterns providing insights about insider threats or compromised credentials can be flagged as anomalies. Moreover, integrating SIEM with third-party threat intelligence feeds, including threat signatures, profiles, and IoCs, can identify patterns, anomalies, and potential threats.
Assessment: Operational and maintenance costs are primarily associated with SIEM. An additional cost, such as licensing, may be incurred if third-party solutions are used for advanced real-time threat recognition. With SIEM-based solutions, there might be possibilities of burnout due to the lack of context needed by plant operators to understand alerts or false positives [21], particularly if the attacker adulterates parameters at random intervals or within normal variability. Therefore, to minimize damage and reduce response time, the solution must adapt to fine-tuning threshold levels to reduce false positives and use context-aware detection techniques to distinguish malicious and benign activities.

Containment and Eradication *Effective solution: SIEM and DT-based simulation mode.*
Justification: We assume the welding process has been integrated into the SIEM system. Following our DT-based IR approach, plant operators may utilize a combination of SIEM and a DT-based simulation mode, as Fig. 6 shows. First, the SIEM collects system logs, network traffic, and user activity (see Fig. 6 step 4) from the physical infrastructure, including HMIs, PLCs, and robotic arms operating in the body shop (see Fig. 1). Second, to allow an in-depth incident investigation, DT simulation mode is integrated with the SIEM system using its log analysis and event correlation capabilities. DT instance(s) are instantiated (see Fig. 6 step 8) based on modeling requirements depicted in Fig. 4. To localize the compromised asset, DT simulation mode can facilitate scenario analysis by modeling different attack scenarios and containment strategies to assess *(a)* how the attack might propagate, *(b)* which containment strategies would be most effective, *(c)* what could be the potential impact of containment actions on the physical system, and *(d)* how to systematically eradicate the root cause of the security incident from the affected system (see Fig. 5). For example, simulating containment strategies (see Fig. 6 step 9) involves *(i)* isolation methods (quarantine affected welding process from other workstations or manufacturing networks)

Table 1. Applying an integration of SIEM with DT-based IR solutions across cyber attack stages (depicted in Fig. 2)

Cyberattack stages	**IR phase**	**Response strategy**	**Details**
Pre-attack	Preparation	Simulation, Replication, SIEM	Simulated training environment for plant operators' training and awareness; Replication mode for continuous digital-physical mapping to detect data inconsistencies; SIEM logs and correlates HMI access.
	Identification	SIEM	SIEM detects suspicious activities like parameter tampering, unauthorized access, or unusual logins by analyzing network traffic and system logs.
During attack	Containment and Eradication	Simulation, SIEM	Simulating containment strategies; SIEM monitors network traffic, logs for unauthorized access to the quarantined welding process, and tracks HMI within the zone for compliance and anomaly detection.
Post-attack	Recovery	Simulation, Replication	Simulated DT instance of the welding process validates the integrity of the restored welding process by comparing current welding parameters against benign operational parameters; DT replication mode synchronizes with the recovered welding process.
	Lesson learned	Simulation, SIEM	Simulated training environment for plant operators to prepare against cyberattacks; SIEM conducts postmortem analysis and ensures compliance through reports and dashboards.

to prevent external cascading inconsistencies or failures across other shops, *(ii)* access controls (disable or restrict access to compromised HMI), or *(iii)* network segmentation (to segregate the body shop's network traffic). For instance, SIEM can support simulation mode in applying isolation methods by monitoring network traffic and logs for unauthorized access attempts to the quarantined welding process. Additionally, they can track HMI within the quarantine zone to ensure compliance with isolation measures and detect anomalies or security breaches. Thus, SIEM can correlate event data from HMI logs with DT simulation mode to reconstruct the timeline of events necessary for identifying the root cause. Once the incident of a compromised HMI is detected (see Fig. 6 step 10), the SIEM should assist in executing incident handling (see Fig. 6 step 12), including remediation, patch management, removal of malware and backdoors, and validation of actions taken to eradicate the root cause.

Assessment: Through real-time monitoring and analysis, SIEM systems can collate multiple events and logs from various assets and systems in the welding shop. Such data ingestion and logged events from multiple sources can support DT simulation mode. During simulation mode, isolation methods can be applied to a range of simulated attack scenarios that can leverage SIEM data to detect security breaches or anomalies. Consequently, integrating SIEM and DT simulation modes ensures the timely identification and triaging of security incidents.

Recovery *Effective solution: DT simulation and replication modes.*

Justification: The plant operators can follow a two-step approach during recovery surveillance, i.e., post-attack stage. First, test patches for a vulnerable asset, i.e., HMI or suspected service in a simulated environment, and monitor the restored system against any anomalies or residual threats. The simulated DT instance of the welding process can act as a sandbox that validates the integrity of the restored welding process by comparing current welding parameters against benign operational parameters. This strategy can help plant operators test, verify, and validate compromised assets before replicating them on the live physical system. Second, upon synchronizing the physical asset or service with DT replication mode, continue monitoring the recovered systems, i.e., HMI through the replication mode to identify the re-occurrence of abnormal behavior or vulnerabilities. The replication mode can be a backup to restore the system to a known stable state in the face of an attack or system failure, thereby minimizing downtime and ensuring resilience.

Assessment: Both DT modes, particularly replication mode, have associated operational and maintenance costs; however, they offer long-term benefits in terms of operational resilience and risk mitigation.

Lesson learned *Effective solutions: DT-based simulation mode and SIEM.*

Justification: The plant operators can harness outcomes from both SIEM and DTs as both solutions have been proposed independently or in combination during the earlier IR phases. For instance, the obtained results can be used to *(i)* carry out post-incident analysis and documentation, *(ii)* provide actionable recommendations and remediation plan outlining steps to address identified weaknesses or gaps (such as weak or default credentials of HMIs) in security measures to prevent similar incidents, and *(iii)* summarize key takeaways and insights gained from IR phases and share findings with other manufacturing industry peers to enforce collaborative learning against evolving threats. Such IR reports might be of additional support to gain insights into future incidents in manufacturing industries. DT simulation mode could be realized as a comprehensive gamification platform (as suggested by [34]). With DTs, a simulated and controlled hands-on learning/training environment can help to train operations personnel and prepare for the next threat wave. For instance, DT instance(s) can present diverse challenges based on varying objectives and scenarios from the welding process. SIEM complements the simulation mode by conducting postmortem analysis and ensuring adherence to regulatory requirements and industry standards through the generation of compliance reports, audit trails, and security dashboards. Moreover, training exercises can vary from *table-top* discussion-based sessions (such as workshops on identifying phishing emails and password security of assets) aimed at familiarizing employees with cybersecurity awareness and best practices, to *live-play* role-based exercises (such as simulated cyberattacks on welding systems followed by IR drills) aimed at enhancing the skill set of security professionals.

Assessment: The investigation results from both security solutions have the potential to identify malicious incidents in a timely manner, minimizing response time and decreasing damage.

4.3 Takeaways

Existing IR solutions like SIEM and underlying technologies such as artificial intelligence/machine learning (AI/ML) or adversary KBs may not suffice independently for effective cybersecurity operations [24]. Consequently, DT as a security-enhancing enabler can serve as complementary security measures alongside existing IR solutions, enhancing the overall security infrastructure and maximizing cybersecurity value. Through the proposed work, such a combined security solution, integrating DTs with conventional IR solutions enables organizations to navigate resource constraints effectively while ensuring the efficacy of their security strategies, thereby enhancing threat visibility, response agility, and defense strategies to fortify the overall security posture.

The capital and operational expenditure on DT-based solutions covers various aspects, including implementation and deployment, operational, maintenance, and training costs. Furthermore, the cost factor in DT modeling is influenced by security objectives such as anomaly detection or cyber deception, which are determined by the required degree of fidelity. For instance, cyber deception demands higher fidelity compared to anomaly detection scenarios [32]. Despite the theoretical nature of the discussion in Section 4.2, the insights on the cost of DT-based solutions offer valuable information that can inform decision-making processes, strategic planning, and resource allocation in practical applications.

In this work, regarding the modeling and deployment of DTs, we follow the composable DTs with component twins that allow a flexible, modular, and agile approach [23,28]. Component twins can provide granularity, thereby supporting the identification and localization of attacks. For example, within the manufacturing context, instead of making system-level DT for the entire body shop, which could be complex and resource-intensive, a more practical approach would be to create individual DTs for components such as PLCs, HMIs, sensors, etc., and then aggregate these instances to form a comprehensive DT. On the one hand, replication mode may incur higher costs as it requires *(a)* modeling DT instance(s) with sufficient fidelity and *(b)* time-dependent synchronization consistency with the physical asset. On the other hand, the continuous feedback loops through replication mode can offer varying degrees of autonomy, ranging from human-in-the-loop to human-on-the-loop to fully autonomous systems. Moreover, it can support system resilience by providing a backup to restore system states. However, given that the security of a DT is a critical concern [4,31], replication modes must be carefully executed. In the manufacturing context, the replication services, in the case of critical processes, ensure availability, reliability, and fault tolerance. Therefore, replication mode must be applied for key operations such as welding processes or automated inspection to minimize the risk of failure and enhance system resilience rather than auxiliary processes such as inventory management. Following a pragmatic approach, a triage decision must be based on attack severity, potential impact, and available resources.

We selected the welding process in the body shop of the automotive manufacturing industry to demonstrate a theoretical application of the DT-based IR framework in a realistic practical scenario. Given that the welding process demands precision and accuracy, the DT-based IR framework as compared to existing security solutions, can provide a viable approach to prevent, detect, and mitigate cyber incidents in a timely and effective manner.

The proposed DT-based IR framework can be extended to other use cases, specifically those involving fixed physical environments characterized by machinery and equipment that remain static. Examples include electronics manufacturing, consumer goods manufacturing, and food manufacturing, which face cybersecurity risks due to their reliance on interconnected systems [3]. It is worth noting that, for the given CPPS use case, the DT-based IR design parameters may vary depending on the security-enhancing use case of the DT and available resources.

5 Conclusion and Future Work

Cyberattacks on manufacturing industries have demonstrated severe consequences for business and economic sectors. This work discusses the significance of integrating security-enhancing DTs for effective IR in the manufacturing industries. First, we present a framework for DT-based IR to guide plant operators in manufacturing industries, mainly focusing on the design and operation lifecycle phases. Second, we analyze an attack scenario within the automotive assembly line to illustrate the practical feasibility of a DT-based IR solution. We discuss viable approaches, such as DT operation modes (simulation and replication), and existing security tools, such as SIEM, while considering various assessment criteria, including cost, damage, and recovery time.

We anticipate that Figs. 4 and 5 can be a reference for plant operators to design and operate DT-based IR solutions. Compared to traditional security solutions, such as SIEM, the DT-based IR framework potentially offers a more effective solution due to its close coupling with attributes of the CPS. This is attributed to its capability to reproduce attack scenarios for in-depth analysis through simulation mode, maintain continuous synchronization of feedback loops with their physical counterparts for real-time updates of system state and behavior through replication mode, and employ component-level DTs to identify cascading inconsistencies (dotted line parts in Fig. 4 and Fig. 5). Furthermore, the proposed approach can guide plant operators in utilizing DT-based IR solutions alongside existing security tools when the system is under attack.

When designing security-enhancing DTs, our approach primarily builds on [14] where we add value by integrating DT architecture (monolithic or composable), data processing services (integration, interoperability), DT performance (quantitative and qualitative), operational constraints (storage and security checks), and DT versioning (maintainability and manageability). This extension to security-enhancing DTs provides plant operators with a thorough consideration of the requirements for designing and developing DT-based IR.

As the proposed DT-based IR framework is yet to be implemented, the next steps in this research anticipate proof-of-concept experimentation through a small-scale testbed to explore practical challenges in integration. Attack and response experiments will be conducted to investigate response times required for the DT-based approach to carry out successful IR functions. Subsequent stages of experimentation and development will focus on tuning the approach for accuracy, for a lab-based use case, aiming to provide broader lessons that can generalized across other use cases. Similarly, comparing DT-based IR solutions with other established security solutions or tools to understand how they can complement security-enhancing DTs in cybersecurity operations is also necessary. Such assessment should include evaluating metrics such as cost, damage, and recovery time. Furthermore, we acknowledge the value of broadening our scope with additional resources, such as those from the DT Consortium [27], to fully consider various requirements for DT design and operation.

Acknowledgements

This work was supported by the North-South Research Programme, which is delivered by Higher Education Authority (HEA) on behalf of the Department of Further and Higher Education, Research, Innovation and Science (DFHERIS) and the Shared Island Unit at the Department of the Taoiseach.

References

1. Saint-gobain press release (2017), https://www.saint-gobain.com/sites/saint-gobain.com/files/03-07-2017_cp_va.pdf
2. Cyber-attack on hydro (2019), https://www.hydro.com/en/media/on-the-agenda/cyber-attack/
3. Alamri, A.H.: Dragos industrial ransomware analysis: Q1 2024 (2024), Dragos, Available at: https://www.dragos.com/blog/dragos-industrial-ransomware-analysis-q1-2024/, *(accessed: May 05, 2024)*
4. Alcaraz, C., Lopez, J.: Digital twin: A comprehensive survey of security threats. IEEE Commun.Surv.&Tuts. **24**(3), 1475–1503 (2022). https://doi.org/10.1109/COMST.2022.3171465
5. Allison, D., Smith, P., Mclaughlin, K.: Digital twin-enhanced incident response for cyber-physical systems. In: Proceedings of the 18th International Conference on Availability, Reliability and Security. ARES '23, Association for Computing Machinery (2023). https://doi.org/10.1145/3600160.3600195
6. Bitton, R., et al.: Deriving a cost-effective digital twin of an ics to facilitate security evaluation. In: Computer Security. pp. 533–554. Springer International Publishing, Cham (2018). https://doi.org/https://doi.org/10.1007/978-3-319-99073-6_26
7. Bécue, A., Maia, E., Feeken, L., Borchers, P., Praça, I.: A new concept of digital twin supporting optimization and resilience of factories of the future. Applied Sciences **10**(13) (2020). https://doi.org/10.3390/app10134482

8. Bécue, A., et al.: Cyberfactory#1 — securing the industry 4.0 with cyber-ranges and digital twins. In: 2018 14th IEEE International Workshop on Factory Communication Systems. pp. 1–4 (2018). https://doi.org/10.1109/WFCS.2018.8402377
9. Dietz, M., Vielberth, M., Pernul, G.: Integrating digital twin security simulations in the security operations center. In: Proceedings of the 15th International Conference ARES (2020). https://doi.org/10.1145/3407023.3407039
10. Doshi, A., Smith, R.T., Thomas, B.H., Bouras, C.: Use of projector based augmented reality to improve manual spot-welding precision and accuracy for automotive manufacturing. International Journal of Advanced Manufacturing Technology **89**(5-8), 1279–1293 (2017). https://doi.org/10.1007/s00170-016-9164-5
11. Eckhart, M., Ekelhart, A.: A specification-based state replication approach for digital twins. In: Proceedings of the 2018 Workshop CPS-SPC. p. 36–47 (2018). https://doi.org/10.1145/3264888.3264892
12. Eckhart, M., Ekelhart, A.: Towards security-aware virtual environments for digital twins. In: Proceedings of the 4th ACM Workshop CPSS. p. 61–72 (2018). https://doi.org/10.1145/3198458.3198464
13. Eckhart, M., Ekelhart, A.: Digital Twins for Cyber-Physical Systems Security: State of the Art and Outlook, pp. 383–412. Springer International Publishing, Cham (2019). https://doi.org/10.1007/978-3-030-25312-7_14
14. Eckhart, M., et al.: Security-enhancing digital twins: Characteristics, indicators, and future perspectives. IEEE Security & Privacy **21**(6), 64–75 (2023). https://doi.org/10.1109/MSEC.2023.3271225
15. Eisenstein, P.A.: European car plants halted by wannacry ransomware attack (2017), https://www.nbcnews.com/business/autos/european-car-plants-halted-wannacry-ransomware-attack-n759496
16. Gehrmann, C., Gunnarsson, M.: A digital twin-based industrial automation and control system security architecture. IEEE Trans. Ind. Inform. **16**(1), 669–680 (2020). https://doi.org/10.1109/TII.2019.2938885
17. Hanrahan, J.: Suspected conti ransomware activity in the auto manufacturing sector (2022), https://www.dragos.com/blog/industry-news/suspected-conti-ransomware-activity-in-the-auto-manufacturing-sector/
18. Kinyua, J., Awuah, L.: AI/ML in security orchestration, automation and response: Future research directions. Intelligent Automation & Soft Computing **28**(2) (2021)
19. Konstantinidis, F.K., Mouroutsos, S.G., Gasteratos, A.: The role of machine vision in industry 4.0: An automotive manufacturing perspective. In: 2021 IEEE International Conference on Imaging Systems and Techniques. pp. 1–6 (2021). https://doi.org/10.1109/IST50367.2021.9651453
20. Lee, R.M., Assante, M., Conway, T.: Crashoverride: Analysis of the threat to electric grid operations. Dragos Inc., March (2017)
21. López Velásquez, J.M., Martínez Monterrubio, S.M., Sánchez Crespo, L.E., Garcia Rosado, D.: Systematic review of siem technology: Siem-sc birth. International Journal of Information Security **22**(3), 691–711 (2023). https://doi.org/https://doi.org/10.1007/s10207-022-00657-9
22. Microsoft: What is azure digital twins?, https://docs.microsoft.com/en-us/azure/digital-twins/overview, (accessed: September 05, 2023)
23. Minerva, R., Crespi, N.: Digital twins: Properties, software frameworks, and application scenarios. IT Professional **23**(1), 51–55 (2021). https://doi.org/10.1109/MITP.2020.2982896
24. Mohsin, A., Janicke, H., Nepal, S., Holmes, D.: Digital twins and the future of their use enabling shift left and shift right cybersecurity operations. In: 2023 5th IEEE

International Conference on Trust, Privacy and Security in Intelligent Systems and Applications. pp. 277–286. IEEE Computer Society (nov 2023). https://doi.org/10.1109/TPS-ISA58951.2023.00042

25. Oettl, F., Eckart, L., Schilp, J.: Cost estimation approach of a digital twin implementation in industry. Procedia CIRP **118**, 318–323 (2023). https://doi.org/https://doi.org/10.1016/j.procir.2023.06.055, 16th CIRP Conference on Intelligent Computation in Manufacturing Engineering
26. Redelinghuys, A., Basson, A.H., Kruger, K.: A six-layer architecture for the digital twin: a manufacturing case study implementation. Journal of Intelligent Manufacturing **31**(6), 1383–1402 (2020). https://doi.org/10.1007/s10845-019-01516-6
27. Pieter van Schalkwyk, Sean Whiteley, M.G.: Digital twin capabilities periodic table (2024), Available at: https://www.digitaltwinconsortium.org/wp-content/uploads/sites/3/2024/04/DTC_Capabilities-Periodic-Table-User-Guide-v1.1.pdf, *(accessed: April, 2024)*
28. van Schalkwyk, P., Isaacs, D.: Achieving Scale Through Composable and Lean Digital Twins, pp. 153–180. Springer International Publishing, Cham (2023). https://doi.org/10.1007/978-3-031-21343-4_6
29. Shah, D.N.: Automatic welding and soldering machine using plc in automobile application. In: 2021 Asian Conference on Innovation in Technology. pp. 1–6 (2021). https://doi.org/10.1109/ASIANCON51346.2021.9544774
30. Shinde, N., Kulkarni, P.: Cyber incident response and planning: a flexible approach. Computer Fraud&Security **2021**(1), 14–19 (2021). https://doi.org/https://doi.org/10.1016/S1361-3723(21)00009-9
31. Suhail, S., Iqbal, M., Jurdak, R.: The perils of leveraging evil digital twins as security-enhancing enablers. Commun. ACM **67**(1), 39–42 (dec 2023). https://doi.org/10.1145/3631539
32. Suhail, S., Iqbal, M., McLaughlin, K.: Digital twin-driven deception platform: Vision and way forward. IEEE Internet Computing pp. 1–9 (2024). https://doi.org/10.1109/MIC.2024.3406188
33. Suhail, S., et al.: Towards situational aware cyber-physical systems: A security-enhancing use case of blockchain-based digital twins. Computers in Industry **141**, 103699 (2022), doi: https://doi.org/10.1016/j.compind.2022.103699
34. Suhail, S., et al.: ENIGMA: An explainable digital twin security solution for cyber-physical systems. Computers in Industry **151**, 103961 (2023), doi: https://doi.org/10.1016/j.compind.2023.103961
35. Tao, F., Zhang, H., Zhang, C.: Advancements and challenges of digital twins in industry. Nature Computational Science **4**(3), 169–177 (2024). https://doi.org/https://doi.org/10.1038/s43588-024-00603-w
36. Yue, T., Arcaini, P., Ali, S.: Understanding digital twins for cyber-physical systems: A conceptual model. pp. 54–71. Springer International Publishing, Cham (2021)
37. Zhou, H., Li, M., Sun, Y., Tian, Z., Yun, L.: Digital twin-based cyber range for industrial internet of things. IEEE Consumer Electronics Magazine pp. 1–11 (2022). https://doi.org/10.1109/MCE.2022.3203202