Paper:

Digital Twin: A Comprehensive Survey of Security Threats

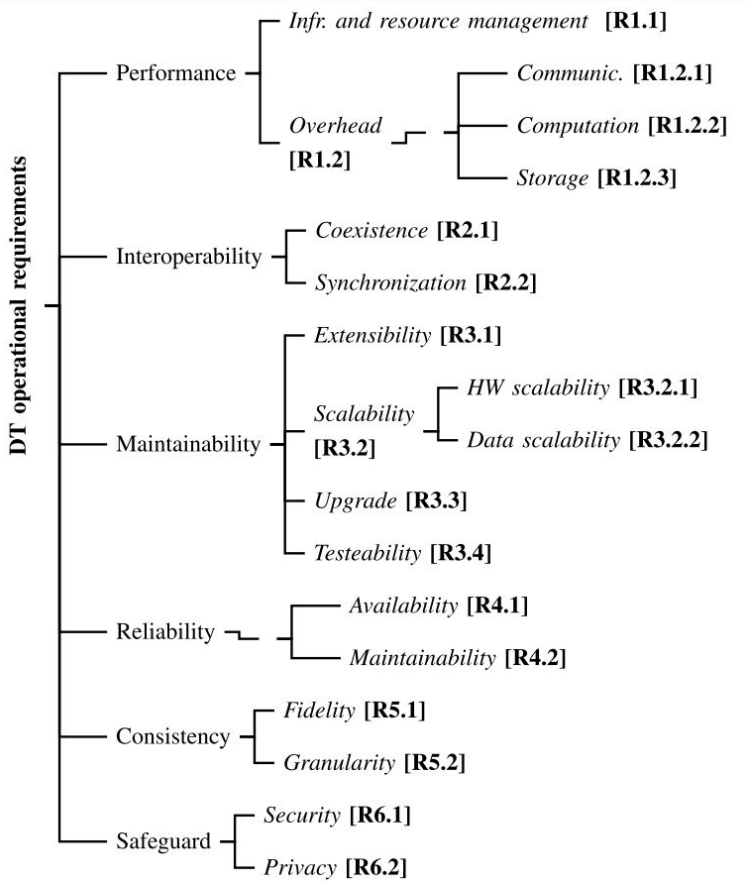
Cristina Alcaraz, Javier Lopez

Presented by: Oscar Llerena

Content

IV. SECURITY THREATS IN THE DIGITAL TWIN

- Threats on **Availability**, **Integrity**, and **Confidentiality** of **data resources**. Also threats to the **privacy** of **Entities** and asset **Location**.
- It is important to address **how security and privacy threats affect operation requirements of DTs** (Operational **Performance** & Reduction of Complexities, **Interoperability** Between Assets & Layers, **Maintainability** of Digital Assets, **Reliability** of Assets & Data, **Consistency** in Reasoning and Representation, **Safeguarding** Virtual Resources, Operations and Data).
- **Security analysis** of a DT must **consider** the four **functionality layers** (Data **Dissemination** & Acquisition, Data **Management** & Synchronization, Data **Modeling**, and Data **Visualization**).

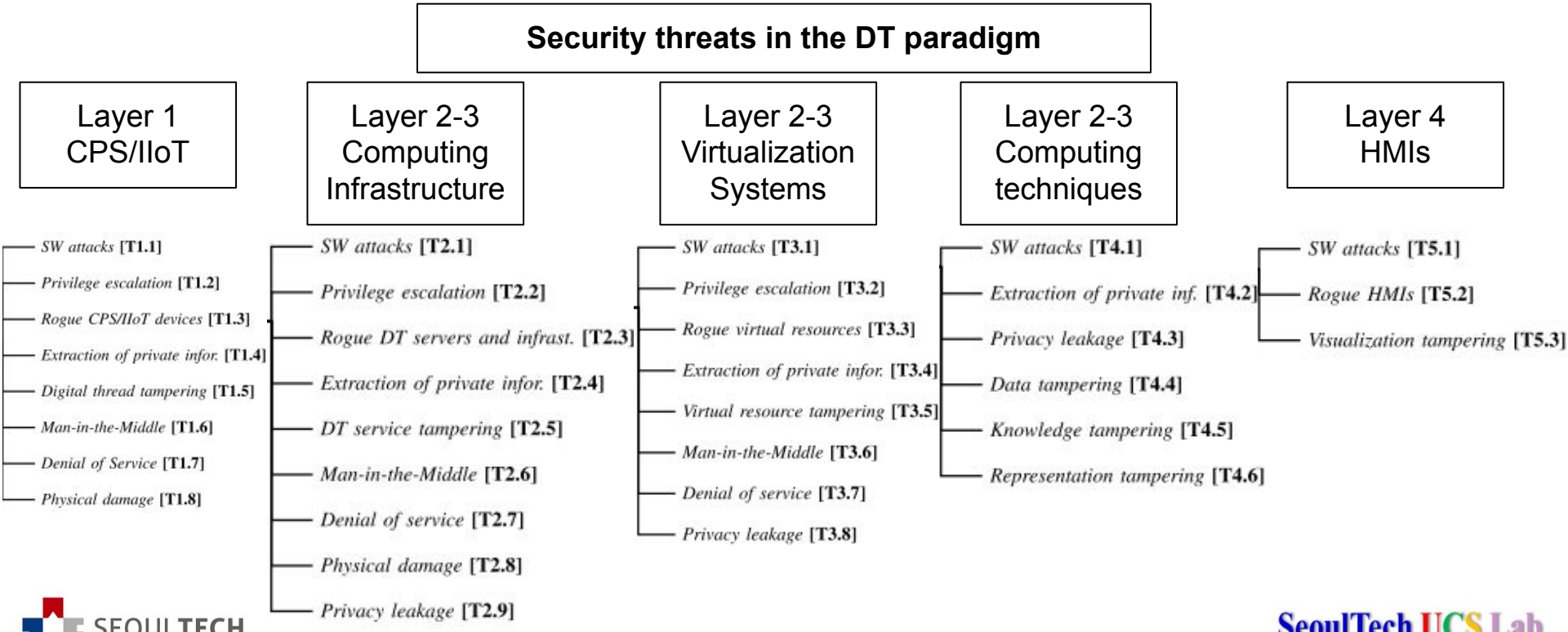


IV. SECURITY THREATS IN THE DIGITAL TWIN

- DTs mostly **rely on digital assets for data processing** (modeling, algorithms, virtualization, and networks).
- In this paper, **two types of attack** surfaces are considered: ***digital*** and ***physical***.
- The **digital attack** comprises all the explorations associated not only with **software** (poor code, default settings, etc.) but also with all the components offering **resources** for computation (networks and information systems).
- The **physical attack** comprises all those security threats associated with **access to endpoints** such CPS/IIoT nodes, communication infrastructures and facilities.
- Attackers may compromise the DT with physical attack surface (L1 to L2-4).
- Also, physical assets may also be at risk when DT is attacked (L4-2 to L1).
- Example: ***Adversaries can penetrate industrial control systems (ICS) and, once inside, can search the location of the DT in order to compromise, learn about the system, extend their technical capabilities, and access the critical system through the DT to exfiltrate information or destroy its resources*** [130].
- The threat would correspond to a typical Advanced Persistent Threat (APT) such as Stuxnet (2009), BlackEnergy (2015-2016), ExPetr (2017) or GreyEnergy (2018) [131].

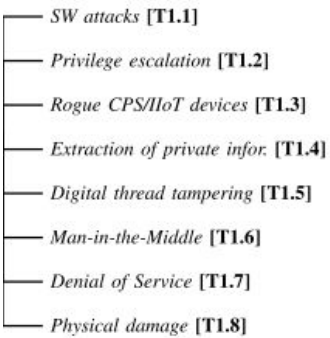
IV. SECURITY THREATS IN THE DIGITAL TWIN

- Figure 4 shows a classification of the different threats that have been identified in the DT paradigm (where [Tx.y] represents the functional layer x and the y-th threat in that layer).



IV. SECURITY THREATS IN THE DIGITAL TWIN

Layer 1
CPS/IIoT



A. Threats at Layer 1

- **[T.1.1] SW attacks:** Operational Technologies (OT) devices rely on proprietary or **third-party SW**, which might present **code bugs**, which may lead to threats such **reverse engineering** [48], **buffer overflows** [134], computing **manipulations** [135], or **alteration** on the **device behavior** [136].
- [134] offers overview on common vulnerabilities and exposures (CVEs) with the support of databases such as ICS-CERT, MITRE and NIST's national vuln. database (NVD).
- [137], manufacturing industries rely on older versions of OSs such as Windows XP (public source code in 2020 [138]).
- [139] and [133] review specific attacks on CPS and IIoT, pointing out malware as a potential SW attack weapon (e.g., PLCBlaster worm [140], Dragonfly, Stuxnet, BlackEnergy 3, LockerGoga, REvil, Industroyer, etc. [141], including rootkits for controllers [142]).
- (i) cause **significant overheads** on the device; (ii) trigger **interoperability** and **maintainability issues** (whether local or remote); (iii) **disturb the synchronization** performance and/or cause consistency issues; and (iv) **cause security** concerns.

IV. SECURITY THREATS IN THE DIGITAL TWIN

Layer 1
CPS/IIoT

- *SW attacks* **[T1.1]**
- *Privilege escalation* **[T1.2]**
- *Rogue CPS/IIoT devices* **[T1.3]**
- *Extraction of private infor.* **[T1.4]**
- *Digital thread tampering* **[T1.5]**
- *Man-in-the-Middle* **[T1.6]**
- *Denial of Service* **[T1.7]**
- *Physical damage* **[T1.8]**

A. Threats at Layer 1

- **[T1.2] Privilege escalation:** Adversaries access the OT domains and aim to **escalate privileges** (admin's permissions) by **exploiting flaws in the authentication and authorization mechanisms**.
- Example: [135] details the influence of **Triton malware** interacting with specific **Triconex controllers** by **exploiting zero-day vulnerabilities** (CVE2018-7522 and CVE-2018-8872). Attackers were able to **escalate privileges on the controller** to gain access to Triconex's memory and **execute arbitrary** codes.
- Similar threats can also occur in DT-based scenarios. Attackers with full rights to access industrial domains could **disconnect Layer 1 nodes**, **change configurations**, generate **false values** or **manipulate network traffic** [47], which would also lead to **significant deviations at Layers 2-4**. If, in addition, these DTs are designed for detection, such as [10], [36], [144], their **systems could handle invalid information**, producing **false positive or negative rates** when comparing the input and output values of the two spaces.

IV. SECURITY THREATS IN THE DIGITAL TWIN

Layer 1
CPS/IIoT

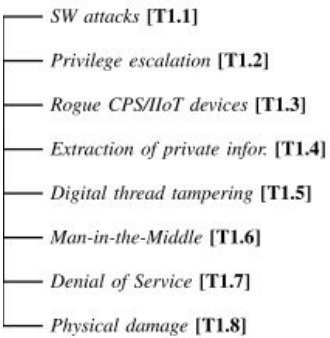
- *SW attacks* [T1.1]
- *Privilege escalation* [T1.2]
- *Rogue CPS/IIoT devices* [T1.3]
- *Extraction of private infor.* [T1.4]
- *Digital thread tampering* [T1.5]
- *Man-in-the-Middle* [T1.6]
- *Denial of Service* [T1.7]
- *Physical damage* [T1.8]

A. Threats at Layer 1

- **[T1.3] Rogue CPS/IIoT devices:** Insiders with full rights to access OT domains may **deploy, clone and replace IT/OT devices**, or **maliciously update SW** components to **take control of the physical space**, and consequently interact or **impact with the digital space**.
- Example: [36] shows how to **configure the flags of a PLC** to hack a hydraulic system.
- Through these rogue devices, adversaries may consequently **lead other attack actions**, such as **man-in-the-middle (MitM)** actions, **disrupt control tasks**, insert a **backdoor for redirection of critical traffic**, or fool the DT itself with **fake output values.**
- Moreover, malicious manufacturers might, for example, **insert compromised parts in CPS/IIoT devices** to achieve specific purposes (e.g., create information leaks, cause malfunctions, or alter the integrity of assets) that can impact not only the normal operation of the system and any DT involved, but also an organization's reputation [145].

IV. SECURITY THREATS IN THE DIGITAL TWIN

Layer 1
CPS/IIoT



A. Threats at Layer 1

- **[T1.4] Extraction of private information:** such as credentials or security parameters shared with the DT. Then, attackers gain access to the DT from the physical space or conduct multiple MitM attacks between both spaces. Another way to extract legitimate information would be through traffic analysis [132]. Adversaries interfere the traffic, eavesdrop the data consumed or produced by the physical space and the virtual plane, or analyze the network flows to map (e.g., by looking at the source and destination IP) and locate the server that hosts the DT and later corrupt the physical space through malicious C&C instructions.
- **[T1.5] Digital thread tampering:** [146] shows the attacker's ability to modify the data exchanged (e.g., synchronization or C&C values) between the physical and digital space of a DT. When insiders take advantage of access privileges to OT domains and freely manage devices (inject malware, produce misconfigurations in the monitoring tasks, or desynchronize the digital space with respect to the physical space) without being supervised.

IV. SECURITY THREATS IN THE DIGITAL TWIN

Layer 1
CPS/IIoT

- *SW attacks* **[T1.1]**
- *Privilege escalation* **[T1.2]**
- *Rogue CPS/IIoT devices* **[T1.3]**
- *Extraction of private infor.* **[T1.4]**
- *Digital thread tampering* **[T1.5]**
- *Man-in-the-Middle* **[T1.6]**
- *Denial of Service* **[T1.7]**
- *Physical damage* **[T1.8]**

A. Threats at Layer 1

- **[T1.6] Man-in-the-middle:** Intruders with the ability to insert rogue devices, and consequently to interfere with communication channels. They launch routing attacks to play with the DT traffic from the physical space [133], [135], [148] and: (i) **create deviations or routing loops** that could **deteriorate the QoS** [149] or the **maintenance processes**; (ii) **inject false data** [30]; (iii) **modify control packets** [10], [36]; or (iv) trace the sequence of traffic flow.
- **[T1.7] Denial of service:** Adversaries **exhaust resources** of IIoT/CPS devices to **limit automation operations** in the **physical space and** the **simulation operations** in the **digital space**. This depletion in CPS/IIoT ecosystems can be carried out from the TCP/IP stack to cause **jamming at the physical layer** [146], [150], **inject malware** at the **application layer**, or provoke **on-the-path attacks** at the **network layer**. Typical DoS attacks in CPS/IIoT routing [135], [148], [151] are **flooding** [47], **replay** [30], **blackhole**, **sinkhole** [151], **wormhole** [152] or **selective forwarding** packets. [150] provides a review of threats in 5G communications. On the other hand, DDoS attack prepare an army of CPS/IIoT botnets. The Mirai attack [153] is an IoT-based botnet example against DNS provider.

IV. SECURITY THREATS IN THE DIGITAL TWIN

Layer 2-3 Computing Infrastructure

- *SW attacks* **[T2.1]**
- *Privilege escalation* **[T2.2]**
- *Rogue DT servers and infrast.* **[T2.3]**
- *Extraction of private infor.* **[T2.4]**
- *DT service tampering* **[T2.5]**
- *Man-in-the-Middle* **[T2.6]**
- *Denial of service* **[T2.7]**
- *Physical damage* **[T2.8]**
- *Privacy leakage* **[T2.9]**

B. Threats at Layers 2-3 - Computing Infrastructure

- **[T2.1] SW attacks:** DT servers are mainly based on systems that compute specific DT services which depend on SW components such databases, ML models, applications and firmware. [156] concludes that OS for cloud-based environments (Windows & Linux) present security vulnerabilities related to authentication, authorization, accounting and privacy. [17] shows malware infections between elements of a DT and between DTs. Also, cloud platforms lack anti-malware measures [157]. Any infected cloud server could complicate cross-space synchronization processes or disable essential functions of the DT.
- **[T2.2] Privilege escalation:** Adversaries who break into the system and try to reach the DT aim to escalate privileges in order to take over the host system. Deficiencies in authentication mechanisms, access control policies, lack of segregation, lack of knowledge or disinterest in the security of the system. [146] and [158] state that cloud-based resources may not be sufficiently isolated in industrial contexts, causing availability, integrity and confidentiality problems.

IV. SECURITY THREATS IN THE DIGITAL TWIN

Layer 2-3 Computing Infrastructure

- *SW attacks* **[T2.1]**
- *Privilege escalation* **[T2.2]**
- *Rogue DT servers and infrast.* **[T2.3]**
- *Extraction of private infor.* **[T2.4]**
- *DT service tampering* **[T2.5]**
- *Man-in-the-Middle* **[T2.6]**
- *Denial of service* **[T2.7]**
- *Physical damage* **[T2.8]**
- *Privacy leakage* **[T2.9]**

B. Threats at Layers 2-3 - Computing Infrastructure

- **[T2.3] Rogue DT servers and infrastructures:** Insiders with full rights to deploy DT servers and related infrastructures may clone and replace components to add malicious servers. This means that data replicates of the physical world may be managed by fake servers, and insiders may take control of the digital thread shared by both worlds.
- **[T2.4] Extraction of private information:** Data privacy is one of the biggest security issues in the DT paradigm [160], mainly because the goal is to protect the intellectual property contained in their servers. Adversaries with access to compromised servers or related infrastructures may extract private information, such as services, dynamics data, configurations, states or security credentials. With this information, they may exfiltrate information for cyber espionage, or identify the main vulnerabilities in the DT (including zero-days) to improve attack techniques. This method of gaining access to sensitive information can even help attackers carry out potential attacks that may result in APTs. The results may range from stealthy manipulations in the DT services to lateral movements between attack surfaces within the computing infrastructure itself.

IV. SECURITY THREATS IN THE DIGITAL TWIN

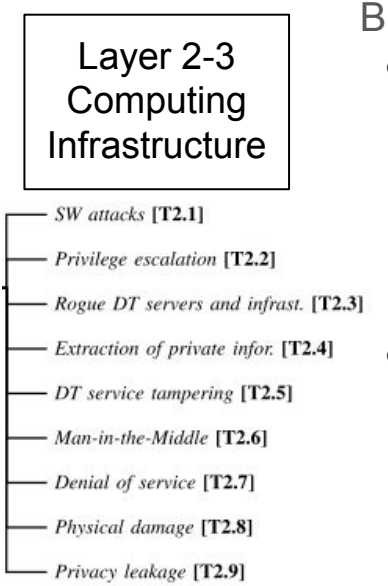
Layer 2-3 Computing Infrastructure

- *SW attacks* **[T2.1]**
- *Privilege escalation* **[T2.2]**
- *Rogue DT servers and infrast.* **[T2.3]**
- *Extraction of private infor.* **[T2.4]**
- *DT service tampering* **[T2.5]**
- *Man-in-the-Middle* **[T2.6]**
- *Denial of service* **[T2.7]**
- *Physical damage* **[T2.8]**
- *Privacy leakage* **[T2.9]**

B. Threats at Layers 2-3 - Computing Infrastructure

- **[T2.5] DT service tampering:** If servers hosting DTs are compromised, either by privilege escalation or abuse, it is very possible that adversaries can manipulate the services of the DT itself.
- **[T2.6] Man-in-the-middle:** MitMs are typical threats in network infrastructures and, in that case, DTs are systems whose logic may be dispersed throughout an entire computing infrastructure. Malicious servers (in the cloud, in the fog and at the edge) can act as MitMs [159] through which DT information flows can pass. Likewise, these MitM servers that execute part of the DT logic can also (i) cause deviations in the knowledge that the DT itself processes; (ii) alter or overflow the databases that the DT manages; and (iii) change the final representation that the DT computes to the end user.
- **[T2.7] Denial of service:** DDoS attacks may also occur in applications that rely on computing infrastructures. However, the extent of the threat may not be so dramatic in edge-assisted contexts. For instance, powerful computing services related to the intelligence and representation of the digital assets (at Layers 2-3) could be deployed within the cloud/fog, and the rest of the services distributed at the edge.

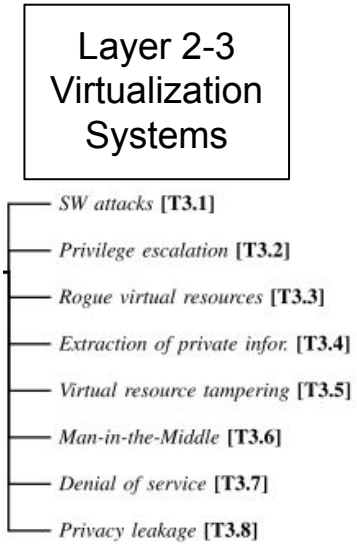
IV. SECURITY THREATS IN THE DIGITAL TWIN



B. Threats at Layers 2-3 - Computing Infrastructure

- **[T2.8] Physical damage:** Operational domains are generally closed systems that require the attacker to be close to the server or its infrastructure. Insiders would therefore be the only ones who could execute this attack as long as they were able to escalate privileges within the facility and gain access to the target, which still constitutes a threat that implicitly causes a DoS and affects the correct functioning of a DT or one of its sub-parts [17].
- **[T2.9] Privacy leakage:** In addition to data privacy, other privacy risks may arise, especially when computing infrastructures adapt intelligence algorithms. Edge paradigms (including cloud and fog) are systems composed of elements capable of computing and storing large volumes of private data. Malicious entities may steal sensitive information (causing confidentiality issues related to [T2.4]) or derive (encrypted) production, logistics or marketing plans, which would undoubtedly put intellectual property at risk [17]. Apart from this, location privacy is also relevant at this point. Hyperconnected servers (e.g., at the edge-cloud [167]) that contain all the DT's logic may be clear targets for adversaries, whose initial purpose may be to trace their locations in order to lead subsequent attacks.

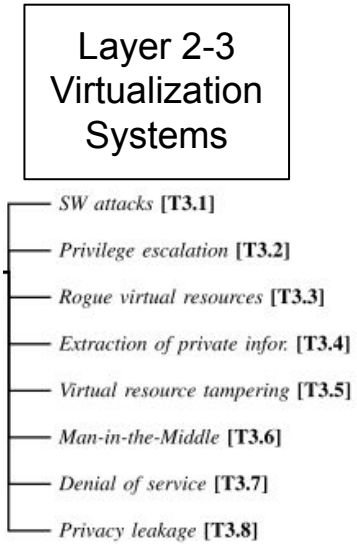
IV. SECURITY THREATS IN THE DIGITAL TWIN



B. Threats at Layers 2-3 - Virtualization systems

- **[T3.1] SW attacks:** Both VMs containing the digital assets, and monitoring and management tools of virtual resources (hypervisors) are SW-based systems that present multiple vulnerabilities. [168] analyzes the security breaches of hypervisors according to real attacks. Through these breaches, adversaries carry out subsequent attacks not only on the VM but also on the host where the VM is running (example: malware penetration into the kernel [169], illicit memory writing, buffer overflow, illegal code execution, memory and information leak, etc. [171]). Although SDN networks benefit the defense against DDoS attacks, the efficiency of packet processing in the communication space still depends on SW components. Compromised SDN controllers result in inefficient data processing, significant overheads or losses of information.
- **[T3.2] Privilege escalation:** VMs, containers and hypervisors managing the DT's logic may present SW vulnerabilities which are attractive for adversaries capable of escalating privileges within the virtualization system [157]. Once inside the system, they can navigate between the virtual resources and launch multiple attacks (e.g., exfiltration, manipulations, overflows or passive analysis).

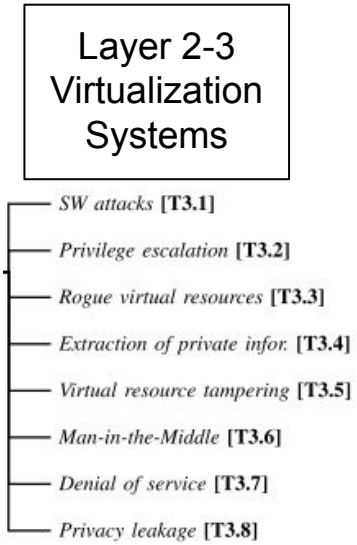
IV. SECURITY THREATS IN THE DIGITAL TWIN



B. Threats at Layers 2-3 - Virtualization systems

- **[T3.3] Rogue virtual resources:** Insiders with the ability to escalate or abuse privileges access the server hosting the DT to insert malicious virtual resources (e.g., VMs/containers), clone legitimate resources or replace the existing ones with malicious resources. The aim is to take control of a part of the DT model contained in a virtual resource or to take control of the entire DT system, including the physical space. Thus, rogue virtual assets may serve to carry out transitive threats between the two DT spaces (from the digital space to the physical space).
- **[T3.4] Extraction of private information:** Malicious virtual resources may extract information from the system host where they are running, and information from other virtual resources running on the same host. For example, the work in [177] shows how to extract private keys by launching a cross-VM side-channel attack. Similarly, malicious hypervisors may not only be able to take control of the VMs running the DT's services [178], but also to execute introspection techniques. As indicated in [179], a hypervisor may execute VM introspection (permit a VM to observe another VM's memory at runtime) or allow the hypervisor to eavesdrop the activities of all the VMs and steal sensitive information.

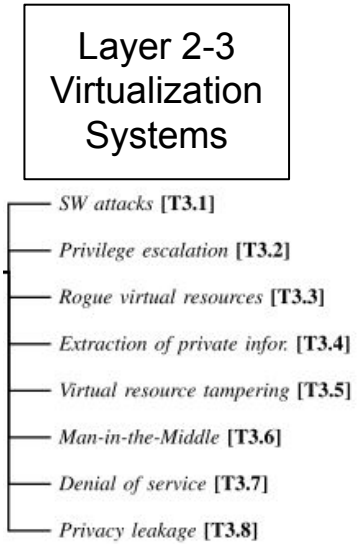
IV. SECURITY THREATS IN THE DIGITAL TWIN



B. Threats at Layers 2-3 - Virtualization systems

- **[T3.5] Virtual resource tampering:** Adversaries with the ability to control the host system, that contains the DT's logic, manipulate sections and actions of the digital assets by compromising their VMs/containers and the hypervisor [179]. For example, they could switch inputs and outputs to corrupt the fidelity level between spaces, desynchronize VMs/containers to impact the interconnection of the digital models, create channels to exfiltrate intellectual property to external entities, inject logic bombs to carry out multiple attacks [87], and saturate shared HW resources such as CPU, cache and memory.
- **[T3.6] Man-in-the-middle:** When VMs/containers need to migrate from one server to another, or replicate their operations at different locations within the system, MitM actions can emerge. This occurs when these operations are carried out through a network infrastructure where adversaries can arbitrate or modify the virtual resources before they are installed on the target node [180]. This last node would include the malicious virtual instances through which adversaries could perform other subsequent attacks, the consequences of which would be similar to those discussed in [T2.6].

IV. SECURITY THREATS IN THE DIGITAL TWIN



B. Threats at Layers 2-3 - Virtualization systems

- **[T3.7] Denial of service:** Any malicious virtual resource (including the hypervisor) can demand additional resources from the server where the DT is deployed [168], [181]. This threat is designed to cause significant overload in terms of communication, computation and storage, such as memory overflow, massive request for HW resources and for connection with other related VMs.
- **[T3.8] Privacy leakage:** VMs and containers can connect to the DT's databases to handle large data volumes associated with the digital and physical assets. If access to these virtual resources is not adequately controlled through strong authentication and authorization mechanisms [94] and through security controls that follow least privilege principles and under regulatory frameworks, multiple attacks against an entity's privacy can occur. In addition, VMs, containers and hypervisors are normally interconnected in a common space, allowing malicious resources to analyze the information flows (e.g., through a cross-VM side-channel attack [177]), in order to locate the most critical virtual resources or derive conduct patterns.

IV. SECURITY THREATS IN THE DIGITAL TWIN

Layer 2-3 Computing techniques

- *SW attacks* **[T4.1]**
- *Extraction of private inf.* **[T4.2]**
- *Privacy leakage* **[T4.3]**
- *Data tampering* **[T4.4]**
- *Knowledge tampering* **[T4.5]**
- *Representation tampering* **[T4.6]**

B. Threats at Layers 2-3 - Computing Techniques

- **[T4.1] SW attacks:** Digital models are an exact SW copy of their physical counterparts, containing specifications, APIs, libraries and source codes. Without a rigorous testing and validation process in terms of design, implementation or adaptation of components (e.g., third-parties' SW pieces), security risks can increase due to bugs caused by bad practices or the cloning of vulnerabilities when copying the SW image of the replicated physical components.
- **[T4.2] Extraction of private information:** Attackers can get sensitive information from the training data and the learning models. This aspect is outlined in [184], which describes how ML models can provide information with respect to a set of training data samples. An attacker can derive sensitive information by: (i) directly accessing the ML model applied and any additional information required (a white-box attack); or (ii) downloading the corresponding model using open APIs together with some information gathered after feeding the inputs (a black-box attack). This also means that once attackers gain access to the target model and its description, they may be able to apply reverse analysis to infer private data.

IV. SECURITY THREATS IN THE DIGITAL TWIN

Layer 2-3 Computing techniques

- *SW attacks* **[T4.1]**
- *Extraction of private inf.* **[T4.2]**
- *Privacy leakage* **[T4.3]**
- *Data tampering* **[T4.4]**
- *Knowledge tampering* **[T4.5]**
- *Representation tampering* **[T4.6]**

B. Threats at Layers 2-3 - Computing Techniques

- **[T4.3] Privacy leakage:** The previous point shows that ML models are susceptible to the extraction of sensitive data through inversion attacks, opening the door to the violation of privacy rights of both the organization and its customers. Here, adversaries may apply reverse engineering to estimate or project new DT states, extract logistical plans and identify vulnerabilities, among other issues. This feature becomes more relevant when the system produces large volumes of data and uses big data techniques with ML algorithms, whose data collectors are able to store such volumes for a long period of time (e.g., edge data centers).
- **[T4.4] Data tampering:** Serious vulnerabilities can arise when data streams are transformed throughout their life-cycle without clear access controls to their structures. Adversaries with previous knowledge of these problems may, for example, prioritize their attack strategies to damage data consistency in terms of fidelity and granularity, and consequently affect the final knowledge.

IV. SECURITY THREATS IN THE DIGITAL TWIN

Layer 2-3 Computing techniques

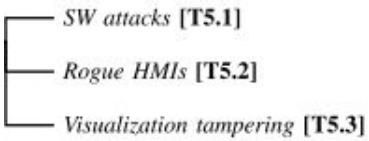
- *SW attacks* **[T4.1]**
- *Extraction of private inf.* **[T4.2]**
- *Privacy leakage* **[T4.3]**
- *Data tampering* **[T4.4]**
- *Knowledge tampering* **[T4.5]**
- *Representation tampering* **[T4.6]**

B. Threats at Layers 2-3 - Computing Techniques

- **[T4.5] Knowledge tampering:** Adversaries with the ability to interfere with a dataset can alter the quality of the classification both in the training phase and in the testing or inference phase. The most notorious threats in the training phase involve injecting malicious samples to generate invalid labels and change the distribution of training data (known as poisoning attack [187]) or directly modify the label values (e.g., through a label contamination attack [188]). The goal is to corrupt the retraining phase by producing malicious samples or reproducing legitimate samples (known as impersonation attack [185]) to consequently redirect the classification or create invalid labels. The result of the threat would correspond with a high rate of false positives or negatives in the classifiers, and an impact on their accuracy.
- **[T4.6] Representation tampering:** Any deviation caused by malware or deliberate disturbances by insiders with abuse of power or escalation of privileges consequently affect the final representation of the data to the end user, such as human operators. Therefore, this threat can be seen as the result of previous threats, mainly focused on changing the fidelity and granularity of digital models and their data.

IV. SECURITY THREATS IN THE DIGITAL TWIN

Layer 4
HMIs



C. Threats at Layer 4 - Human Machine Interfaces

- **[T5.1] SW attacks:** HMIs are systems mainly supported by SW components (e.g., OS, applications and dashboard services) capable of managing and displaying results, and interacting with the physical space, data centers and external infrastructures/systems. The latter characteristic makes them particularly susceptible to penetrations and malware infections. These threats can vary significantly, for example:
  a. producing overheads to disrupt or delay Layer 3 representations;
  b. modifying the level of fidelity and granularity of such representations;
  c. altering specific HMI configurations to complicate extensibility and update processes; or
  d. exfiltrating data, among other security issues. Specifically, these security concerns are detailed in [146], but with a particular focus on AR technology.

IV. SECURITY THREATS IN THE DIGITAL TWIN

Layer 4
HMIs

SW attacks **[T5.1]**
Rogue HMIs **[T5.2]**
Visualization tampering **[T5.3]**

C. Threats at Layer 4 - Human Machine Interfaces

- **[T5.2] Rogue HMIs:** Insiders with full rights to access the IT or OT domains may insert, replace, configure or clone HMIs with a connection to the DT. Through these rogue devices, they may, for example: (i) modify or disable the inputs/outputs values from/to the connected DT; (ii) alter the final data representation in the HMI to conduct invalid conclusions; (iii) block or hinder maintenance of HMIs; or (iv) exfiltrate information to other illicit sources. In [189], the authors demonstrate the influence of a rogue engineering workstation on S7 Simatic PLCs, which impersonates an HMI to later inject malicious messages and execute operations on the control logic.
- **[T5.3] Visualization tampering:** As mentioned above, adversaries with the ability to modify specific HMI settings and services may also modify the final visualization of the Layer 3 representations, as also stated in [146] but with a particular focus on AR. Adversaries may, for example, hide information, show erroneous or inconsistent data, or change the data integrity (e.g., C&C instructions). An example of a deception attack can be found in [154] and [190]. The authors demonstrate how to fool an HMI by stealthily changing the PLC register values to zeros, causing the HMI to present a different reality and forcing the worker to make an incorrect decision.

IV. SECURITY THREATS IN THE DIGITAL TWIN

Summary of section IV:

1. Threats with the **greatest impact** are related to the **deployment of rogue components** followed by MitM, SW attacks and DDoS attacks.
2. Almost **all threats** have some **influence** on the **final consistency of the data**, either in terms of fidelity or granularity. This demonstrates the great weakness of DT technology in critical contexts, where high accuracy in data handling is essential.

TABLE II
IMPACT ON THE DT OPERATIONAL REQUIREMENTS AFTER A THREAT

Threats	Operational requirements of a DT														
	R1.2.1	R1.2.2	R1.2.3	R2.1	R2.2	R3.1	R3.2.1	R3.2.2	R3.3	R4.1	R4.2	R5.1	R5.2	R6.1	R6.2
[T1.3]	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	
[T1.5]					✓					✓		✓	✓	✓	
[T1.6]	✓	✓	✓	✓	✓				✓	✓	✓	✓	✓	✓	✓
[T2.5]				✓	✓					✓		✓	✓	✓	
[T3.5]	✓	✓	✓	✓	✓				✓	✓	✓	✓	✓	✓	
[T4.1]	✓	✓	✓	✓	✓	✓				✓		✓	✓	✓	
[T5.1]	✓	✓	✓			✓			✓	✓	✓	✓	✓	✓	✓
[T5.2]	✓	✓	✓			✓	✓		✓	✓	✓	✓	✓	✓	
[T5.3]										✓		✓	✓	✓	
[T1.1, T1.2]	✓	✓	✓	✓	✓	✓			✓	✓	✓	✓	✓	✓	
[T1.8, T2.8]					✓					✓		✓	✓	✓	
[T2.3, T3.3]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
[T2.6, T3.6]	✓	✓	✓	✓	✓			✓	✓	✓	✓	✓	✓	✓	✓
[T1.7, T2.7, T3.7]	✓	✓	✓		✓				✓	✓	✓	✓	✓	✓	
[T4.4, T4.5, T4.6]												✓	✓	✓	
[T2.1, T2.2, T3.1, T3.2]	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	
[T1.4, T2.4, T2.9, T3.4, T.3.8, T4.2, T4.3]														✓	✓

IV. SECURITY THREATS IN THE DIGITAL TWIN

Summary of section IV:

Table IV, which shows the cascading effect of threats on the functionality layers. For example, a threat [T1.1] in Layer 1 may involve a synchronization variation that implies significant changes in the final management of the digital models included in Layers 2 and 3, with relevant impact on the final representation of the DT (Layer 4). The table also reveals that Layer 1 (included as part of the physical space) is the most affected layer due to the bidirectional link between spaces.

TABLE IV
CASCADING EFFECT OF THREATS ON FUNCT. LAYERS

Threats	Layers			
	1	2	3	4
[T1.4]	●			
[T5.3]	◐			◑
[T2.4, T3.4, T4.2]	●	◑	◑	
[T2.9, T3.8, T4.3]		◑	◑	
[T1.1, T1.2, T1.3, T1.5, T1.6, T1.7, T1.8]	●	◑	◑	◑
[T2.1, T3.1, T5.1, T2.2, T3.2, T2.3, T3.3, T5.2, T2.5, T3.5, T4.1, T4.4, T4.5, T4.6, T2.6, T3.6, T2.7, T2.8, T3.7]	◐	◑	◑	◑

●: affects the physical and digital asset ◐: only affects the physical asset
◑: only affects the digital asset

V. EXPLORATION OF SECURITY APPROACHES

- Research works focusing on protection-related recommendations for the DT paradigm.
- This section explores security approaches that are needed to enhance the protection of DTs.
- Some of those approaches have a strong technical nature while others are more closely related to security management and procedures.

A. Hardware and Software Security

- Interconnection between DT models and Layer 1 may add security gaps.
- These gaps usually originate from HW/SW vulnerabilities.
- Vulnerabilities may be due to a lack of an appropriate design or inadequate validation.
- It is recommended that design approaches:
    a. Ensure a root of trust from the HW (e.g., by using a trusted platform module (TPM) or a trusted execution environment (TEE));
    b. Provide secure programming and good practices;
    c. Establish security design patterns; and
    d. Force verification processes and testing.

V. EXPLORATION OF SECURITY APPROACHES

B. Hardening of DT Infrastructures and Decoupling

- There is a particular need to protect the infrastructures that make up the DT itself.
- In this case, defense in depth constitutes the basis of approaches for protecting DT systems.
- As a first line of defense, isolation and segmentation could be good approaches to bring about the decoupling of simulation functions from illicit or external access.
- DT services spread across the entire computing infrastructure (cloud, fog, edge) may be managed by different network administrators under different security policies.
- It is also essential to pre-establish access limits and the degree of trust of each entity interacting with such DT services.
- In Section III-B, it is discussed the fact that DT connections at Layer 1 and digital assets at Layer 3 must coexist with the environment in which they are deployed.
- This coexistence requires not only understanding the communication protocols and their QoS, but also understanding the type of security that these protocols implement.
- Security hardening also means constantly monitoring the actual usage of DT resources, especially those deployed at Layers 1 and 2.

V. EXPLORATION OF SECURITY APPROACHES

C. Identity, Authentication and Authorization

- DTs are complex systems that characterize real-world physical assets and networks.
- They comprise interfaces and processes, all interacting with each other to achieve a common goal.
- This kind of coexistence, especially for dynamic environments, requires data authentication.
- A DT can add an authentication approach in a local service outside the OT domain or rely on an external one established somewhere at the edge.
- This service would force entities to verify their access from the IT domain, further protecting the underlying operational infrastructure.
- Authorization approaches are also needed mainly because multiple and heterogeneous entities may request access to restricted DT resources.
- These resources can range from IIoT/CPS devices to servers, digital assets (e.g., models, VMs, containers) and databases.
- There are already several approaches that control access rights and privileges in critical systems.
- Hyper-connected DTs may also require access control frameworks based on standardized languages.
- These access protocols can be combined with decision and policy enforcement points.

V. EXPLORATION OF SECURITY APPROACHES

D. Deception, Intrusion Detection and Situational Awareness

- Security risks can arise if preventive approaches are not applied to detect intrusion attempts and penetrations.
- DT technology contains important pieces of intellectual property that must be protected.
- Advanced honeypots could be a suitable approach to protect access to critical OT domains.
- For example, a federated industrial honeypot is proposed to simulate real Modbus devices.
- Situational awareness is the ability to understand what is happening at all times and with a high degree of detail.
- Real-time traceability of attacks should also be implemented.

E. Response and Recovery

- No paradigm is free from errors or completely secure, including DT technology.
- This creates a need to implement resilience measures capable of safeguarding simulation operations.
- Resilience is a relevant protection area for the DT paradigm.
- NIST identifies five protection areas, two of which are specific to response and recovery.

V. EXPLORATION OF SECURITY APPROACHES

F. Event Management and Information Sharing

- Security operations centers (SOCs) are specialized systems overseen by cybersecurity experts.
- SOCs can be based on security information and event management (SIEMs) systems.
- Forensic techniques can recover configurations, states and data.
- DLT technology can be a suitable option to leave immutable traces of the actions taken by digital assets.
- Event management systems could manage shared information belonging to computer emergency response teams (CERTs) CERTs could maintain a shared log of the latest threats and vulnerabilities.
- This information can even be shared across a DLT network.

G. Trust Management

- Establishing trust between collaborative components of a DT is fundamental for creating trustworthy environments.
- Sun et al. present a trustbased aggregation model for DT-driven IIoT scenarios.
- Distributed or centralized trust approaches may, in turn, require a high level of computation and storage.
- The integration of trust mechanisms could improve the decision-making in the DT and facilitate the detection of anomalous conducts.

V. EXPLORATION OF SECURITY APPROACHES

H. Privacy

- Privacy leakage (in terms of data, location and usage) can take place in several ways.
- For example, processing of large volumes of data using big data techniques without appropriate control over the use of these data can lead to significant leaks of relevant information.
- DTs need to be able to determine what information can be shared.
- The type of deployment and the level of access in DT computing infrastructures, together with their databases, are also critical.
- Adversaries can increase their awareness by taking control of several computing subdomains.
- The dynamic nature of the new industries forces us to consider some other aspects in the approaches.
- Human operators, operational processes and CPS devices (e.g., robots) generally perform the same operations following routine movements and actions.
- This allows adversaries to derive behavior patterns or the availability of resources or areas.

V. EXPLORATION OF SECURITY APPROACHES

I. Governance and Security Management

- Since DT technology is used in critical systems, organizations must consider protection measures.
- It is thus essential to establish security controls under regulatory frameworks.
- In this regard, DT-specific standards, such as ISO 23247 [50], [220], [222], should also be broadly considered.
- ISO/IEC 27000 family (for information security management systems) and ISA 62443 must be considered.
- The implicit complexities of industrial contexts and the new relationships that the DT adds to that context create the need to automate the risk management processes to prevent potential threats.
- All these procedures must be part of the security policies that will make it possible to control any access to DT systems.

V. EXPLORATION OF SECURITY APPROACHES

J. Traceability, Auditing and Accountability

- As mentioned in Section II and shown in Figure 2, DTs are composed of multiple layers and technologies.
- If these data are stored correctly, it is possible to track all the activities, events and changes of DTs.
- The concept of traceability can also be applied for contextawareness, in order to explain the contextual states through which a DT (or a part of it) transits.
- These states can vary depending on the application context, where incidents, conflicts, anomalies or attacks can emerge and force the DT to change.
- Traceability is a technique that allows other essential services to be implemented, such as auditing and accountability.
- DLT networks combined with DT technology can be very useful.
- With regard to implementation, traceability (including data provenance), auditing and accountability present serious storage problems.
- Large data volumes produced at Layers 1-4, and significant computational and communication overheads.

V. EXPLORATION OF SECURITY APPROACHES

K. Training and the Human Aspects

- In the OT area, there is a particular lack of training, interest and education in the new ITs.
- Many stakeholders who manage OT systems have a very specific acquaintance of their environments.
- Automated activity controls are recommended to determine the degree of know-how, competence and skills in the appropriate use of DT technology.
- These controls involve monitoring compliance with security policies.

VI. FINAL REMARKS AND FUTURE WORK

- A DT is based on the composition of technologies such as cyber-physical systems, edge computing, virtualization infrastructures, artificial intelligence and big data.
- The confluence of all these technologies when deploying a DT, together with the implicit interactions with its corresponding physical counterpart in the real world, generate multiple security issues that have not yet been sufficiently studied.
- This has motivated us to survey the potential threats associated with the DT paradigm.
- Each layer establishes a set of essential services provided by multiple interfaces, technologies and computation systems.
- A DT is a critical system that can be of great interest to adversaries.
- The fulfillment of its operational requirements must be considered.
- The paper looks at the four functionality layers, where the composing technologies reside, all of them prone to different types of attacks.
- The authors proposed a set of security recommendations and approaches that can help to ensure an appropriate and trustworthy use of DTs.
- Next steps of the research will include a more detailed set of security approaches as well as their specific mapping with the classification of threats.
- Authors intend to implement lightweight defense solutions that help to protect the DT and its deployment.